Shifting the Burden: towards new tests for Language Analysis in the Asylum Procedure

James Alexander Hoskin

Doctor of Philosophy
(PhD)

University of York
Language and Linguistic Science

September 2022

Abstract

The purpose of Language Analysis in the Asylum Procedure (LAAP) is to produce a linguistic assessment of claims to origin by asylum applicants who cannot or will not prove their identity by documentary means. It is used by a number of governments, particularly in Europe.

Considering the discordant state of the literature on LAAP, remarkably little directly relevant experimental work has so far been conducted. The core of this thesis is thus empirical in nature. The ultimate objective of the experiments is the development of novel, supplementary tests for LAAP. These experiments principally investigate the perceptions of native speaker non-linguists (NSNLs), a category to which the vast majority of asylum applicants belong. Perception is an aspect of applicants' linguistic competence that has hitherto been overlooked in LAAP, which analyses only their language *production*.

At present, applicants must demonstrate the authenticity of their language production in a one-shot interview. Interview practices, and the assessment of the derived data by NSNL analysts, have provoked particularly sustained criticism. The best-developed of the proposed tests would both diverge from and complement current LAAP practice by eliciting judgements from applicants themselves, not an NSNL analyst, in identifying their own (claimed) language variety.

The experiments herein primarily concern Syrian Arabic. Results demonstrate, *inter alia*, significantly superior accuracy for Syrian over non-Syrian NSNLs in identifying Syrian speakers. This engenders optimism about the development of the projected new tests. These could be founded on the empirically validated performance of the target test groups—(Syrian) Arabic-speaking NSNLs vs. (non-Syrian) Arabic-speaking NSNLs—in the present thesis and/or related work yet to be conducted. Such tests, in confirming or casting doubt on the assessment of data emerging from the LAAP interview, would add at least one extra layer of validation to extant LAAP practice, thereby enhancing its fairness, rigour and transparency.

1. TABLE OF CONTENTS

2. LIST OF FIGURES

3. LIST OF TABLES

Acknowledgements

My thanks go first to my supervisor, Paul Foulkes, for his counsel, direction and sense of humour over the last five years.

I extend my gratitude also to: Vince Hughes and Sam Hellmuth (TAP members); Vince Hughes and Anders Eriksson (viva panellists); Jeremy Jones and Eleni Petraki for their recommendations; Roderick Martin for the start in the field and subsequent back-up; Tina Cambier-Langeveld for advice and co-authorship; Sascha Schäfer for generosity and patience; Ourooba Shetewi, Ghada Khattab and Rana Almbark for assistance with Arabic; Priska Hubbuch and Simonette Favaro-Buschor for statistics and discussion; Mihaela Stan for statistics; and the masters of the canon for solace and stimulation when ideas ran short.

I owe love and special thanks: to Mum and Dad for the ethos; to my brother, Al, for tending the fire despite all; to The Blob for the beautiful drawings, to The Marn for the cricket and to both for the sunshine and fresh air (may it be so evermore); to Nanione for her loyalty and the many helping hands (may you ever be my right-hand HAND), and to Buni for the crucial month of good cooking and peace. Without the two of you, I genuinely never would have made it.

Author's Declaration

I, James Alexander Hoskin, declare that this thesis is a presentation of original work and that I am the sole author. This work has not previously been presented for an award at this or any other university. All sources are acknowledged as references.

Some material within this thesis has been presented previously in the following publications from the author:

Hoskin, J.A. (2018). Native speaker non-linguists in LADO: an insider perspective. In Nick, I.M. (Ed.), *Forensic Linguistics: Asylum-seekers, Refugees and Immigrants* (pp. 23-40). Malaga: Vernon Press.

Sole author of chapter on the involvement of NSNLs in LAAP at the Swedish agency Verified; contributions to Chapters 2, 3 and 4.

Hoskin, J., Cambier-Langeveld, T. & Foulkes, P. (2020). Improving objectivity, balance and forensic fitness in LAAP: a response to Matras. *International Journal of Speech, Language and the Law*, 26(2), pp. 257–277.

Lead author of article on forensic factors in LAAP; contributions to Chapter 3.

But yet they that have no science are in better and nobler condition with their natural prudence than men, that by mis-reasoning, or by trusting them that reason wrong, fall upon false and absurd general rules.

— Thomas Hobbes, Leviathan

Chapter 1: Introduction

1.1. Preface

Language Analysis in the Asylum Procedure (herein LAAP) has been used by various governments, especially in Europe, since the mid-1990s. Its general purpose is to produce a linguistic assessment of claims to origin by asylum applicants who cannot or will not prove their identity by conventional documentary means.

More specifically, the proper task of LAAP—as in other branches of forensic science—is assessing the probability of the respective evidence for two competing hypotheses (Aitken et al. 2011). The first hypothesis is that an asylum applicant of uncertain origin is an authentic speaker of the variety he/she claims to speak; the second is that he/she is *not* an authentic speaker of the variety in question. In short, LAAP should be (but apparently is not always, at least explicitly) concerned with the forensic evaluation of the relative likelihood of the evidence for two hypotheses. In forensic science generally, the result of this evaluation is known as the likelihood ratio (Aitken et al. 2011; Morrison 2009). The likelihood ratio may be expressed verbally: it need not be strictly quantitative (Aitken et al. 2011). These matters are expanded upon in Chapters 2 and 3.

The literature on LAAP has, since its inception in the late 1990s, been stalked by discord. Accusations of poor and/or obscure practice are persistently renewed and sporadically countered. The proper role of sociolinguistic and forensic considerations in posing and addressing the core LAAP task has also provoked trenchant disgreement. The potentially fraught nature of the LAAP interview is a further object of continuing controversy. Most durable of all have been assertions and counter-assertions as to the competence, or lack thereof, of native-speaker non-linguists (NSNLs) in assessing the claims of asylum seekers.

In this thesis, I take the view that none of these vital questions, or the quality of evidence mustered to support their varied interpretations, should be avoided. The last comprehensive and critical investigation of issues in and approaches to LAAP was attempted by Wilson (2016).

At a remove of six years, a renewed and critical consideration of similar range, updated to comprehend developments since, is fully in order.

Considering the discordant state of the literature on LAAP, remarkably little directly relevant experimental work has so far been conducted. The core of this thesis is thus empirical in nature. Empirical validation, with a focus on demonstrating the performance of various types of listener, is the sole way in which the polemical tone of the LAAP literature—a tenor due in part to the scarcity of LAAP-like experimental work—might be resolved.

The ultimate objective of the experiments herein is the development of novel, supplementary tests for LAAP. These experiments principally investigate the perceptions of native speaker non-linguists (NSNLs), a category to which the vast majority of asylum applicants belong. Perception is an aspect of applicants' linguistic competence that has hitherto been overlooked in real-world LAAP, which instead analyses their production of language.

As I conclude in Chapter 3, it is unlikely at this stage that practitioners will abandon their respective approaches to LAAP. It is for this reason that I frame the projected new tests as supplementary to, rather than as substitutes for, extant LAAP practice. At present, applicants must demonstrate the authenticity of their language use in a one-shot interview. This aspect of LAAP practice, along with the involvement of NSNL analysts, has been subject to sustained criticism. The practice of LAAP is therefore potentially open to the supplementary tests proposed, and partially developed, in this thesis.

The first test—the best developed of the two—would demonstrate the applicant's language perceptions: their capacity to distinguish samples of the claimed linguistic variety from those of others. A test of this type would both diverge from and complement current LAAP practice in eliciting judgements from the applicant him/herself, not from an NSNL analyst. The second test, at an inchoate stage of development, would engage the applicant's production of language: his/her ability to perform multiple accurate repetitions of highly marked speech local to the claimed area of origin. Thus, in the perception test, the asylum applicant would not even be required to speak, while the production test would demand only that the applicant accurately repeat audio-recorded language samples.

It is only in free (albeit guided) language production, such as occurs in the current LAAP interview, that the applicant may produce linguistic features unexpected by the analyst. The applicant's production of unexpected features in the interview may in some cases be misconstrued by the analyst as evidence of language imitation, rather than—for instance—code switching or accommodation to the speech of the interviewer, and thus lead eventually to incorrect rejection of the applicant's claim. Assessment errors of this kind would not obtain, however, if the applicant him/herself were required to judge the provenance of audio samples, and/or to repeat tightly controlled samples of his/her (claimed) own linguistic variety.

In the proposed tests, then, no such errors of judgement could arise. Unlike present LAAP practice, the tests would not be based on an interview, and thus would not require appreciable intervention by an NSNL interlocutor. Neither, since it could be directly compared with validated baseline results on the same task(s) by reference populations of other NSNLs who speak the same variety, would the applicant's performance have to be assessed by an NSNL analyst. In this respect the proposed test designs simultaneously acknowledge and bypass criticisms of the potentially distorted nature of the speech material emerging from LAAP interviews (e.g. accommodation), as well as the controversy over the role of NSNL analysts in judging the material.

The five experiments described in this thesis cover two languages: two experiments are on (Yorkshire) English and three on (Syrian) Arabic. Because of its direct relevance to LAAP, however, the primary focus is on Syrian Arabic. Results demonstrate significantly superior accuracy for Syrian listeners, compared to non-Syrians, in identifying Syrian speakers. Yorkshire born and raised listeners also evince a significant advantage over other listeners in the equivalent, English language task. The former finding in particular engenders considerable optimism as to the utility of future LAAP tests, based on conceptually similar tasks and involving applicants who claim Syrian origins.

In summary, then, the use of multiple tests for LAAP would have at least four benefits. First, the experimental subjects in the thesis are representative of asylum applicants in the sense that they, too, are predominantly NSNLs from a part of the world where applicants commonly

originate. Thus the envisaged novel tests could, if developed further, be founded on the *empirically validated* performance of the target test groups—Syrian Arabic-speaking NSNLs vs. non-Syrian Arabic-speaking NSNLs—in the present thesis and/or in related work yet to be conducted.

Second, supplementary tests, in assisting to confirm or cast doubt on the assessment of data emerging from the initial LAAP interview, would add at least one extra layer of validation to extant LAAP procedure. The incorporation of one or more tests of applicants' perceptions into existing procedure—as well as a possible new test of production—would enhance LAAP's fairness, rigour and transparency.

Third, some of the burden of judgement would be shifted away from NSNL analysts and onto the applicant. In this way, the heated debate on the manner and extent of NSNL analysts' involvement in LAAP—as well as the controversy over the character of speech data derived from the LAAP interview—might be somewhat cooled. This 'shifting of the burden' would be completely in harmony with the burden of proof in LAAP cases, which rests squarely on the applicant (cf. Chapter 2).

Fourth, similar tests could be developed which target Arabic varieties other than Syrian, as well as further varieties and languages of principal interest to LAAP. These could be used in a uniform and standardised manner by any LAAP practitioner.

In the interests of clarity, the two following sections describe in further detail the background to LAAP, as well as its history and some aspects of its practice. In the final section of this chapter I provide an outline of the thesis.

1.2. Language analysis in the asylum procedure: general background

Societal breakdown in the Third World is among the most acute international problems of the first decades of the 21st century. Ultimate causes of the current mass movement of people to Europe, Australasia and North America include poverty, corrupt and oppressive governments,

internecine armed conflict and a rapidly increasing population, especially in Africa—projected to double to approximately 2.6 billion by the year 2050 and to continue growing until at least 2100 (UN 2015, p. 4).

The number of asylum claims has of late experienced a decline compared to its peak in 2015, when nearly 1.3 million people from outside the European Union sought asylum within it, a six-fold increase from 2008 (Figure 1.1). However, it seems unlikely that this relative lull will be of long duration. Estimated overall migration of non-EU citizens to Britain increased by a third between March 2019 and June 2020, though it is unknown what proportion of the total consisted of asylum seekers (Figure 1.2).

Figure 1.1: Asylum applications in the EU, 2008 to 2020 (Eurostat 2021)
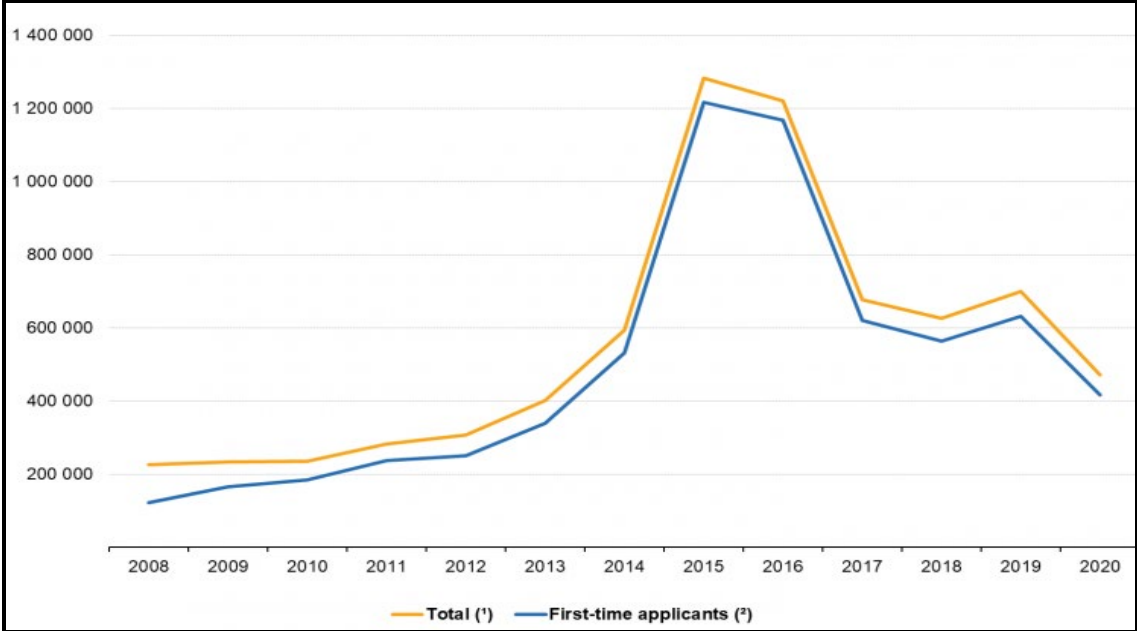
Figure 1.2: Estimated non-EU immigration to Britain, 1991-2021; dark blue line represents Office of National Statistics estimates of LTIM—i.e. long-term international migration ([Migration Observatory 2022](#))
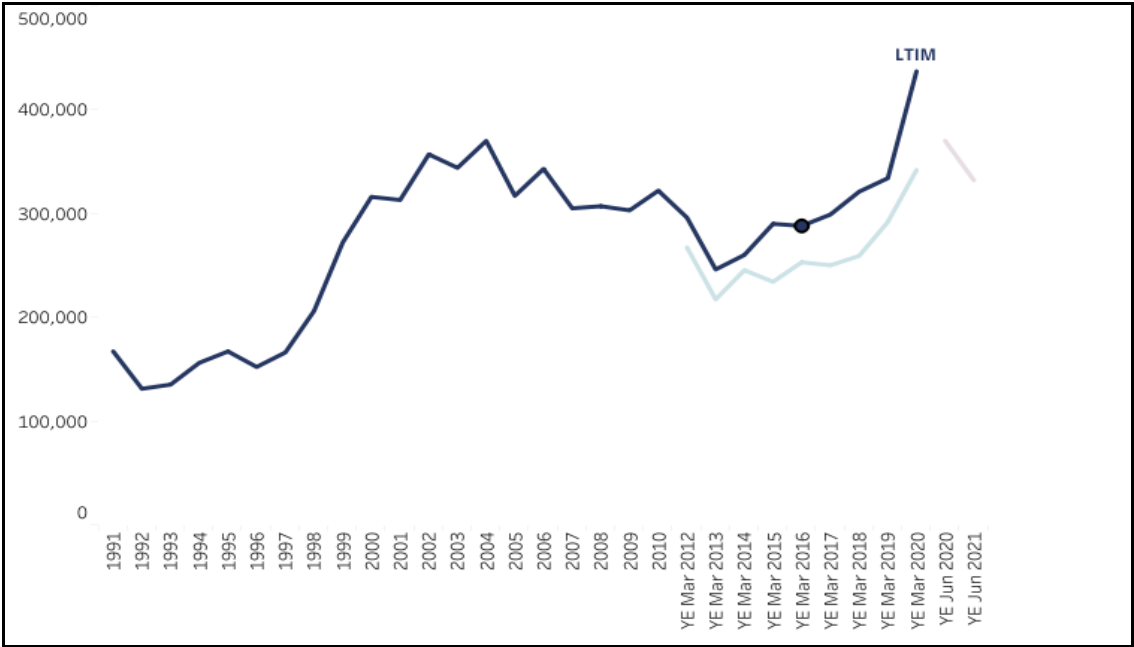
These trends considered, population movement on a similar scale to that of 2015 may be expected to resume, and even to intensify, throughout the third decade of the current millennium. It also seems probable that many people arriving in destination countries will seek asylum under the aegis of the relevant United Nations convention (of 1951) and protocol (of 1967). The conflict in Ukraine which commenced in February 2022, for example, had induced over six million people to seek asylum in the EU by the beginning of August of the same year, with more than 3.5 million of these subsequently registered for temporary protection ([UNHCR Data Portal 2022](#)).

The ultimate causes mentioned above are, however, only partial explanations of the rising trend of mass migration. One proximate cause may be the need for protection from persecution of various kinds. Another may be the desire for improved employment opportunities or access to social welfare services—in short, economic migration. Migration of the latter type is not considered compelling grounds for the granting of asylum under the UN provisions, unless the receiving country deems it so ([UNHCR 2019](#)). Governments which

knowingly grant asylum to economic migrants are, naturally, vanishingly rare. Furthermore, for claims to be considered, most governments (see e.g. Home Office 2018) mandate that asylum applicants must originate in one of a designated range of locations; not all troublespots are considered legitimate sources of asylum seekers.

In other words, a person may seek asylum on the basis that they claim origins in a location in which the ultimate causes of migration are endemic, but this in itself is not enough. They should be from a country, region or ethnic group whose people are deemed by the receiving government to be in need of protection. As the volume of migration into Western countries has grown, accompanied by a large number of claims for asylum, so has the imperative to determine which migrants meet the criteria imposed by the receiving goverment and which are fraudulently claiming a national, regional or ethnic origin identity other than their own. Language analysis was one of the methods developed to assist in the sorting of claims.

1.3. What is Language Analysis in the Asylum Procedure?

1.3.1. A note on nomenclature

Language analysis of asylum claims is today generally known as Language Analysis for the Determination of Origin (LADO). However, some in the field (e.g. Cambier-Langeveld 2016; Hoskin, Cambier-Langeveld & Foulkes 2020, after Reath 2004), have begun using the alternative designation Language Analysis in the Asylum Procedure (LAAP). This acronym is used in this thesis for two reasons, both having to do with the inaptness of the phrase '*Determination of Origin*'.

The first is that LAAP is not normally used as the sole instrument for assessing an asylum seeker's likely origins but is one among many means used to evaluate asylum applications. It is not determinative in itself (Fraser 2019, p. 75).

The second reason is that LAAP is employed in the first instance to investigate to what extent the applicant's language use seems to substantiate their claim to origin, not to determine origin

ex nihilo (Cambier-Langeveld 2016). In this connection Foulkes, French & Wilson (2019), characterising LAAP as a speaker profiling task, point out that only when the asylum claim is found to be unlikely does the task change from *verification* of the speaker's claims to answering the open-ended question of the *classification* of his/her linguistic identity. The latter, not the former, is the type of question implied by the designation 'LADO'. Yet the former question is by far the more commonly addressed of the two, meaning that 'LADO' is an inapposite description of the procedure in most cases.

1.3.2. Definition of the term 'native speaker'

The definition of the term 'native speaker' used throughout the thesis follows two proposed previously in the literature on LAAP. The first is that of Broeders (2010, p. 52), for whom a speaker's native identity derives from "... the group of speakers in which he was socialised and learnt to speak his first language."

The second is that of Cambier-Langeveld (2010b, p. 22):

> "...a native speaker can be defined as a speaker who has first-hand, extensive and continuous experience with the language area and with other speakers of the language and the relevant varieties, starting from an early age"

According to these complementary definitions, 'native speakerhood' is fundamentally a socially acquired, and not a learned or individually elective, category. Thus a speaker of a language, no matter how fluent, cannot be considered a native unless he or she was born and raised in a community in which the language in question is natively spoken and unless he/she has retained contact with some branch of the social organism in which the language is spoken.

The assumption made in this thesis, and by most LAAP practitioners (cf. LINGUA 2021, Cambier-Langeveld 2010b, Verified 2012), is that 'native speakerhood' comprises a set of language competencies and resulting behaviours which, asssuming some degree of enduring embeddedness in the community of socialisation, remain with the individual for life and are thus identifiable via LAAP. A further crucial assumption, which animates the hypotheses

advanced in later chapters of this thesis, is that among these competencies is a generally greater ability (*vis a vis* non-natives) to perceive other native speakers by the way they speak (Cambier-Langeveld 2010b, p. 22). Unlike aspects of language *production*, this superior ability to *perceive* cannot be imitated. Thus, both the perception and the production tests proposed in Chapter 12 of this thesis are based on perceptual data.

1.3.4. Overview of the history and practitioners of LAAP

As far back as the mid-1990s governments were developing novel linguistic means of responding to the then-accelerating trend of mass population movement by assessing the credibility of asylum claims. An ancestor of LAAP was first used in 1993 by the Language Section of the Swedish Immigration Authority to assist in the processing of a rising number of asylum applications associated with civil conflicts in Yugoslavia, Sri Lanka and Somalia, as well as with the end of the USSR (Wilson 2016, p. 18).

The basic assumption of this earliest form of LAAP and its latter-day variants is the same: the uncontroversial principle that the way a person speaks can reveal a great deal about their origins (Eades 2005, p. 506; LNOG 2004, p. 262). However, as demonstrated in Chapters 2 and 3, the nature of the question LAAP should seek to answer, and the resulting approaches to the LAAP task, are hotly contested.

In the years since 1993 immigration agencies in a number of countries—besides Sweden, which pioneered the use of LAAP and continues to use it today—have administered varieties of LAAP, either on their own account or via private companies (Eades & Arends 2004, p. 180; Reath 2004, p. 1; Wilson 2016, p. 21). Among these countries are (or were) Australia, Austria, Belgium, Denmark, Finland, Germany, Iceland, the Republic of Ireland, The Netherlands, New Zealand, Norway, Switzerland and the United Kingdom.

At present five language analysis agencies are known to be in operation. Three of these— Germany's Bundesamt für Migration und Flüchtlinge (BAMF), as well as LINGUA in Switzerland and the Office for Country Information and Language Analysis (OCILA) in The Netherlands—are

government agencies. The other two are private companies: Verified and Sprakab, both based in Sweden. A third private operator, De Taalstudio, was based in The Netherlands but ceased performing LAAP in January 2023.

The Turkish government was, as of 2022, developing a form of LAAP on which it was being advised by various international experts (including the author). In its pilot form, it involves the application of a kind of language recognition software. It is understood that the piloting is being performed on individuals who claim origins in the Xinjiang Uyghur Autonomous Region of Northwest China. At the time of writing, LAAP in Turkey was not yet known to have progressed beyond the pilot stage.

A range of independent linguists, whose number and identity are not fully known, conduct 'counter-analyses' and act as expert witnesses in appeal cases. Some counter-analysts were formerly commissioned by or under contract to De Taalstudio (Verrips 2010), but with the recent cessation of De Taalstudio's operations all presumably now work independently, at the request of the asylum applicant's legal counsel (Lucas 2017; Matras 2018).

Counter-analysis reports are alternative assessments of the applicant's language use, usually founded on the same, interview-derived speech material examined by the LAAP agency whose report was commissioned by a national immigration authority (e.g. Verified for the British Home Office). Counter-analyses are requested by the applicant to contest the findings of a LAAP report which does not confirm the applicant's claims. Among known counter-analysts are: Professor Yaron Matras, who works on Arabic and Kurmanji; Dr Chris Lucas and Professor Enam Al-Wer, each of whom works on Arabic; Professor Derek Nurse, who works on Swahili/Bajuni; Dr Georgi Kapchits, who works on Somali; Professor Peter Patrick, who critiques the general methodology of LAAP reports (Patrick 2019); and Professor Monika Schmid, who has worked on cases thought to have involved language attrition (Schmid 2019).

1.3.5. Overview of the practice of LAAP

Many asylum applicants can produce the sorts of documentation considered suitable to support their identity claim. Some applicants, however, are unwilling or unable to produce any officially acceptable form of identification. Some undocumented applicants additionally assert what the relevant immigration authority suspects to be a false personal history, thus casting further doubt on the legitimacy of their asylum claim. In such cases, LAAP may be requested by the authority as a test to assist in determining the validity of the claim (for a thorough account of UK Home Office guidelines which may lead to a request for LAAP, see Home Office 2018; for a detailed discussion, see Wilson 2016, pp. 12-16).

In LAAP, a recorded interview takes place in which the asylum applicant is given the task of demonstrating linguistic knowledge, and in many cases also local and cultural knowledge. The applicant is asked to speak in one or more language(s) known to be used in the region in which he/she claims to originate. The resulting speech data is then used by the commissioning authority, in conjunction with other evidence (again consult Home Office 2018 and Wilson 2016 for details), to assess the plausibility of his/her claim to origin.

Eades (2005) states that, at the time her article was written, immigration authorities in Germany and The Netherlands were conducting some interviews in a *lingua franca*, such as English, rather than in the applicant's declared native tongue. Whatever its status in the past, this practice is not at present widespread, except in the case of interviews with some African asylum claimants—hardly surprising given the presence of well-established, natively-spoken varieties of English in a number of African states (Bobda et al. 1999) and the cautiously estimated presence of more than 2,000 languages on the continent, a third of the world's linguistic stock (Heine & Nurse 2000).

There are, however, several reasons that it is not always straightforward to use linguistic information, derived from an interview in the applicant's claimed native language(s), to assess an asylum claim. First, language and nationality—or even regional background—do not in all instances have a one-to-one correspondence. Many languages are spoken across national

borders, and it is frequently the case that a specific regional or ethnic sub-variety of a language is relevant to an asylum claim.

Second, the claimant might speak a naturally-occurring mixed variety, or they may switch among two or more linguistic 'codes'. For example, speakers of pronouncedly diglossic languages, such as Arabic, are known to interpolate acrolectal elements in their native variety (Khattab & Foulkes, in press).

Third, due to mobility, applicants' native variety may have undergone attrition (Schmid 2019). Applicants might also accommodate their speech to the variety spoken by the interviewer, or to (their perception of) the official context of the interview (Singler 2004, Channon et al. 2018).

Fourth, there can be difficulty in establishing what language the claimant speaks. The official or scholarly name of the language might not coincide with ordinary usage, or a single language may be referred to popularly and even officially by several names, as is the case with varieties spoken in parts of west-central Africa (Mufwene 2009), Central Asia (Maryns 2004) and the 'Kurdish belt' in the Middle East (Findahl 2018). In addition, there might be no up-to-date documentary records of the language at issue, especially if fieldwork has recently been impossible in troubled regions (e.g. Kurdish and Arabic in Syria).

Speaker-specific factors referred to above, such as code switching and language attrition—as opposed to problems of inconsistent nomenclature and knowledge deficits consequent on the inaccessibility of up-to-date data in areas of interest—are referred to herein as indicators of secondary socialisation, in contradistinction to the primary socialisation which confers native speakerhood.

Speaker-specific factors have been plausibly cited, alone or in combination, as in-principle reasons that authentic applicants can be mistaken for 'fakers'. Equally, however, as Cambier-Langeveld (2010a) asserts, knowledge of such phenomena may serve to deceive language analysts as to the identity of claimants who falsely profess native knowledge of a language and speak it only as a second language or who imitate it when questioned. Distinguishing among authentic and inauthentic applicants, who may evidence a superficially similar range of speech

behaviours and/or claimed language backgrounds, is the central problem for LAAP as presently practiced.

Foulkes & French (2012, p. 559) state, with reference to forensic speaker comparison, which is used in criminal cases:

> "...analysis of vocal features can certainly yield results that have crucial evidential value, although in most cases such evidence is used in a corroborative role alongside other information."

This is also true of LAAP: the literature evidences that, in The Netherlands, Switzerland and the United Kingdom, LAAP is never used as the sole test of a claim's veracity (Cambier-Langeveld 2010b; Baltisberger & Hubbuch 2010; Maryns 2004; Home Office 2018). In addition, my personal experience as a former LAAP practitioner for Verified (2015-17) indicates that authorities in Sweden, Norway, Denmark, Iceland, the Republic of Ireland and Finland use LAAP as one instrument among several in making decisions on asylum applications. The procedures currently used by the Belgian, Australian and New Zealand governments—and, indeed, whether they still employ LAAP at all—are unknown.

At the time of writing Verified is the main provider of LAAP to the British, Danish, Finnish, Icelandic, Irish, Norwegian and Swedish governments, while Sprakab produces reports for the Austrian, British, Danish, Finnish and Swedish governments (Findahl 2018, p. 59). Before its folding, De Taalstudio was principally engaged in conducting counter-analyses on cases where the asylum applicant appeals the findings of an initial language analysis report, usually by the OCILA (Verrips 2010).

Practitioners vary in their approaches to LAAP. In the role of analyst, LINGUA exclusively uses linguists, defined as people with advanced training in linguistics. These analysts are not required to also be native speakers of the language at issue. NSNLs are employed at LINGUA as interviewers and/or *ad hoc* advisors to its analysts (Baltisburger & Hubbuch, 2010). De Taalstudio employed, case by case, linguists with some competence in the relevant language,

though it remains uncertain whether there was any requirement that they also be native speakers (cf. Verrips 2010).

Others—the OCILA, Verified and Sprakab—use a 'team approach' (Cambier-Langeveld 2014; Prokofyeva 2018; Findahl 2018), in which linguists (i.e. those with a university degree in a linguistics or, in some agencies, a language-related field) supervise NSNL analysts. NSNLs are subject to various types of training and supervision by linguists at OCILA and Verified. For reasons of commercial confidentiality, the precise nature of this training is largely unspecified in the literature. However, representatives of all three of these agencies more or less broadly describe a training program for NSNL analysts by which is inculcated knowledge of forensic and/or (socio)linguistic factors relevant to the conduct of LAAP (see e.g. Cambier-Langeveld 2010a, 2010b, 2012; Prokofyeva 2018, Sprakab 2022). NSNL LAAP analysts are therefore emphatically *not* naive informants. This point receives further emphasis in Chapter 4, and the nature of the training given to NSNLs at the three agencies is detailed (to the extent possible) in Chapter 2.

1.4. Outline of the thesis

This chapter has provided a brief description of the background to and practioners/practice of LAAP. In Chapters 2, 3, 4 and 5 I review the relevant literature in detail.

Chapter 2, intended as an expansion on the relevant parts of Wilson 2016, examines the approaches to LAAP of the various known practitioners. Chapter 3 reviews and investigates the nature of the LAAP question and the debate over the 'forensic' and 'sociolinguistic' conceptions of LAAP. Chapter 4 deals with the predominantly theoretical (and some empirical) literature on native speaker perceptions. Chapter 5 examines past empirical findings as to the accuracy and confidence of NSNL judgements. I endeavour throughout to introduce novel evidence, arguments and perspectives on each of these four topics.

Chapter 6 specifies the hypotheses whose strength the experiments are designed to evaluate. These three hypotheses have a dual foundation: general principles of native speaker perception

derived from Chapter 4, and the findings of LAAP-related empirical studies reviewed in Chapter 5. Chapters 7 to 11, the core of the thesis, detail the experimental work conducted as the first steps towards the development of novel supplementary tests. In Chapter 12, preliminary suggestions are made as to how such tests could be designed and their results interpeted. Chapter 13 concludes the thesis with a summary of results and a discussion of possible directions for future work.

Chapter 2: Approaches to LAAP

2.1. Overview

As explained in Chapter 1, Section 1.1, the ultimate aim of the empirical work presented in later chapters is to develop novel, supplementary tests for LAAP. These tests are so designed that the applicant would not be required to engage at all in (free) language production, the domain in which either secondary linguistic influences or language imitation may be manifested. The tests instead mainly examine the applicant's linguistic perceptions, as well as his/her ability to reproduce tightly designed speech samples. In accordance with the premises advanced later in this thesis (cf. Chapter 4), language perceptions are conditioned predominantly by the applicant's identity as a native speaker, or not, of the target variety (cf. subsection 1.3.2). It is predicted that tests centred on perception might help reveal whether unexpected linguistic features emerging from the interview for the primary LAAP test are more likely attributable to secondary linguistic influences or to language imitation.

It is important that the reader understand how the proposed tests are different from (and thus potentially complementary to) existing LAAP practice, which is naturally founded on a range of theoretical assumptions. The remainder of Chapter 2 is devoted to developing the reader's understanding of current approaches to LAAP and the various theoretical assumptions underpinning them, as well as to setting the scene for the review of the relevant literature undertaken in Chapters 3, 4 and 5.

A number of factors are involved in the various approaches to LAAP. Among the most important are the following: the types of personnel recruited and their respective functions; the length and structure of the interview; the composition of reports (including the testing of evidence against hypotheses and the presentation of conclusions); and the manner and degree to which claims of secondary socialisation are accounted for. These factors, along with practitioners' attitudes to the involvement of NSNLs in LAAP, are examined in this chapter. Statistics, where available, are also explored. These reveal interesting patterns which, among other things, illuminate differing rates of acceptances/rejections by the various practitioners.

Section 2.2 reviews the varying approaches to what I have termed 'primary-phase LAAP' (i.e. of the kind commissioned in the first place by various national governments) among the agencies who have to a greater or lesser extent made their practices publicly known: LINGUA, the OCILA, Verified and Sprakab. Section 2.3 details in turn the respective approaches to 'secondary-phase' LAAP, (i.e. counter-analysis): those of De Taalstudio, Yaron Matras and Chris Lucas. Section 2.4 is concerned with alternative approaches to LAAP, as revealed by what is known about the practices of the German and various Southern and Eastern European governments, and contrasts them with the better-known practices addressed in sections 2.3 and 2.4.

2.2. Approaches to primary-phase LAAP: LINGUA, the OCILA, Verified, Sprakab

Detailed information on approaches to LAAP was for the first 20 or so years of its practice exceedingly scarce. As Fraser lamented in 2009:

> "LADO agencies operate under extreme requirements of confidentiality and security ... Very few are willing to reveal their methods or processes even in confidence."

(p. 129)

Since 2010, however, several individuals attached to one or another LAAP agency have published papers outlining agencies' approaches in greater or lesser detail. Similar material produced by counter-analysts has also emerged. Below I discuss the approaches to LAAP adopted by its known current practitioners.

2.2.1. LINGUA

LINGUA, the branch of the Swiss immigration department charged with conducting LAAP, was founded in 1997.

Personnel: recruitment and functions

After some experimentation during the late 1990s and early 2000s with the use of NSNL analysts, LINGUA has settled on the employment of "external analysts": specialist linguists who are competent in the language at issue and who work case by case (Baltisburger & Hubbuch 2010, p. 9). LINGUA does not require that its external analysts be native speakers of the linguistic variety at issue. External analysts perform LAAP on the basis of a sample drawn from an interview conducted either by themselves or by a native speaker of the language in question, who is ususally an NSNL and who also works case by case (Baltisburger & Hubbuch 2010, p. 17). NSNLs are otherwise confined to sometimes advising supervising linguists of their "impressions and observations" (Baltisburger & Hubbuch 2010, pp. 17-18). LINGUA favours the use of trained linguists over NSNLs as analysts due to the complex sociolinguistic profiles which might result from the claimed life histories of many asylum applicants (Baltisburger & Hubbuch 2010, p. 13).

LINGUA also employs full-time "supervising linguists", educated to at least MA level, who have expertise in languages of recurring interest. Supervising linguists train NSNLs in the conduct of interviews and recruit external analysts according to the case at hand (Baltisburger & Hubbuch 2010, p. 15). They test prospective external experts' report-writing skills, using recordings from cases previously conducted, for the quality of their " ... argumentation, the weighting of the different points and the explanations of the linguistic phenomena observed in a subject's language." (Hubbuch, pers. comm.).

After external experts are approved for employment, supervising linguists check that their reports are forensically fit-for-purpose, that they are

> " ... non-ambiguous, that they do not contain any inconsistencies or unfounded statements, that they are relevant for the case at hand and that they are comprehensible also for non-linguists."

(Baltisburger & Hubbuch 2010, p. 16).

A major overarching function of supervising linguists is thus to guarantee forensic rigour. As Baltisburger & Hubbuch recognise, "...not every qualified linguist is by definition a qualified expert. One of the reasons for this might be that many linguists are not used to work in a forensic context" (2010, p. 17). Forensic rigour is further secured by the use of a standard format for LINGUA reports, including a conclusion scale. This assists interpretation of results by judges and asylum case-workers (LINGUA 2021).

Reports are sometimes (at an unspecified frequency) assigned to native speakers of the language at issue—i.e. persons other than those who conducted the initial analyses—for cross-checking of conclusions. It is unknown whether these cross-checkers are in all instances linguistically-trained. Academic linguists are sometimes engaged to perform further quality assurance (Baltisberger & Hubbuch 2010, p. 13).

Interviews: structure and content

Interviews are conducted over the telephone by an NSNL in the applicant's stated or inferred primary language variety. The interviewer and the applicant are linguistically matched according to the applicant's claim to socialisation in a particular linguistic community. Interviews are digitally recorded and last for 45 to 90 minutes (Hubbuch 2019, p. 45).

The interview is constructed to avoid mention of the reasons for the applicant's seeking asylum or his/her other personal data, both of which are considered irrelevant to—or even liable to bias—the analysis of his/her language use (Hubbuch 2019, p 43). Applicants are informed before the interview of its purpose, and scope is given them to speak all the languages that

they claim to know (generally in separate interviews). They are also instructed to speak in their native dialect/variety (LINGUA 2021).

Interviewers adjust the topics discussed in interviews to the "ethnicity, family origin, education and chosen profession or job" of the applicant (Hubbuch, pers. comm.). The applicant's knowledge and speech competence are evaluated with reference to his/her stated background. Interviews are in a "form that is oriented towards a conversation", which "avoid[s] the form of interrogation" (Hubbuch 2019, p. 42).

Reports: composition and hypotheses

Following the interview, the external analyst analyses the occurring linguistic features for their (in)congruence with the variety expected to be spoken by the applicant based on their declared life history. A report on the findings is then produced by the external analyst. In it,

> "…at least eight relevant aspects must be analysed (from at least 2 different, linguistic areas like phonology, morphology, syntax, vocabulary)…Each of these linguistic elements is compared with the variety which is normally to be found in the subject's alleged region/milieu of origin."

(LINGUA 2021)

However, the gathering of linguistic data is not the sole objective of the LINGUA interview; information gleaned from it is also used to assess applicants' cultural and local knowledge of their claimed place of origin. These include,

> "…at least five different aspects of everyday life (e.g. eating habits, geography, administration, religion, clothing habits)."

(LINGUA 2021)

The respective weight assigned in the Swiss government's decision-making to, on the one hand, narrowly linguistic and, on the other hand, wider cultural factors is unspecified in LINGUA's documentation.

The questions LINGUA's reports seek to answer are:

"Does the claimant speak the languages and/or linguistic varieties that are expected on the basis of his/her biography? If not: are there indications that point to an alternative region of socialisation?"

(LINGUA 2021).

To address these two questions—the first a matter of speaker verification and the second of classification, as characterised by Foulkes, French & Wilson (2019)—LINGUA uses a four-point conclusion scale. This includes the ratings 'definitely', 'definitely not', 'most likely' and 'most likely not'. In the event that the applicant's language variety remains unidentifiable and the agency is thus unable to reach a conclusion on this formal scale, it is

" ...explicitly stated that there are not enough elements to come to a conclusion concerning a claimant's main socialization."

(Hubbuch, pers. comm.)

LINGUA's reports are said to test a single hypothesis,

" ... stating what linguistic variety/varieties, language repertoire etc. is expected on the basis of the biographical data supplied by the claimant and on the sociolinguistic profile of the region in question ... "

(Hubbuch, pers. comm.)

Whether this hypothesis is formally stated in LINGUA's reports is unknown. It is clear, however, that LINGUA explicitly takes into account the applicant's entire *claimed* personal history. If, for example, the applicant claims to have grown up in Damascus but to have spent several years in Cairo at some stage, the hypothesis would reflect the possible occurrence of secondary influence from Cairene Arabic.

Some types of secondary influenceare considered more plausible than others:

> "...depending on the biography, certain influences might be acceptable (even expected in some cases) – in other cases however, the influence might be too strong to be explained by the given biography, or they occur on all linguistic levels when, on the basis of the biography, you'd have expected them e.g. mostly on the lexical or maybe also on the phonological level but not really in morphology."

(Hubbuch, pers. comm.)

If the interviewer speaks a variety other than that expected from the applicant, the possibility of speech accommodation is also considered in reports (Hubbuch, pers. comm.). As mentioned above, LINGUA considers that NSNLs are generally unable to take such episodes of secondary socialisation into satisfactory account and that this is a task for which trained linguists are uniquely qualified.

Accounting for possible episodes of secondary socialisation

The possible difficulty for experts in evaluating the authenticity of indications of secondary socialisation was identified in Chapter 1 as the central problem in LAAP. LINGUA is plainly aware that episodes of secondary socialisation may influence the linguistic repertoire of the applicant—whether authentic or inauthentic—and this fact is taken to necessitate the deployment of specialist knowledge.

Attitude to NSNL involvement

The preceding shows that LINGUA's approach reveals no principled objection to some degree of NSNL input, even at the decision-making level, as long as it is subject to the ultimate arbitration of trained linguists with knowledge of the forensic requirements of the Swiss legal system. While their various approaches are otherwise similar to LINGUA's, the remaining three LAAP agencies diverge from it in their allocation of functions: native speaker competence and specialist linguistic knowledge each derive from two distinct individuals, who are separately

responsible for the weighing of evidence, on the one hand, and the reporting of the resulting conclusions on the other (this was referred to above as the 'team approach').

Statistics: confirmations, disconfirmations, no decision

Table 2.1 shows the percentage of cases, referred to LINGUA between January 2016 and September 2018, resulting in confirmations (categories 1 and 2), disconfirmations (category 3) and non-decisions (category 4) as to the applicant's claimed identity. The data show that in all three years approximately as many cases were disconfirmed as confirmed, with a minority (under 15%) of cases left undecided.

Table 2.1: confirmations, disconfirmations, non-decisions by LINGUA analysts, 2016-2018 (adapted from Favaro-Buschor, pers. comm.)

Year	2016	2017	2018 (Jan. - Sept.)
1. Confirmed to one degree or another	37%	44%	39%
2. Partially confirmed	4%	6%	11%
3. Not confirmed	49%	42%	36%
4. Undetermined	10%	8%	14%

A similar pattern holds for 2018-19, in a subset of cases in which the Arabic-speaking applicant claimed origins in Syria. Of these, only eight (i.e. approximately half) were confirmed, and none of the remaining six was the subject of a wholly uncertain conclusion (Hubbuch, pers. comm.).

These statistics are of interest for a number of reasons. First, they suggest that linguistic training (which all LINGUA analysts must have) does not necessarily result in reluctance to reach a decision. Between c. 85% and 90% of all cases in each of the three years were either confirmed or disconfirmed, to one degree of certainty or another. This is notable because, as discussed in detail in Chapter 5, it has been repeatedly claimed in the LAAP literature that trained linguists are uniquely willing "...to say 'I'm not sure'." (Fraser 2019, p. 74). If this were so, one might expect to see the fact reflected in a high percentage of 'undetermined' (i.e. non-decision) cases.

The low percentage of non-decisions is still more remarkable because, as reviewed above, LINGUA is emphatic in its preparedness to weigh carefully applicants' accounts of secondary socialisation against their language use in the LAAP interview. Again, one might expect this be reflected in the form of a relatively high percentage of non-decisions.

Because LINGUA does not specify the relative weighting of linguistic and other factors in its reporting, some or all of the decisions cited above may have been heavily influenced by non-linguistic factors (e.g. applicants' strong or weak knowledge of culture and geography of the region in which they claim to originate). It is, of course, also possible that a majority of the cases involved applicants with relatively straightforward life histories and little in the way of mobility or other factors to confound the analyst. But this would be difficult to reconcile with the insistence of many commentators on LAAP (e.g. Patrick 2012, Eades 2005, Jacquemet 2009) that complex life histories, and correspondingly complex linguistic repertoires, are the norm among immigrants to Europe; see also the even stronger claim by Blommaert (2009) and Arnaut & Spotti (2015) of prevalent linguistic "superdiversity" in this context.

In any case, the above statistics suggest that LINGUA's analysts are far more likely than not to reach a conclusion (to one or another degree of certainty) and that confirmations of the applicant's account are (slightly) more common than disconfirmations. What is *not* known is whether these decisions were correct. It is in this connection that the experiments described in the thesis, with their aim of initiating the development of supplementary tests to increase the accuracy (or at least shore up the certainty) of conclusions, may be especially relevant to LAAP practitioners, including LINGUA.

2.2.2. The OCILA

The Dutch government's LAAP bureau, the OCILA, was instituted in 1999. It uses a team approach to LAAP, and therefore utilises NSNLs to a greater degree than does LINGUA. Otherwise, the respective approaches of the two agencies are similar. Between 2000 and 2008,

the OCILA produced 15,000 reports. This means that in this period LAAP was performed on 10% of all asylum claims made in The Netherlands (Cambier & van den Boogert 2008, p. 1).

Personnel: recruitment and functions

As of 2010, four"supervising linguists", academically trained to MA or PhD level, were directly employed by the OCILA (Cambier-Langeveld 2010b, p. 23-4). Each linguist specialises in a particular language, with responsibilities for other languages as dictated by caseload. One, for example, is an Arabist who also works on East African and Kurdish languages (Cambier-Langeveld 2010b, pp. 23-4). There is no known requirement that linguists be native speakers of the language(s) for which they are responsible.

Like at LINGUA, linguists are in charge of recruitment of native speakers of these languages (Cambier-Langeveld 2010b, p. 24). Supervising linguists are responsible for data collection and research, from "all available sources", about the languages in their purview. Sources include the website Ethnologue, country reports produced by the Dutch ministry for foreign affairs, published dialectological surveys and information reported by native speakers reliably known to be from the area in question (Cambier & van den Boogert 2008, pp. 4-6).

Each linguist works with NSNLs ("language analysts") of the relevant language(s), who are employed case by case. Language analysts are native speakers of at least one dialect of the relevant language and are knowledgeable about related varieties (Cambier-Langeveld 2010b, p. 23). They must also have continuing contact with speakers of their native variety and be well versed in its sociolinguistics (Cambier & van den Boogert 2008, p. 11). All are subjected to criminal background checks.

NSNL language analysts undergo testing for their talent in recognising by ear and analytically describing particular linguistic varieties. Once appointed, they are continuously trained in linguistics, as well as in forensic principles, by the specialist linguist, who must have a particularly keen awareness of the latter. If candidate language analysts are judged to be over- or under-confident in identifying linguistic varieties but are deemed otherwise capable, they are "coached" into greater forensic awareness (Cambier-Langeveld 2010b, p. 25).

"Impossible cases" involving border varieties are sometimes presented to analysts to test their forensic awareness. Other aspects of training include transcription, identifying accents in verified speech recordings and acquaintance-building with the dialectological literature. Analysts and linguists are paired and work on all cases in tandem. Analysts are assessed by linguists in order to determine the geographical range of their expertise: some are suited to working on urban varieties clustered around their home region, while others are capable of working on a wider variety of dialects.

Interviews: structure and content

Cambier-Langeveld (2018b) reports that, in 2016, OCILA began performing 'language indications' on all applicants claiming origins in Syria. 'Indications' involve brief (eight- to 10-minute), in-person interviews of applicants as soon as practicable after their claim is registered. The relatively brief interval between claim and 'indication' (which takes the form of a monologue, its content conditioned by a document listing topics suggested by OCILA) is considered to reduce the risk of the applicant's speech having been tainted, either by preparation or "otherwise". The use of language indications since 2016 is reported as having resulted in a substantial increase in the number of claims being confirmed by OCILA, without a full interview having to be conducted. Statistics are discussed in closer detail below.

In cases where 'indications' are not performed, interviews conducted at OCILA are similar in structure and content to those of LINGUA. References to background information on the asylum claim itself are avoided (Cambier-Langeveld 2010a, p. 81). Interviews last 45-60 minutes (Cambier & van den Boogert 2008, p. 7). Applicants are guided in discussion of five broad topics: geographical features of their birthplace; ethnic background and neighbouring ethnic groups; the applicant's native and other languages; languages spoken in their home region; and aspects of daily life—cuisine, work, education and so on (Cambier & van den Boogert 2008, p. 8).

Like at LINGUA, applicants are strongly encouraged to speak in their native dialect/variety. Unlike at LINGUA, however, interviews are performed by trained civil servants, accompanied by an interpreter who is a speaker of the same broad variety claimed by the applicant; this is

designed to avoid an analyst's both interviewing an applicant and working on his/her case. Prior to the interview, applicants are told of its purpose, and they are encouraged to speak all the languages they claim to know. In the case of "minor languages" (i.e. those where expertise required for a full LAAP analysis are unavailable), the applicant is instructed to read items from a wordlist (Cambier & van den Boogert 2008, p. 9). The channel via which interviews are conducted (in-person, telephone etc.) is unknown.

Reports: composition and hypotheses

NSNL language analysts are charged, according to Cambier-Langeveld (2010a, p. 73), with *judging* the interview data; they also *analyse* the data and compose the resulting report, in conjunction with a supervising linguist. The NSNL analyst working on a given report is matched to the applicant's claim to origin—e.g. a native Damascene (not merely a Syrian native) NSNL will analyse and report on the speech of an applicant claiming to originate in a speech community located in Damascus.

The analyst signs the report, and the linguist provides his/her initials (Cambier & van den Boogert 2008, p. 10). The supervising linguist does final checking of the report for cogency and consistency, and cross-checks are conducted at random by other linguists working in the bureau.

Reports are presented in a standard template. They concentrate on comparison of segmental features (i.e. phonology, lexicon, morphology/syntax and 'other') occurring in the interview with those known to occur in the variety the applicant is expected to speak on the basis of his/her claim. There is no known minimum imposed as to the quantity of features that must be listed to substantiate the conclusions reached in the report (Cambier & van den Boogert 2008, p. 28). Like LINGUA, the OCILA evaluates the applicant's local knowledge (geography, culture etc.) of the area in which they claim origin, but the weight assigned to this evidence, relative to the narrowly linguistic data, is unknown.

As at LINGUA, OCILA uses a fixed conclusion scale: definitely, definitely not, probably, either/or and 'the available data are insufficient to formulate a result' (Cambier & van den Boogert 2008, p. 13). A fixed conclusion scale is employed for forensic reasons—i.e. it is calculated to assist judges in reaching a decision on the question:

"Do the results of the language analysis take away existing doubts about the claimed origin?"

According to Cambier & van den Boogert (2008, p. 13), this question has its foundation in Dutch law. Indeed, by the provisions of The Netherlands' Aliens Act, the burden of proof rests on the asylum seeker:

"An application for the issue of a residence permit for a fixed period as referred to in section 28 shall be rejected if the alien has not made a plausible case that his application is based on circumstances which, either in themselves or in connection with other facts, constitute a legal ground for the issue of the permit."

(quoted in Zwaan 2010, p. 217).

Thus, according to Dutch asylum law, "A lack of evidence...means that the asylum seeker has a problem making his case" (Cambier & van den Boogert 2008, p. 13).

Somewhat surprisingly, given the emphasis in the organisation's documentation on awareness of the forensic context in which LAAP is situated, there is no explicit evidence that a hypothesis is formally stated and tested in OCILA reports.

Accounting for possible episodes of secondary socialisation

At OCILA,

> "The objective of language analysis is *not* to evaluate whether the results can be explained or accounted for by referring to the applicant's own account, but to examine whether the results are consistent with what may reasonably be expected from a person given his claimed background and origin."

(Cambier & van den Boogert 2008, p. 13)

OCILA takes as an additional reasonable expectation the premise that

> "An individual who grows up within a particular language area and who spends his formative years there must become a member of the speech community of that area."

(Cambier & van den Boogert 2008, p. 13)

Explicitly, then, the OCILA does not allow the applicant's declared personal history—i.e. the possible influence of secondary socialisation—to bear on its judgements. Rather, the focus is on identifying his/her primary (i.e. native) linguistic variety. As discussed above, while LINGUA does not specify the extent to which the applicant's story is given *a priori* credence, the phrase "...expected on the basis of his/her biography" (LINGUA 2021) suggests that the applicant's entire declared life history is taken into account. In this, OCILA and LINGUA apparently differ.

Attitude to NSNL involvement

At OCILA, NSNL language analysts are responsible for reaching conclusions (both 'judging' and 'analysing' language). They are supervised by linguists, who assist in the task of analysis but not that of judgement.

Given the centrality of the NSNL role at OCILA, there is no perceptible objection in either principal or practice to their involvement in LAAP, provided they have been proven to have a good 'ear' for language and are sufficiently aware of the forensic context. The latter is undefined, though presumably it includes making appropriately cautious evaluations of the

evidence that are interpretable by linguists in such a way that they can be expressed according to a fixed conclusion scale for use by the court. As at OCILA, however, their role is limited and mediated by the complementary functions allocated to trained linguists.

Statistics: the effect of 'language indications'

Cambier-Langeveld (2018b) reports that in 2014 the OCILA produced c. 100 reports, in 2015 c. 200 reports and, in 2016, in the months prior to October, 350 reports on applicants claiming Syrian origins. From October 2016, language indications were introduced, resulting in the production of 500 reports on Syrian-identifying applicants. Between 2017 and 2020, c. 4,000 additional reports based on indications were produced.

Cambier-Langeveld (2018b) attributes to the introduction of language indications a sharp increase in the number of Syrian claims confirmed via LAAP—from 80% in 2014 to over 99% in 2018—as well as greater speed in assessing claims at a time when considerable pressure is being exterted on OCILA resources. A negative result following the analysis of indications results only in recommendation of a full LAAP assessment of the type described above but never, in itself, disconfirmation of the claim. All indications are performed by NSNL analysts, further emphasising the trust vested by the OCILA in their abilities.

Further notable in these OCILA statistics is the high rate (80%) of confirmations of claims by Syrian applicants, even before the introduction of indications. This is especially interesting when contrasted with the c. 50% confirmation rate by linguist native speaker-analysts at LINGUA. Further observations and discussion of this comparison arises below, in connection with statistics provided by Verified.

Again, though, as discussed above with reference to the available LINGUA statistics, it is not known whether OCILA's indication-based decisions—or, indeed, those arrived at via more thorough LAAP analyses—were correct (but see Cambier-Langeveld 2010a for discussion of a small set of pre-indication cases in which OCILA's conclusions were in every instance confirmed as correct). Once more, this fact, combined with the favourable view of NSNL competence implied by the central place it occupies in OCILA's approach, suggests that the experimental

work described in this thesis has the potential to bear fruit in the real-world practice of LAAP by reinforcing or casting doubt on conclusions reached via existing methods.

2.2.3. Verified

Verified, a private contractor founded in 2004, approaches LAAP (which it refers to as LOID: Language of Origin Identification) in a similar manner to the OCILA. This similarity includes its emphasis on the unique knowledge of the NSNL informant-as-analyst and deployment of a team approach involving NSNLs and trained linguists. Between 2015 and 2017 I was employed by Verified as a linguist. By 2018, Verified had produced more than 100,000 LAAP reports for government clients, including the United Kingdom, Norway, Denmark, Finland, Iceland and Ireland (Prokofyeva 2018, p. 41). It is reported that, as of the same year, LOID had been in use for 'about six years' (Prokofyeva 2018, p. 47).

Personnel: recruitment and functions

To be eligible for employment (on commission) as analysts, NSNLs must present evidence of their erstwhile residence in the country and region of whose language variety or varieties they claim to have grown up a native speaker (Prokofyeva 2018, p. 49). Verified defines a 'native speaker' as, "...a person who was raised with the linguistic variety under analysis being used in the home" (Verified 2012, p. 3).

Analysts are tested and carefully selected for their personal integrity, their aptitude for abstract reasoning and their awareness of and 'feel' for language. They also undergo intensive initial training by academically-qualified linguists, and this continues by various means as their employment progresses. In-house training is aimed at honing analysts' cognisance of the linguistic and forensic factors involved in analysis and interviewing (Hoskin 2018, p. 27). Prokofyeva (2018) provides a thorough overview of the training given to analysts at Verified.

Cross-checking of conclusions—i.e. analysis of the same case by two NSNL analysts of the same linguistic background—is regularly conducted, in the course of recruitment/ training and

casework (Prokofyeva 2018, p. 50). In consideration of the sensitivities of the "forensic context" in which LAAP is conducted, analysts are also subject to police checks (Prokofyeva 2018, p. 49).

Linguists recruited by Verified, who are often not speakers of any of the languages analysed in the agency's reports, must be in possession of at least a Bachelor's degree in linguistics. The definition of 'linguistics' is rather broadly construed; some of the linguists recruited specialise in languages rather than linguistics as a field in itself (Prokofyeva 2018, p. 51; Matras 2018, p. 59). As at OCILA, linguists are engaged in a mainly supervisory role. At Verified this means checking the forensic and formal linguistic rigour of NSNL analysts' conclusions and—on the basis of analysts' judgements—composing and signing LAAP reports.

Also like at OCILA, linguists are responsible for research on language varieties of interest. Similarly, too, sources consulted include [Ethnologue](), dialectological surveys, academic monographs and information from NSNL analysts. This last is considered generally reliable, often carrying as it does greater currency, if not necessarily greater weight in Verfied's reporting, than does the literature (Hoskin 2018, p. 36). In instances where NSNLs and published dialectological sources differ in describing formal properties of key features—or their geographical distribution—linguists must sometimes adjudicate as to the weight these two sources of information should be assigned in the drawing and reporting of conclusions (Hoskin 2018, p. 36). As a general rule, however, reputable published accounts are given priority over native speaker accounts for reference and reporting purposes (Hoskin 2018, p. 36).

Interviews: structure and content

As at LINGUA and the OCILA, interviews at Verified are designed to elicit discussion of cultural and geographical knowledge, and no mention is made of the asylum claim itself. Similar also to practice at LINGUA and the OCILA, the interview at Verified is of a broadly 'sociolinguistic' type: the NSNL interviewer prompts the applicant in discussion of topics of general interest in an attempt to generate natural speech data in the applicant's native dialect/variety.

 Interviews are conducted via videolink, with no picture, either by the 'client' (e.g. a civil servant working for the UK Home Office) or an NSNL working for Verified—though with increasing

frequency by the latter. Interviews should result in at least 15 minutes of speech by the applicant him/herself (Prokofyeva 2018, p. 53). As at LINGUA and OCILA, in order to counteract possible bias, the interview and the analysis are conducted by two different people (Prokofyeva 2018, p. 51).

Also as at LINGUA and the OCILA, applicants are afforded the opportunity to speak in all the languages of which they claim knowledge. Cases involving bilingualism in Kurdish and Arabic have in the past been especially common.

Reports: composition and hypotheses

Final reports are composed collaboratively, between NSNL analyst and linguist (Hoskin 2018; Prokofyeva 2018). The NSNL analyst's native linguistic background and the applicant's claim of origin is matched in the same way at Verified as it is at LINGUA and the OCILA.

In cases commissioned by the Home Office, preliminary reports, based solely on the analyst's intuitions, are returned within 30 minutes of the necessary materials being submitted to Verified. This report is intended as an initial guide for the case officer only, and is never taken into account in evidence; the final report follows within five working days (Wilson 2016).

The final report is based on the conclusions reached by the analyst, presented in a document citing predominantly segmental features occurring the applicant's speech. At least eight discrete features must be included, which can be any combination of observed phonological, morphological, syntactic or lexical units (Verified 2012, p.3). The analyst also gives a rating on a formal conclusion scale of the likelihood that the applicant belongs to the linguistic community most plausibly associated with his/her claim to origin. Analysts frequently incorporate instructive remarks covering characteristics of the applicant's speech beyond the segment (e.g. intonation).

Hoskin (2018) cites several examples; the example reproduced below relates to a non-corroborative conclusion (bolding in original).

"The applicant's way of speaking, his **intonation** at the end of each answer, **repeating** and **unintelligible** words led me to a conclusion that the applicant's dialect is more likely deviating than consistent with the tested hypothesis."

(2018, p. 31)

Further examples of the same general type are catalogued in Chapter 4.

The document produced by the analyst is then interpreted by the linguist. For example, cited features are transliterated into the IPA, and the rating given by the analyst is sometimes amended in consideration of a review of the reference recording, features described in dialectological accounts held in-house and/or further consultation with the analyst. The linguist (but not the analyst) is named in and signs the final report.

Like at LINGUA and the OCILA, Verified reports follow a standard template. They include sections on phonology, morphology, syntax and lexicology, as well as 'other comments', under which heading may appear, e.g., notes on inconsistencies in the realisation of some features referred to in other sections of the report.

On the nature of the hypothesis against which the evidence is considered, Verified reports:

"The purpose of the linguistic analysis is to try the hypothesis that the subject of the linguistic analysis speaks a language or dialect which is consistent with the given place of origin…The linguistic behavior of the subject displayed in the data is examined for consistency with the given speech community being tested. Where the given place of origin is refuted by objective analysis, another alternative hypothesis is generated for testing against the speech data."

(Verified 2012, p. 2)

Verified's reports feature a rating scale, which is also used by analysts in their preliminary ratings. This is reproduced in Figure 2.2.

Figure 2.2: Rating scale, with explanatory preamble, used at Verified (Hoskin 2018)

To the extent given below the person's language use (in the sample) is consistent with the person's claimed linguistic community. Please note that the numbers are used for reference purposes only and do not constitute a linear scale.

+3   The language analysis shows with certainty that the results obtained are clearly consistent with the linguistic community as stated in the hypothesis.

+2   The language analysis clearly suggests that the results obtained most likely are consistent with the linguistic community as stated in the hypothesis.

+1   The language analysis somewhat suggests that the results obtained more likely than not are consistent with the linguistic community as stated in the hypothesis.

0   The language analysis can neither confirm nor refute the hypothesis, as the results obtained do not constitute a basis on which to assess the linguistic community as stated in the hypothesis.

-1   The language analysis somewhat suggests that the results obtained more likely than not are inconsistent with the linguistic community as stated in the hypothesis.

-2   The language analysis clearly suggests that the results obtained most likely are inconsistent with the linguistic community as stated in the hypothesis.

-3   The language analysis shows with certainty that the results obtained are clearly inconsistent with the linguistic community as stated in the hypothesis.

The sole explicitly stated hypothesis evaluated via the above rating scale—which incorporates both verbal and numerical elements (cf. Aitken et al. 2011)—is that the applicant belongs to the linguistic community specified in the report. In the initial (i.e. verification) phase, this hypothesis is formed according to the applicant's claim—as submitted to the LAAP-commissioning government authority—of geographical origin and/or kin-group membership (Prokofyeva 2018). It is worth adding that while the formation of hypotheses by the expert is generally held to be forensically unsound, being as it is susceptible to the intrusion of bias (Aitken et al. 2011), such objections do not properly apply here. Verified's hypotheses in individual cases are merely *worded* by its experts, while their *substance* is determined by the LAAP-commissioning government (i.e. 'Is this person an authentic speaker of their claimed variety?' or similar)—a procedure which is likely applicable also to LINGUA and the OCILA.

The wording of the hypothesis at Verified follows the format:

> "The hypothesis is that the linguistic behaviour displayed is consistent with a [X language] linguistic community that is represented in [Y region or town], [Z country]."

In the verification phase, no alternative hypothesis is considered against the 'null' hypothesis: that the applicant's account is true. Verified's rating scale thus assesses the likelihood of the evidence for *one* hypothesis as to the applicant's likely linguistic background. But this scale on its own is not capable of generating a likelihood *ratio*. This can only emerge from considering the respective probability of the evidence in favour of *two* hypotheses (cf. Aitken et al. 2011; see Section 1.1), for which purpose the inclusion of two rating scales would naturally be required.

A second hypothesis is, however, tested if a rating of -1 or below is entered in connection with the original analysis (i.e. in cases where a classification task is called for). In classification tasks the second hypothesis is suggested by the NSNL in whose judgement the initial hypothesis was not supported. He/she determines impressionistically the linguistic community in which the applicant may have originated. Verified then refers the case to an NSNL analyst of the presumptively relevant variety, after which a new report is composed. It is here, in the classification phase, that the evidence for the second hypothesis as to the applicant's

background is tested. Again though, as in the verification phase, this testing occurs without reference to an alternative hypothesis (Hoskin 2018). The possible shortcomings of this aspect of Verified's approach are discussed extensively in Hoskin, Cambier-Langeveld & Foulkes (2020), as well as in Chapter 3 of this thesis.

Accounting for possible episodes of secondary socialisation

Verfied's published documentation explicitly mentions the potential operation of accommodation, as well as the variable use of more or less prestigious varieties according to context (Verified 2012, p. 4). In order to counteract the possible occurrence of such phenomena in the interview, Verified aims to achieve as close a match as possible between the applicant's hypothesised native variety and that of the interviewer (Prokofyeva 2018, p. 53). As discussed in Hoskin (2018), NSNL analysts are aware of such factors, and are conscientious in alerting linguists to their occurrence.

Secondary linguistic socialisation is thus accorded *some* weight in Verified's reports. However, as Lucas (2017) observes, there is no explicit accounting of *how much* weight is given to such factors.

Overall, the nature of the hypotheses tested shows that the focus of Verified's reporting remains squarely on assessing the likelihood that the applicant is a *native speaker* of a particular variety. In this respect, Verified's approach is fundamentally in agreement with that of the OCILA.

Attitude to NSNL involvement

The same remarks entered above, on the attitude of OCILA to NSNL involvement, apply here. NSNLs are essential to LAAP as performed at Verified, with the caveat that they must be trained and their work checked and refined by linguists.

Statistics: confirmations of Syrian background over a 12-month period

According to a Verified linguist (Stan, pers. comm.), over an unspecified 12-month period— taken to have occurred in the last five years—Verified produced c. 1,600 LAAP reports. A total

of 1,248 (78%) were on applicants who claimed Syrian origins in a specific linguistic community/part of the country (all were either L1 Arabic or L1 Kurmanji speakers). A random survey by Verified of a subset of these 1,248 cases indicates that LAAP confirmed the appplicant's claimed or inferred linguistic background at a rate of 94.5%. The remainder were found, via testing of a second hypothesis, to belong to linguistic communities located in the countries specified in Table 2.2.

Table 2.2: Identified national origins of Syrian-identifying applicants found by Verified not to belong to the hypothesised Syrian linguistic community

Country	Percentage of applicants
Egypt	38%
Iraq	36%
Other Syrian dialects or inconclusive result	20%
Tunisia	2%
Jordan	2%
Morocco	2%

These statistics demonstrate that in the period in question the majority of reports on Syrian applicants substantiated the origin claimed. This occurred in a context in which NSNLs are ultimately responsible for making the vital judgements—with, as discussed above, demonstrated awareness of at least some of the possible effects of secondary socialisation on the applicant's repertoire.

This high rate of confirmation—found similarly in the OCILA statistics previously quoted—is especially notable when compared to the approximately 50% confirmation rate by LINGUA's trained linguist-analysts. The ready inference is that critics of 'team-approach' LAAP, as a source of poorly-motivated rejections of asylum applicants' claims to origin (e.g. Matras 2018), rely for their critique on a very small and selective subset of cases. This amounts to under 6% of Syrian cases in both The Netherlands and the countries under Verified's purview. Given these comparisons, it is striking that LINGUA's practice has attracted very little critical examination in

the LAAP literature, perhaps because of its use of trained linguists as analysts and/or the fact that appeals of its decisions are not well-publicised.

Of course, as elsewhere, there is no guarantee that Verified's conclusions were correct in every instance. In combination with the agency's favourable view of NSNL competence, it seems possible that Verified's approach might amenably accommodate supplementary tests of the type proposed in this thesis. The caveat is that some or all of the national migration boards that contract with Verified would have also to agree to the use of additional testing. The likelihood of such agreement being reached is at this stage impossible to determine.

2.2.4. Sprakab

Information on the approach of Sprakab to LAAP is much scarcer than on those of the three agencies previously reviewed. It was founded in 2000, and as of 2008 had conducted 40,000 'linguistic analyses'. Some of these were presumably of the LAAP type, although the company also produces 'forensic phonetic' reports, apparently including speaker comparison (LINGUA 2009)

Personnel: recruitment and functions

As of 2018, Sprakab was engaged by the British, Danish, Finnish and Swedish governments (Findahl 2018, p. 59). Like the OCILA and Verified, the agency employs a version of the team approach. This involves NSNLs and linguists (who do not speak the language analysed) working in tandem (Findahl 2018, p. 59). No further details are available on the nature of collaboration between NSNL analysts and linguists.

As at all the agencies whose approaches have been examined so far, NSNL analysts must have had wide exposure to the varieties spoken in their region of origin. Subsequent to their engagement, they are trained on the job for at least several months.

Interviews: structure and content

No details are available.

Reports: composition and hypotheses

Each report is the product of the combined work of four individuals, though the functions performed by NSNLs and linguists, and how many of each are involved, is unknown (Wilson 2016, p. 113). Reports are checked in some cases by a second linguist, though the question of how and for what remains open. Reports are also sometimes validated by staff at the OCILA (Wilson 2016, p. 113).

On its [website](#) Sprakab numbers among its services 'linguistic indications', which appear to be similar to those conducted by the OCILA:

> "Linguistic indication is a less ambitious version of language analysis. Linguistic traits are not documented or analysed in detail. The purpose is to get a quick indication of the speaker's adherence to a language/dialect group and geographic region."

[(Sprakab 2022)](#)

Sprakab also claims that "Our forensic phonetics analysis practice adheres to the methodological and ethical guidelines of the International Association for Forensic Phonetics and Acoustics." ([Sprakab 2022](#)). How this bears out in its practice of LAAP remains unspecified.

Accounting for possible episodes of secondary socialisation

No information is available.

Attitude to NSNL involvement

As an exponent of the team approach, there is no suggestion that Sprakab has any objection to NSNLs being involved in LAAP, though as at the OCILA and Verified their role is presumably circumscribed by the complementary functions of linguists.

Statistics

None are available.

2.3. Counter-analysis: De Taalstudio, Yaron Matras, Chris Lucas

A distinction was drawn above between 'primary-phase' and 'secondary-phase' LAAP. The former refers to the work of agencies who produce the initial reports on behalf of national governments (LINGUA, the OCILA, Verified and Sprakab). The latter is undertaken by counter-analysts, at least in part on the basis of the recording of the applicant's speech resulting from the interview held for the primary-phase report. Counter-analysis reports are generally requested by legal representatives of applicants whose claims are not confirmed by primary-phase LAAP. Three organisations/individuals acting as counter-analysts have made material available describing their approach: De Taalstudio, Chris Lucas, and Yaron Matras. None employs NSNLs as analysts, preferring to rely on their own expertise or (in the case of De Taalstudio) that of others engaged to work on case-by-case assignment.

2.3.1. De Taalstudio

De Taalstudio's foundation dates to 2003 (Verrips 2010, p. 280). In September 2022, the organisation's founder, Maiike Verrips, announced on its website that De Taalstudio would cease performing language analysis from 1 January 2023 (De Taalstudio 2022).

Personnel: recruitment and functions

At De Taalstudio a handful of linguists—trained to at least Masters level and specialising in a particular range of linguistic varieties—is charged with the case-by-case engagement of experts in the languages/dialects under their purview to re-examine (or counter-analyse) primary-phase LAAP reports, usually by the OCILA. De Taalstudio's linguists proofread reports produced by the counter-analysts engaged and check the reports for conformity to De Taalstudio's quality criteria (Verrips 2010, p. 284). They also perform cross-checking by, for example, comparing the

conclusions of different experts on the same case (Verrips 2010, p. 288). It is unknown whether De Taalstudio's linguists are required to be native speakers of the languages for which they are responsible.

Experts recruited must have the qualifications specified in the *Guidelines for the Use of Language Analysis in relation to questions of national origin in refugee cases* (henceforth the *Guidelines*). This document was composed and signed by 19 linguists, including Verrips herself, and published in 2004. I refer to the *Guidelines* several times in subsequent chapters; see Wilson (2016) for a comprehesive examination. For the present, it is sufficient to note that the *Guidelines* were expressly intended to

"...assist governments in assessing the general validity of language analysis in the determination of national origin, nationality or citizenship"

(LNOG 2004, p. 261).

As Verrips reports, the *Guidelines* also specify that

"Judgements about the relationship between language and regional identity...should be made only by qualified linguists...holding higher degrees in linguistics [and having a history of] peer reviewed publications..."

(LNOG 2004, p. 262).

In line with the above, "many" experts engaged by de Taalstudio are trained to PhD level and have published in their chosen field. They must also be capable of "a neutral, objective judgement, and...be trustworthy and discreet" (Verrips 2010, p. 281). As with De Taalstudio's linguists, it is unknown whether experts must be native speakers of the languages on which they are engaged to report (though it is implied in, e.g., Verrips 2011 that there is no such requirement).

Verrips claims that de Taalstudio is the only LAAP organisation in whose reports the responsible expert is identified by name (2010, p. 281). However, as discussed above, Verified's analysts are also named in and sign reports, while the OCILA's supervising linguists provide their initials.

Further, in 2010 De Taalstudio switched to anonymous reports, and to using codes in place of the names of experts (Hoskin et al. 2020).

Interviews: structure and content

Most interviews informing the production of De Taalstudio's reports are conducted by the OCILA for primary-phase LAAP and made available for counter-analysis. If insufficient speech material is located in the recording, an additional recording is made by the applicant him/herself. This recording is admissible as evidence in Dutch asylum courts "under certain conditions" (Verrips 2010, p. 282). These conditions are unspecified.

Reports: composition and hypotheses

Like those of the primary-phase agencies, De Taalstudio's reports follow a template. This states, among other things, the applicant's claimed origin, the expert's assessment of the likelihood that this claim is substantiated and sections detailing examples of the applicant's phonology, lexicon, morphology, syntax and proficiency in the languages covered by the interview recording(s) (Verrips 2010, p. 283). The latter suggests that De Taalstudio may produce more than one counter-analysis for each language the applicant claims to speak. Aside from their language production, the applicant's "cultural and geographical knowledge" is also assessed (Verrips 2010, p. 283). Beyond specification of the applicant's claimed origin, there is no evidence that a formal hypothesis is stated in De Taalstudio's reports.

Quantitative expression of conclusions is not permitted, because the *Guidelines* advise that

> "...language analysis does not lend itself to quantitative statistics such as are often found in some others kinds [*sic*] of scientific evidence."

 (2004, p. 263).

Conclusions may be (qualitatively) phrased as experts wish, "...as long as the conclusion is clear, and supported by the arguments presented in the report" (Verrips 2010, p. 283). Examples of

conclusions expressed by counter-experts who have worked for De Taalstudio are reproduced in Chapter 3.

Accounting for possible episodes of secondary socialisation

According to Verrips (2010, p. 289), reports produced by the OCILA, on the basis of exactly the same interview data as assessed by De Taalstudio's experts, reach essentially opposite conclusions at an estimated rate of between 30% and 60%. Verrips also remarks on the frequency with which OCILA reports express a greater degree of certainty than do De Taalstudio's counter-analyses (but see section 2.2.2, above, in which it is shown that, as of 2014, the OCILA confirmed the applicant's claimed linguistic background in c. 80% of cases).

The high rate of past disagreement between OCILA's and De Taalstudio's conclusions is worthy of consideration here (see Hoskin et al. 2020 for an alternative discussion of this subject). As we have seen, both the OCILA and De Taalstudio assess, in addition to language, the applicant's cultural and geographical knowledge. The extent to which this is taken into account in either the counter-analysis or the original report is not known. Yet it is difficult to imagine that the conclusions of the original report and the counter-analysis could differ in their evaluation of the accuracy of such knowledge, making this more likely a source of agreement than of disagreement. The fact that some applicants are assessed on their competence in more than one language makes the frequency of opposing conclusions even more difficult to comprehend, since the assessment, by the OCILA and De Taalstudio respectively, of the applicant's competence in *more than one language* would have to diverge.

For her part, Verrips (2010) suggests that the frequency of disagreement is attributable in part to insufficient knowledge of languages spoken in troubled regions. But it is difficult to understand why both OCILA's NSNLs and De Taalstudio's counter-analysts would not be similarly handicapped in this regard. A more plausible explanation—to which Verrips also alludes (2010, p. 289)—refers to the fundamentally different nature of the question addressed by the two organisations.

The OCILA, as we have seen, largely dismisses the applicant's life-history account subsequent to his/her early linguistic socialisation, on the grounds that the account could be partly or entirely fictional. Instead, the focus is on discerning the authenticity of the applicant's claim to speak a (hypothesised) primary language variety or varieties. As Verrips points out, this leads OCILA to regard the mixing of features in some cases (she uses the example of Somalia) as evidence of the applicant's inauthenticity. Many counter-analysts, however, regard phenomena such as language mixing to be completely non-aberrant. In fact, in some situations (e.g. southern Somalia), counter-analysts would positively predict their occurrence. Furthermore, "Many linguists are...critical of the idea that a 'non-authentic' variety of Southern Somali can be detected reliably" (Verrips 2010, p. 290).

Plainly, then, De Taalstudio's counter-analysts and the OCILA have radically different views of—and a correspondingly different way of accounting for—what I have termed 'episodes of secondary socialisation'. The disagreement extends even to epistemology, with some counter-analysts and sympathetic parties disputing the very notion of authenticity, at least where language is concerned (see for example Blommaert 2009, Maryns & Blommaert 2001, Matras 2018).

This chapter is not the place for in-depth commentary on the relative merits of the two Dutch agencies' positions; this is undertaken Chapter 3. It is admittedly doubtful whether language mixing properly belongs in all cases to the shorthand category 'secondary socialisation' (e.g. it can hardly be said to be 'secondary' where it occurs among native bilinguals). What is worthy of note here is that De Taalstudio's counter-analysts are entirely typical of LAAP counter-analysts generally, in three respects: first, the high degree of trust they vest in the applicant's life-history account; second, and relatedly, the nature of the question that they seek to answer (as stated by, e.g., Patrick 2012); and third, the fact that they often treat evidence of secondary socialisation as tending to favour the applicant's claims.

Attitude to NSNL involvement

De Taalstudio employs only linguists. Though experts working for the company are sometimes native speakers of the languages analysed, they are never NSNLs, and Verrips contends that the

deployment of native speaker competence is not a necessary condition for reliable LAAP (2011, p. 132).

Statistics

No statistics are available.

2.3.2. Yaron Matras

Matras's description of his approach (see Matras 2018) appears broadly consistent with that of De Taalstudio as described by Verrips. Indeed, Matras has performed counter-analyses in The Netherlands as an expert engaged by the company (Hoskin et al. 2020). Until his retirement in 2020, he was Professor of Linguistics at the University of Manchester.

Personnel: recruitment and functions

Matras personally conducts counter-analysis on cases involving Arabic and Kurmanji. He is not a native speaker of either language, though he claims fluency in both (Matras 2021, p. 234). However, he states that in performing counter-analyses he has depended, in an unspecified manner and degree, on the opinions of native speakers trained to postgraduate or postdoctoral level in linguistics (Matras 2021, p. 249).

Interviews: structure and content

Like De Taalstudio, Matras apparently uses the interview recording made by the primary-phase LAAP bureau to inform his conclusions. A second stage is conducted in some cases, with new material elicited from the applicant via a set of questions structured in the manner of a sociolinguistic interview, in order to elicit the sort of fluent and natural narrative style that might be lacking in a formal asylum interview (Matras 2018, p. 57). It is not known whether this second interview is performed in person or through other channels.

Reports: composition and hypotheses

Matras's approach appears to represent an application of the "basic LADO question", proposed by Patrick:

> "How does an applicant's linguistic performance in a LADO context correlate with their history of speech community membership and language socialization?"

(2012, p. 536)

Through this question, the available data is used to assess

> "…the plausibility of the applicant's claim to have been socialised in a particular location or locations"

(Matras 2018, p. 57).

Matras calls his approach 'inductive-dialectological'. Inductive reasoning proceeds from the bottom up, with data gathered and a hypothesis developed to explain them. Deductive reasoning is the converse: the analysis starts out with a hypothesis that gives rise to certain expectations if the hypothesis is true, and data are tested against these expectations. Like De Taalstudio, Matras uses the applicant's life-history account as the *de facto* null hypothesis. His approach to the question at issue is thus better described as deductive than as 'inductive'. I return to consideration of this issue in Chapter 3.

Matras begins his analysis by making observations about linguistic features in the original, pre-recorded LAAP interview, with a focus on those features known to be dialectologically variable. These materials are then compared to reference recordings from other speakers of Arabic and Kurdish, drawn from a wide geographical area. The reference recordings include examples of specific diagnostic phrases and words believed to reveal regional isoglosses. The two databases

of reference recordings are the 'Database of Arabic dialects' and the 'Database of Kurdish dialects'.

As of 19 August 2022 the Kurdish reference recordings were available at https://web.archive.org/web/20220303110723/http://kurdish.humanities.manchester.ac.uk/database-of-kurdish-dialects/). The Arabic recordings, which in early August of the same year were to be found at http://www.arabic.humanities.manchester.ac.uk/database-of-arabic-dialects/, were no longer available on the web.

The Arabic speakers in the corpus are described as "bilingual" (Matras 2018, p. 70). Some samples in the database "were collected in the Middle East while others in the UK" (Matras 2018, p. 75). No location is represented by more than two or three speakers, and Syrian Arabic is represented by three speakers in total. No further demographic information is offered on the Arabic speakers, but the Kurmanji speakers are largely male (78%) and university educated (69%), with 48% aged between 20 and 29 years. It would appear, then, that both the Arabic and the Kurmanji databases are rather too limited to serve as the sole reference for a conclusion to be derived inductively "in relation to the intersection of isoglosses", as Matras suggests (2018, p. 69).

In conformity with the format of reports issued by De Taalstudio, Matras's conclusions are freely-worded. No formal scale is used in his reports, and the range of possible conclusions is in consequence unspecified.

Accounting for possible episodes of secondary socialisation

As we have seen, Matras claims to work with an "open"—i.e. inductive—hypothesis. His discussion of cases (2018, pp. 65-71), however, demonstrates that a further "contextualisation" takes place, in which the linguistic findings are related to the applicant's declared life history, including possible secondary influences. Unlike (to varying degrees) the primary-phase LAAP agencies, Matras disputes whether secondary and primary socialisation can reliably be separated:

"No linguistic theory denies the existence of variation, and no serious and modern theory subscribes to the view of a single-layer 'primary' linguistic socialisation."

(Matras 2021, p. 243)

This statement is consistent with his approach to hypothesis formation/testing, in which the applicant's claimed life-history is taken as the starting point.

Attitude to NSNL involvement

Matras argues that "LADO reports should be authored by qualified linguists who are experts in the language that is being analysed" (2021, p. 236). However, as we have seen, he claims to have relied on the opinions of trained native speakers in conducting counter-analyses. He is thus apparently receptive to the involvement of linguistically-trained native speakers in LAAP but not to input by NSNLs.

Statistics: Overturning of primary-phase LAAP decisions on the basis of Matras's counter-analyses

Matras states that he provided counter-analyses to Verified reports in 17 appeal cases in England & Wales between 2017 and 2021; all involved Syrian appellants, and in all instances the court rejected the primary-phase conclusion and found in favour of the applicant (2021, p. 238). The total number of appeals (and, indeed, cases referred to LAAP) that occurred in this period is unknown.

Nevertheless, given the low number of rejections by the Home Office's main supplier, Verified (under 6%, as covered in subsection 2.2.3), as well as the fact that LAAP is performed only in a minority of asylum cases (c. 5% prior to 2015, according to Lucas 2017), one might infer that the statistics submitted by Matras represent a non-trivial percentage of the total number of appeals. This suggests that, as Matras claims (2021, pp. 339-40), his method and conclusions are accorded considerable juridical credence, at least in England and Wales. Indeed, Matras refers to the "lower threshold of evidence" required in asylum cases (2021, p. 239). This, as he

implies, may be one explanation for the receptiveness of decision-makers in England and Wales to his conclusions.

2.3.3. Chris Lucas

Lucas, currently Senior Lecturer in Arabic at SOAS, University of London, conducts counter-analysis on reports in which the applicant speaks Arabic, exclusively in the British context (Lucas, pers. comm.). In practice, this means that his counter-analyses mostly dispute reports produced by Verified. Other than a seminar uploaded to Youtube and a paper presented at a conference, Lucas has not yet published any material on his approach to LAAP.

Personnel: recruitment and functions

Lucas apparently relies primarily on his own fluency in and knowledge of Arabic, though he has collaborated with native Arabic speakers trained to postgraduate level in linguistics (e.g. Lucas & Ismail 2018).

Interviews: structure and content

Unike De Taalstudio and Matras, Lucas does not conduct a second, sociolinguistic-style interview of the applicant (Lucas 2017). In its place, he deploys a novel assessment device: a lexical test of the applicant.

According to Lucas & Ismail (2018), lexical items are selected for use in this test on the dual basis of their frequency of occurrence and their degree of imitability. The reasoning is that words less frequently used in everyday speech (the example of 'plum' is given) are much less susceptible to imitation than are words encountered with greater frequency (e.g. 'now'). This reasoning is disputable: nouns such as 'plum' are perhaps *more* likely to be the subject of speaker attention than adverbs such as 'now'. Nevertheless, an empirical study produced by Lucas himself suggests that his test has merit in separating true local speakers from non-locals (see below).

Lucas (2017) describes the administration of this test. Pictures of c. 20 objects to which 'lower-frequency' content words refer (*butcher* in Arabic among them) are presented to the applicant, via videolink. In each instance he/she is asked a question along the lines of, 'What is this called in your dialect?' Responses are recorded and compared with the isoglosses attested in a lexical survey, the *Wortatlas der arabischen Dialekte*, written by the Arabic dialectologists Peter Behnstedt and Manfred Woidich and published in installments between 2010 and 2014 (Lucas 2017). Finally, on the basis of a cross-referencing of the applicant's responses with the lexical isogloss maps in the *Wortatlas*, Lucas arrives at a further assessment of the likelihood of his/her claim to origin.

Lucas & Ismael (2018) report the results of a study they conducted to determine the efficacy of the test in separating native-dialect speakers from others. In the first task, 64 native Arabic speakers, whose origins were known, were directed to name images in their native dialect; these were compared with forms found in the *Wortatlas.* In the second task, they were requested to name a smaller selection of these images as would, in turn, a speaker of Syrian, Iraqi and Yemeni Arabic. A total of 42 non-Syrians did the Syrian lexical test, 19 non-Iraqis did the Iraqi test, and three non-Yemenis did the Yemeni test. Over 90% of responses resulting from the first task matched those expected on the basis of the *Wortatlas*, while in the second task matches occurred at 50% or lower. It appears, therefore, that Lucas's test reliably separates genuine from non-genuine speakers of Syrian, Iraqi and Yemeni Arabic.

This finding acted as an encouragement to the experiments conducted for this thesis, since Lucas & Ismail's study similarly examines the linguistic competence of NSNLs. Given the apparent reliability of Lucas's test, it is surprising that he has not yet advocated its use more widely.

Reports: composition and hypotheses

The first part of Lucas's counter-analysis is his response to the findings of the primary-phase report. The second part comprises a detailed summary of the applicant's responses to his lexical test (Lucas 2017). No further detail is available on the composition of his reports.

Accounting for possible episodes of secondary socialisation

Lucas assesses the likelihood of secondary influences in the applicant's speech—principally possible accommodation to the interviewer (Lucas 2017)—in the first part of his counter-analyses. By avoiding free language production, however, Lucas's lexical test is so designed as to bypass secondary influences on the applicant's speech, drawing instead on his/her native-speaker knowledge (or lack thereof) of the target variety.

Attitude to NSNL involvement

As mentioned above, Lucas is known to have worked only with Dr Hanadi Ismail, an Arabic native speaker with PhD-level linguistic training.

Statistics: number of cases offered by appeal lawyers vs. number accepted

Between 2013 and 2017, Lucas had received approximately 20 inquiries from appeal lawyers about his willingness to perform counter-analysis; of these requests, Lucas accepted four (Lucas 2017). All involved Syrians, and three of these concerned reports by Verified.

Lucas presents counter-analyses only when he disagrees with the conclusion of the original report (Lucas, pers. comm.). The above figures are therefore consistent with Lucas's estimate that he has agreed with the conclusions of primary-phase reports by Verified in "most (but not all)" cases (Lucas, pers. comm.). These simple statistics reinforce inferences made above as to the frequency at which asylum claims are confirmed via primary-stage LAAP, as well as the prevalence of appeal cases and resulting confirmation of the original conclusion—at least in the British context.

As we have seen, Verified confirmed the applicant's claimed linguistic background in 94.5% of cases drawn from a random survey of more than 1,200 involving Syrians (although it is unlikely, given Verified's client portfolio, that all of these were British cases). Moreover, Lucas (2017) states that LAAP was conducted on only 5% of applications for asylum in Britain prior to 2015, not all of which were likely lodged by Syrians. On these dual bases it is reasonable to surmise the following. First, the number of claims to Syrian origins found to be fraudulent on the basis

of primary-phase LAAP is negligible. Second, many of the very small minority of cases of this type that proceed to the appeal stage end in reversal of the primary-phase finding. Thus, while false rejections of asylum claims certainly do violence to the truth, in Britain they are probably very uncommon at both the initial and the appeal stage.

2.3.4. Discussion

For the purposes of the work presented in this thesis, an encouraging general finding emerges from the preceding review of the four primary-phase LAAP agencies: the fact that in all instances NSNLs occupy a stable, recognised role. However, the extent of NSNL involvement varies among the four agencies, from a greater role as judges and (secondarily) analysts of language at the OCILA, Verified and Sprakab to a lesser as interviewers and informal consultants at LINGUA.

A second finding of note is that all four primary-phase agencies seek *principally* to identify whether the applicant is an authentic speaker of the language variety/varieties associated with the relevant claim to origin. The main emphasis is never on discovering whether possible secondary linguistic influences can be interpreted as matching the applicant's claimed life history.

However, this emphasis is variable in degree. LINGUA is concerned to take into account the applicant's entire life history ('biography') in forming hypotheses but is cautious in distinguishing likely authentic from likely inauthentic linguistic manifestations of secondary socialisation—though it is unspecified how secondary influences considered likely are distinguished from those thought unlikely. Verified's (single) hypothesis explicitly assesses the likelihood of the evidence that the applicant is an authentic native speaker of the linguistic variety most likely associated with his/her claim to origin, while also taking into some account evidence of secondary socialisation inferred on the basis of the applicant's claimed life history. Again, though, the full range of plausible types of evidence and the manner and degree of their consideration is unknown. The OCILA is most explicit in insisting on the endurance in the applicant's speech of identifiable indications of primary socialisation (i.e. native-speakerhood),

with little-to-no overt weight accorded the applicant's life-history account or the influence of secondary socialisation on his/her speech.

A third finding is that the primary-phase agencies are alike in expressing their awareness of the forensic context in which LAAP is situated. Eriksson (2010a, para. 2) defines "forensic science" as designed to "aid the legal system by providing scientifically based evidence". By this definition, LAAP is indeed ultimately forensic in character: it is conceived to assist immigration officers and/or judges in making decisions. LAAP reports are a type of admissible evidence, gathered in a formal setting and evaluated in an official context—and, in instances of appeal, before a court—to support or dispute a claim: the applicant's account of his/her background.

LINGUA, the OCILA and Verified all provide overviews of the training administered to various categories of employee to develop their understanding of the forensic nature of LAAP. Agencies' awareness of forensic practice is also evident in their use of fixed verbal/numeric likelihood scales, akin to those used in forensic speaker comparison, which assist decision-makers in assessing the strength of the evidence relative to a specified range of alternative conclusions.

The four agencies' awareness of forensic factors is further apparent in their principal emphasis on attempting to determine the applicant's primary language variety (or, in the case of Verified, linguistic community). This is in line with the argument that,

> "…there is good reason to doubt declarations, by asylum seekers or by experts acting on their behalf, which would serve to lessen or explain away the potential differences between true and false claimants. The applicant's story may be tailored to explain possible mismatches between the linguistic data and what one would normally expect given the applicant's claimed origin."

(Hoskin et al. 2020, p. 266).

In their acknowledgement of the forensic nature of LAAP, the approaches of the primary-phase agencies appear at odds with those of counter-analysts. No counter-analyst is known to use fixed conclusion scales; instead, freely-worded conclusions, without reference to the range of

alternatives, are preferred, at least by De Taalstudio and Matras. Neither do counter-analysts generally accept the premise that claimed episodes of secondary socialisation—as suggested by the applicant's life-history account and possibly reflected in their language repertoire—should be subordinated to the aim of identifying the applicant's primary language variety. Some (cf. Verrips 2010; Matras 2021) dispute that there is a meaningful distinction between primary and secondary language socialisation. This means that the question addressed in many counter-analyses reports is a version of the "sociolinguistic LADO question" proposed by Patrick 2012 (see above, 2.3.2, and further discussion in Chapter 3).

The fourth observation relates to the comparison of statistics made available by the various LAAP practitioners. In Britain and The Netherlands at least, this indicates that rejections of Syrian claims represent a very small fraction of both the total number of Syrian applicants and those subsequently referred to LAAP. Furthermore, where an appeal is lodged, the primary-phase LAAP report seems very commonly to be overturned.

It has been argued that LAAP is a "gatekeeping" instrument, employed by governments to deny asylum (Patrick 2012, Campbell 2013). Moreover, it is alleged, the premeditated exclusionary results of LAAP are achieved "at low cost" Patrick (2012, p. 544), through an "economy of scale" (Matras 2018, p. 61). It has even been suggested that

> "...it is problematic to make political and/or bureaucratic decisions on the basis of what is, primarily, a social matter, namely the particular language variety spoken by an individual."

(Eades 2005, p. 507)

In response to this assertion, the general observation could be made that restricting access to public services of all kinds is standard practice in modern bureaucratic states, with access being granted or denied on the basis of numerous social criteria. In Anglophone countries, for example, English-language competence is a requirement for the granting of citizenship or admission to university. But the more germane point is this: if the statistics reviewed above are at all representative, the best that can be said in support of assertions that LAAP is a form of

gatekeeping is that its effectiveness is extremely limited, and that the money expended on it is largely wasted.

It is further notable that the primary-phase assessment which results in such a minor proportion of rejections (c. 5% over one year by Verified, the main supplier to the British government, and c. 80% by OCILA for the Dutch goverment as of 2014) invariably involves an NSNL analyst. In Switzerland, however, where LAAP is conducted solely by trained linguists, the rate of rejection is much higher: between 36% and 49% in each of the three years from 2016 to 2018. Given this evidence, it is curious that so much of the literature critical of primary-phase LAAP is concerned with the danger of false rejection of applicants' claims by NSNL analysts while rarely mentioning the danger of false rejections by academically-trained linguists (e.g. Corcoran 2004, Eades et al. 2003; Eades 2005, 2009; Fraser 2009, Patrick 2010, 2012; Matras 2018, 2021). Notably, too, the possibility of false acceptances of claims, by either kind of analyst, is almost never discussed (see Cambier-Langeveld (2018a) and Wilson & Foulkes (2014) for exceptions). As the experimental work of this thesis demonstrates, however, this kind of error is by no means negligible in its frequency, especially in the judgements of Syrian NSNLs (see especially Chapter 11).

A further dominant, and intertwined, theme in the LAAP-critical literature is the near-unanimity of objections to NSNLs' employment as analysts. The sources of these misgivings—the crux of the longstanding controversy over LAAP—are covered in Chapters 3 and 4. First, however, the known range of alternative approaches to 'LAAP proper' used by various Central, Southern and Eastern European countries is reviewed and discussed.

2.4. Alternative approaches to 'LAAP proper'

2.4.1. Germany: the Federal Office for Migration and Refugees (BAMF)

It is known that BAMF uses a form of LAAP, though little is known of its details. Cambier-Langeveld (2010b, p. 21) reports that BAMF utilises "a linguist specialized in the language in question" rather than a version of the team approach. However, in March 2017, *Deutsche Welle*

reported that, beginning in 2018, BAMF was planning to use "speech analysis software" (SAS; apparently a type of language recognition software) to assist in assessing the origins of asylum seekers. The software proposed was reported to be "based on the same voice authentication technology used by banks and insurance companies". BAMF had been using more conventional LAAP, conducted solely by human analysts, since 1998. The article does not report that LAAP proper was to be replaced by SAS but notes merely that the latter "would help migration officers review the applicants' sources of origin as one of several 'indicators'" (quotes in original).

Professor Monika Schmid, then of the University of Essex, is quoted in the same article as criticising the use of SAS on the grounds of her doubts that suchs software would be fit to distinguish between the authentic repertoire of a subject on the one hand and, on the other, language use resulting from priming by an interviewer, or from accommodation to the interviewer's speech. She also cites a study of her own in which both computer programs and humans were unable to reliably identify native speakers of German when they had been resident for at least five years outside Germany.

While SAS would presumably reduce the costs and labour expended in the asylum process, a number of criticisms, besides Schmid's, may be made of its use in LAAP. Schmid argues that a speaker's authentic repertoire may conceivably be distorted by accommodation to the speech of an interviewer. What she does not mention, however, is the possibility of attempts by asylum seekers at imitatation of language varieties other than their own. This brings us to perhaps the most important limitation of the use of computer programs in LAAP. Eriksson (2010b, p. 89) points out that, mainly because reference databases of the necessary size and complexity are difficult to compile, computer programs are unable to detect voice disguise in criminal cases. A similar objection can be made to the ability of such programs to discern deliberately deceptive language imitation in LAAP.

In principle, there are a number of further possible problems with SAS. First, it is doubtful that, again without input of a very wide range of reference data, it would have the capacity to detect the sorts of fine segmental and suprasegmental distinctions that are thought to separate

authentic speakers of one variety from those of another at the speech community level (Nolan 2012).

Second, detection of the sorts of *individual* phonetic variation, typical both within and among speakers of a given variety (i.e. at the idiolectal and intraspeaker level), would likely be beyond the range of SAS. This is especially so considering the "continuous" nature of speech evidence (Rose & Morrison, 2009), in which indexical features are "statistical rather than absolute, phonetically gradient, or simply not present in a particular speech sample" (Foulkes, French & Wilson 2019, p. 83). SAS may therefore confound speech variation at the individual level with that found more typically at the level of the community, resulting in misleading findings as to the group identity of speakers.

Many of the above criticisms have also been levelled at LAAP as performed by human analysts. But they are still more potent when applied to SAS, depending as it does on necessarily limited data input which cannot replicate the competence acquired by human beings over a lifetime of natural language use and exposure.

2.4.2. Southern and Eastern Europe

A number of national immigration agencies in Southern and Eastern Europe report that they employ alternative forms of linguistic screening, of an apparently perfunctory and impressionistic kind. Cambier-Langeveld (2018a, pp. 8-9) establishes on the basis of a 2013 report by the European Migration Network that in Greece, Hungary, Italy and Spain a verbal judgement by the interviewing interpreter on whether the asylum applicant speaks the expected language variety (as defined by the interpreter) may be used to evaluate the asylum claim's probity.

This characterisation broadly applies also to Spain (Morgades 2010, p. 171), for Spain. While LAAP proper is not officially used by the Spanish authorities in assessing an asylum applicant's background, a verbal judgement may be used informally to corroborate the opinion of other parties involved in the determination of a claim.

Writing of the Italian context, Pretto (2010) says nothing of a role for interpreters in judging the veracity of asylum seekers' claims, except that the then Italian Minister of the Interior did not recommend the use of language analysis *per se* (p. 188) and that the national constitution regards language as a means for the individual to access the right to legal defence rather than as a method for the authorities to make determinations of origin (p. 187). In explaining what this means in practice, Pretto asserts that government-issued literature on the subject of asylum, which applicants are by law entitled to receive, is in fact often unavailable in the language the applicant claims to speak (p. 194). She reports, moreover, a case in which a group of asylum seekers alleged that they had not understood the language in which an interpreter had conducted their primary asylum interview (p. 193).

In any case, as Cambier-Langeveld (2018a, pp. 13-14) argues, a number of problems, beyond the fundamental types Morgades and Pretto mention, are associated with relying on translators or interpreters alone, instead of employing one of the variants of LAAP proper discussed above. Such problems may arise, for example, from translators' lack of forensic training or from the fact that their impressions are neither recorded in writing nor subjected to a second, analytical, opinion.

The comparative rigour and transparency of LAAP is apparent in two further ways. Firstly, it involves the systematic analysis of the asylum seeker's language performance (phonological, morphological, syntactic and lexical) by at least one person who is trained in linguistics. Secondly, in most cases, the population against which the asylum applicant's language use is assessed has its basis in the literature on the language varieties of the region where he/she claims origin. While this, too, is imperfect (it may be out of date or somewhat simplified), it is likely to be a more reliable tool than, for example, the impressionistic judgements of language interpreters, whose level of forensic training is unknown.

2.4.3. Discussion

When the alternatives encountered in Southern and Eastern Europe are considered—and reservations entered about the use of SAS in Germany—it is beyond dispute that LAAP is a relatively rigorous and transparent, though imperfect, form of language assessment. This is especially so given that, as Cambier-Langeveld suggests, infallible methods can be approached but never attained (2018a, p. 7).

However, in common with the alternatives in Southern and Eastern Europe, LAAP is usually performed only on the basis of the applicant's linguistic performance in a single, high-stakes interview. Numerous researchers argue that it is dangerous to rely on the LAAP interview as the sole source of information on an asylum seeker's linguistic repertoire, because of the interrogatory tone of some interviews, the risk of applicants accommodating to interviewers' speech and the inbuilt imbalance in power relations (Blommaert 2009; Eades & Arends 2004; Channon et al. 2018; McNamara, Verrips & van den Hazelkamp 2010; Maryns 2010; Maryns & Blommaert 2001). Assertions have also been made (by e.g. Corcoran 2004; Eades et al. 2003) that there is insufficient knowledge of the dialectology of some languages and that the available sources are largely unreliable.

Such issues, combined with the dispute over the proper importance of secondary socialisation, and arguments as to NSNLs analysts' (lack of) expertise, suggest that the new kinds of auxiliary LAAP tests developed in this thesis are worth pursuing. These tests would reduce the involvement of NSNL analysts, consider only the applicant's primary (i.e. not secondary) socialisation and not involve a human interlocutor—meaning that the possible effects of accommodation, codeswitching, diglossia and other secondary linguistic sociolinguistic phenomena would be reduced to negligibility. Adoption of such tests may have the additional benefit of enhancing the forensic rigour of LAAP, a matter considered in depth in Chapter 3.

Chapter 3: Forensic and sociolinguistic factors in LAAP

3.1. Overview

In Chapter 2 I reviewed the approaches of the various LAAP practitioners. I detailed fundamental differences between the respective approaches of primary-phase agencies and counter-analysts, which reside ultimately in divergent conceptions of the LAAP task.

In this chapter I adopt a critical view of differing views of the LAAP task. I show that, notwithstanding such differences, there is general agreement among practitioners and commentators that LAAP is a forensic endeavour in the sense that it takes place within the legal system. I assert, however, that the forensic 'embeddedness' of LAAP has critical *methodological* consequences which are rarely acknowledged in the practice of counter-analysis—or, in some respects, of primary-phase LAAP. I argue that counter-analysts' practice of matching observed linguistic features to (alleged) episodes of secondary socialisation in an applicant's claimed life history is inconsistent with the essentially forensic nature of LAAP, with its overriding duty to assist decision-makers. Having in mind the potential for better fulfillment of this same principal duty, I show that practitioners ignore (or are unaware of) the necessity of testing the evidence for two competing hypotheses—instead of a single hypothesis—and so producing a likelihood ratio of the two propositions' respective probabilities (cf. Aitken et al. 2011; Morrison 2009).

I then demonstrate that the supplementary tests proposed in this thesis are forensically valid in the sense that they disregard the applicant's claimed life history and focus instead on verifying his/her declared or reasonably inferred primary language variety (herein also referred to as 'substrate' or 'substratal linguistic variety'). On this basis I again conclude, as in Chapter 2, that the tests envisaged have considerable potential utility for improving the forensic rigour of LAAP.

3.2. The contested nature of the primary LAAP task: 'sociolinguistic' or 'forensic'?

In Chapter 1, I defined the primary LAAP task as assessing the likelihood that an asylum applicant of uncertain origin is an authentic speaker of the variety he/she claims to speak. I also cited some of the limitations of scholarship which may make this task challenging in some cases. These include:

- the possibility of non-correspondence between, on the one hand, scholarly or official and, on the other hand, popular names for language varieties; and
- shortfalls in dialectological surveys, often occasioned by a lack of recent fieldwork.

In the same chapter I identified a number of sociolinguistic factors that further compound the difficulty of the task, including:

- the fact that nationality and language are not always co-determinative categories; and
- the possible occurrence in the applicant's repertoire of code-switching, accommodation, mixing or attrition (i.e., broadly speaking, evidence of secondary socialisation), all of which *can* be mistaken for deliberately deceptive language imitation.

As discussed in Chapter 2, the potentially complicating effect of such factors has led some commentators (e.g. Patrick 2010, 2012; Matras 2018) to argue that LAAP should be interpreted through a dominantly sociolinguistic lens, an inquiry that only linguists are equipped to address. This assertion has been countered over the last decade or so by the view that LAAP is in essence a type of forensic speaker profiling, a procedure to which NSNLs are eminently capable of contributing, provided suitable training, checks and supervision by native speakers are in place (e.g. Cambier-Langeveld 2014, Foulkes et al. 2019).

In a limited sense, the divergence between the 'sociolinguistic' and the 'forensic' views of LAAP can be interpreted as a difference in emphasis rather than as a matter of substance. Nobody denies that LAAP reports are legally admissible evidence, and no-one disputes that

sociolinguistic factors must be taken into account in the submission of such evidence. Yet the respective positions taken in the dispute are inextricably bound to differing conceptions of the nature of the question LAAP should set out to answer. Four important methodological distinctions proceed from fundamental differences between the 'sociolinguistic' and the 'forensic' perspectives on LAAP.

3.3. The nature of the LAAP question: attitudes to the applicant's account

The first distinction lies in the framing of the primary question to be addressed in LAAP, which is in turn bound up to divergent attitudes to the applicant's account. Peter Patrick is foremost among those who propose that the "basic LADO question" should be primarily sociolinguistic. As set out in Chapter 2, Patrick (2012, p. 536) frames the question as follows:

> "How does an applicant's linguistic performance in a LADO context correlate with their history of speech community membership and language socialization?"

In this version of the LAAP question, the applicant's entire life-history account is taken to be implicitly credible, and his/her language use is weighed against it. Thus, in counter-analysis reports, manifestations of linguistic performance that might otherwise be construed as evidence of a false claim are interpreted as reflecting episodes of secondary socialisation, consistent with the applicant's story. As shown in Chapter 2, this has been the approach adopted in counter-analysis reports produced by De Taalstudio and Yaron Matras.

Cambier-Langeveld (2014, pp. 377-8) disputes Patrick's framing of the LAAP question as essentially sociolinguistic. She makes the point that the applicant's personal history is precisely the matter in doubt, which it is the task of LAAP to help resolve (bold in original):

> " …the basic LADO question is not a sociolinguistic one but a **forensic** one, the 'history of speech community membership and language socialization' being unknown and subject to investigation."

From this it follows that the applicant's account, other than his/her declared (or inferred) substratal language variety, must be regarded with appropriate balance and objectivity. The

identification of his/her linguistic substrate should therefore be prioritised, with secondary factors being treated critically. This is the approach taken by the four primary-phase LAAP agencies, as detailed in Chapter 2.

The sociolinguistic LAAP question accords no priority to identifying the applicant's substratal language variety, on the grounds that an individual's speech may reflect the fact that "contemporary speech communities increasingly comprise translocal, complex, multilayered, polycentric, and socioeconomically stratified semiotic spaces" (Stroud 2009, quoted in Patrick 2012, p. 537). For such reasons, Matras (2018, pp. 56-7) contends, there is

> "...an emerging consensus among researchers that LADO should address individuals' socialisation history and their way of communicating in a particular setting (that of the interview) rather than place of origin."

Matras also claims that Verified's approach denies "...the possibility of variation, layering and indexicality of linguistic forms" (2018, p. 72). This is inaccurate. In fact, as reviewed in Chapter 2, all primary-phase agencies, to varying degrees, take into account the possibility of secondary socialisation.

The 'sociolinguistic' LAAP question is founded on a series of secure observations derived from sociolinguistic studies dating back to the 1960s (see, e.g., Patrick 2010), as well as simple common sense. Patterns of speech are individually variable, even within recognised communities of speakers. Among individuals, the deployment of variants is crucially contingent on the context of interaction. Further, the range of variants available to the individual may be altered by personal mobility or by other influences such as close family members who speak another language or variety.

None of these general principles is controversial. But it does not necessarily follow that LAAP should emphasise the (secondary) speech characteristics of the individual asylum applicant to the detriment of identifying the (primary) patterns shared by the group to which he/she belongs. Nor should sociolinguistic precepts in themselves lead counter-analysts to doubt whether the applicant's linguistic substrate can (or should) even be detected (cf. Chapter 2).

Indeed, when viewed from a forensic perspective, claims of secondary socialisation cannot be tested, since their effect is inherently unknowable; neither can they be relied upon, since they may be false (Hoskin et al. 2020). For example, there is no way of securely predicting that personal mobility would produce any notable effect on individuals' speech repertoires. This is especially so in LAAP, since such influences are merely *claimed* by the applicant—and it is precisely the task of LAAP to assess the truth of such claims.

Contrasting the very different contexts in which sociolinguistic studies and LAAP occur, Moosmüller (2010b) points out that asylum applicants subjected to LAAP cannot be assumed in all cases to 'cooperate' (i.e. to tell the truth) as do consenting participants in a sociolinguistic study. For example, the applicant may cooperate in that he/she participates in the interview but be uncooperative in the sense of faking a life history and associated acquisition of secondary linguistic features. The latter is commonly reflected in the sort of unsystematic confusion of individual features that arises from lack of control over the variety being imitated (Moosmüller 2011).

By way of a relevant illustration, consider the fact that in the Arabic of Amman (capital city of Jordan), the reflex of Modern Standard Arabic /q/ is [g], while in Syrian urban varieties it is generally [ʔ]. In the course of a LAAP interview, a Jordanian attempting to imitate Syrian Arabic might unsystematically use their native [g] in some instances and urban Syrian [ʔ] (and even [q]) in others. This variability of forms could be interpreted as, e.g., evidence of individual mobility and resulting exposure to Syrian Arabic; it could also be construed as arising from attempted imitation. In Chapter 12, I attempt to describe in detail Moosmüller's typology of imitator behaviour. For the present, it is sufficient to note that it is difficult to separate unstructured language mixing, which arises out of 'partially cooperative' behaviour in the LAAP interview, from the structured type that *may* be associated with fully cooperative applicants (Moosmüller 2011).

As the LAAP practitioner's task is to aid truth-seeking and assist the decision-maker, the aim of objectively and reliably separating fully cooperative from partially cooperative applicants (i.e. imitators) is central (Hoskin et al. 2020). This means that if the applicant's claims of secondary

socialisation are in any way dubious—and, owing to the essentially forensic nature of the LAAP task, they must always be considered so—they cannot be put on an equal forensic footing with the identification of his/her linguistic substrate.

Considering the centrality to LAAP of distinguishing structured from unstructured language mixing, and thereby identifying partially cooperative applicants (i.e. probable imitators), it is puzzling that the *Guidelines* (LNOG 2004) do not mention the possibility that unstructured language mixing may occur in the speech of asylum applicants. This omission is especially surprising because the *Guidelines* are in large part "a rather general treatise on language variation" (Eriksson 2008, para. 3).

Points 9 and 11 of the *Guidelines* are particularly concerned to increase awareness of the possible occurrence in LAAP of structured language mixing—e.g. code-switching and accommodation, as well as other types of mixing that might arise as a result of multilingualism. Neither of these points, however, alerts the reader to the possibility that such behaviours may be manifestations of language imitation. The failure of the *Guidelines* to acknowledge this fact amounts to a severe limitation on their forensic utility; indeed, as Cambier-Langeveld (2010a, p. 87) observes, the *Guidelines* do not even recognise that LAAP occurs in a forensic context.

As stated in Section 3.2, there is no dispute that LAAP takes place in a forensic setting; even the two most forceful proponents of the 'sociolinguistic LADO question' acknowledge as much (Patrick 2018, p. 3; Matras 2021, p. 248). The forensic *context* of LAAP granted, it follows that practitioners should observe their basic forensic duty, which is to the decision-maker above all else.

The only way to fulfill this duty in good faith is to use the tools available to LAAP—including those informed by sociolinguistics—in order to detect the applicant's likely linguistic substrate, rather than to substantiate dubious personal histories (Cambier-Langeveld 2018a). This is especially so because judges, as non-linguists themselves, are not qualified to assess whether claimed secondary influences are genuinely attributable to the applicant's individual life history, or assumed in order to deceive. According to this view, the 'sociolinguistic' LAAP question

implies *a priori* acceptance of the applicant's story. From a forensic standpoint, this must be regarded as misguided.

3.4. Who should 'do' LAAP?

The second methodological difference between the 'forensic' and 'sociolinguistic' approaches concerns the question of who is qualified to perform LAAP. It is a truism that only linguists (though not, of course, *all* linguists) are trained in sociolinguistics. It follows from this that an essentially sociolinguistic LAAP question must be addressed by a trained linguist with at least some knowledge of sociolinguistics, and not by an NSNL. This is the view propounded by—among others—Patrick (2010, 2012, 2016), Eades (2005, 2009) and Singler (2004), all of whom were among the co-authors of the *Guidelines* (LNOG 2004). The *Guidelines* urge that LAAP be performed by "expert" linguists with postgraduate academic training in relevant languages and a record of peer-reviewed publications. Overlooked in this insistence on academic credentials, however, is a specification of *precisely* what kind of expert knowledge linguists should possess.

As Wilson (2016, pp. 52-4) points out, possession of the types of credentials specified in the *Guidelines* would not, without further specification, render a linguist equal to executing the central LAAP task: deciding on the authenticity of the applicant's language use in the LAAP interview. The aptness of the type of expertise vaguely defined in the *Guidelines* is open to doubt on several fronts. Oddly, there is no mention of training to postgraduate level in sociolinguistics, let alone phonetics or its forensic applications. Neither is it specified that a LAAP-ready linguist should be a native speaker of the relevant language or proficient in the particular variety associated with the applicant's claim.

On the question of the expertise optimal for LAAP, a 2009 resolution by the International Association for Forensic Phonetics & Acoustics (IAFPA) offers a counterweight to the *Guidelines* in the following terms:

> "It is not a valid assumption that a native speaker, linguist or specialized linguist is by definition also a qualified analyst, capable of performing the type of analysis [required in LAAP]. Language analysis is a form of forensic analysis that requires additional skills and competence…Specific training and testing is therefore recommended."

(IAFPA 2009)

Moreover, according to Foulkes et al. (2019, p. 103), it is not a legal necessity that witnesses possess certain academic credentials in order to be considered an expert by a judge, at least in the United Kingdom. Relevant knowledge or skill suffices to meet the definition of an expert before a court.

Cambier-Langeveld (2010a, p. 73) goes further than the IAFPA resolution, arguing that only a (forensically trained) native speaker is capable of making the kinds of judgements crucial to LAAP, regardless of their formal linguistic credentials. She is supported in this belief by a leading British forensic phonetician, Francis Nolan (2012, pp. 282), who writes convincingly, in theoretical terms, of the ability of NSNLs to distinguish among otherwise similar dialects by perceiving features which are so subtle and multidimensional that they are below "the threshold of consciousness" and therefore only apparent to native speakers. In this connection Nolan cites studies of NSNLs' perceptions of pitch accent variability in local varieties of Connemara Irish, as well as studies investigating retroflexion in Hindi vs. Dravidian languages and voicing of plosives in Chinese.

This favourable attitude to the exercise of native speaker competence in LAAP is reflected in the 2009 IAFPA resolution, an organisation to which both Cambier-Langeveld—its current president—and Nolan belong. It states:

"In cases involving the analysis of language and speech for the determination of national identity IAFPA recognises the contribution to be made by:

1. Linguists and trained native speakers with the latter working under the guidance and supervision of the former;

2. Linguists with in-depth research knowledge of the language(s) in question."

A positive view of native speaker competence is widely shared in the forensic phonetic community (e.g. Foulkes et al. 2019). NSNL informants were used to decisive effect in the Yorkshire Ripper hoax case in the late 1970s, and they have been employed also in other, lesser-known criminal cases. Stanley Ellis, an authority on English dialects working on the case, operated on the basis of the same theoretical premise advanced by Nolan (2012) and Klein (1988): it is likely that only local lay listeners are capable of perceiving qualities of accent sufficiently subtle to pinpoint a speaker's linguistic substrate with the requisite precision. By consulting such informants Ellis was indeed able to identify, within a matter of a few miles, the origins of the suspect.

Here, then, NSNL judgements were interpreted by a trained linguist—much the same procedure as in the team approach employed by OCILA, Verified and Sprakab (as well as in other applications of forensic phonetics; Foulkes et al. 2019). It is obvious that in this case an excessive preoccupation with secondary influences on the subject's speech would have served only to obscure his true origins and so to limit the forensic utility of Ellis's work. This is precisely the argument made by Cambier-Langeveld in the LAAP context.

3.5. Testing: evidence and hypotheses

The third methodological difference between the 'sociolinguistic' and 'forensic' approaches to LAAP concerns the testing of evidence against one or more hypotheses. While all practitioners

claim to weigh the evidence against at least one hypothesis, the varying ways in which they do so reflect differing conceptions of the LAAP question and the degree of credence given to the applicant's story.

As mentioned in Chapter 1, experts with a forensic background have characterised LAAP as closely akin to speaker profiling. In the first instance, where the asylum applicant makes a specific claim to a particular national origin, this takes the form of a speaker verification task. According to Hoskin et al. (2020, p. 4), speaker verification in LAAP involves:

> "...developing, by deduction, a specific alternative hypothesis for comparison with the initial hypothesis based on the asylum claimant's story."

The other kind of forensic speech profiling task, speaker classification:

> "...employs inductive reasoning to identify the asylum seeker's most likely country of origin, e.g. for repatriation purposes, consequent upon an inability to verify the applicant's account of his/her origins."

(2020, p. 4)

In Chapter 2 it was specified that the primary-phase LAAP agencies LINGUA, the OCILA and Verified are known to perform both speaker verification and speaker classification tasks. At Verified, each of these tasks is executed in relation to consideration of the evidence for a single formal hypothesis: the speaker belongs to linguistic community [X]. A similar though less formal approach to the testing of the evidence is discernable in the approaches of the other two agencies.

Discussing the OCILA, Verified and LINGUA respectively, Hoskin et al. (2020, p. 264), Prokofyeva (2018, pp. 43-45) and Baltisberger & Hubbuch (2010, p. 15) suggest that evidence for the alternative hypothesis—i.e. that the speaker *does not* belong to linguistic community [X]—is informally considered in determining what kind of language expertise is required to conduct each individual case. However, the evidence for an alternative hypothesis is not (known to be) formally evaluated in the reports issued by any of these agencies.

For his part, Matras claims to conduct LAAP (which in counter-analysis consists solely of a verification task) by inductive instead of deductive means. He argues, from the 'sociolinguistic' perspective, that the inherent variability of language use (code-switching, constraints deriving from interview techniques and so on) demands an "open hypothesis", which must be:

" ... formulated in relation to the intersection of isoglosses, giving a holistic perspective on the data that derives inductively from that contextualisation rather than in relation to a pre-set hypothesis ... " (2018, p. 69).

This allows the expert to make,

" ... an assessment of the co-occurrence of features and the extent to which they render the applicant's statement about their own life history and socialisation plausible or not."

(2018, p. 74)

Referring specifically to Verified's approach to hypothesis testing, Matras further contends that

"…framing the question of linguistic background as two juxtaposed hypotheses…risks biasing the results by excluding the possibility that speech can be varied and multilayered"

(2018, p. 53).

However, as pointed out in Chapter 2, Matras is mistaken in this regard. Verified does not 'juxtapose' two hypothesis but introduces a second—single—hypothesis only in classification tasks, where the initial question of verification has returned a sufficiently doubtful conclusion to warrant it. Neither is Verified's hypothesis in the classification phase "pre-set", contrary to Matras's assertions (2018, p. 61; 2021, p. 246). In reality, it is determined by the NSNL analyst who worked on the initial classification task (Hoskin 2018, p. 36). If, for instance, the linguist believes that the applicant is from Egypt, the case is referred to an Egyptian analyst—who, if he does not natively speak the relevant regional variety, would in turn refer it to a third analyst who does. The fact that evidence for an Egyptian origin is assessed in all classification tasks seen by Matras (2018, p. 61) does not mean that the same occurs in all tasks of the same type

or—still less—that the hypothesis is "predetermined" to fit an "economy of scale" (Matras 2018, p. 73).

Also set out in Chapter 2 is the argument that Matras does not, in fact, employ an inductive ('open') hypothesis. In this sense, the distinction he makes between his and Verified's approach to hypothesis formation is without substance. His approach, no less than Verified's in the first instance, is addressed to the task of (deductive) verification, not (inductive) classification, of the speaker's likely linguistic identity.

In spite of Matras's claims, then, Verified (along with the primary-phase agencies in general) and counter-analysts are alike in formulating a single hypothesis to address their respective versions of the LAAP question, whether 'forensic' or 'sociolinguistic'. In spite of Matras's claims, then, Verified (along with the primary-phase agencies in general) and counter-analysts are alike in formulating a single hypothesis to address their respective versions of the LAAP question, whether 'forensic' or 'sociolinguistic'. The difficulty with this, however, is that no LAAP practitioner is thus capable of generating likelihood ratios, which would weigh the evidence for two propositions in turn, thereby better aligning LAAP practice with broadly agreed forensic precepts (Cf. Aitken et al. 2011, Morrison 2009) and fulfilling the expert's fundamental duty to the decision-maker.

The real difference in the approaches of primary-phase agencies and counter-analysts lies instead in divergent attitudes to the applicant's story. While Matras and De Taalstudio appear to take a non-skeptical reading of the applicant's entire claimed life-history as the single hypothesis to be addressed, the primary-phase agencies construct a single hypothesis against which to weigh evidence of the applicant's (probable) linguistic substrate. These differing attitudes to the applicant's account were comprehensively examined in 3.3. The remainder of this subsection addresses the possible consequences of single hypothesis testing in the verification cases most commonly encountered in LAAP, both at the primary phase and in the course of counter-analysis.

Hoskin et al. (2020) argue that verification cases logically involve the testing of the respective evidence for two hypotheses, without which procedure an opinion on the strength of the

evidence cannot be properly expressed. In the field of forensic speech science, it is widely accepted that evaluating the strength of evidence requires consideration not just of evidence that would be expected if the hypothesis being evaluated is supported but also of evidence that would be expected if the hypothesis is *not* supported (Hoskin et al. 2020). Evidence adduced in favour of a single hypothesis is of limited value at best, in that it confirms the claimant's version of events only to the extent that its plausibility cannot be entirely dismissed (Hoskin et al. 2020). As emphasised repeatedly above, it is only by considering the relative strength of the evidence for *two* hypotheses that the expert can generate a likelihood ratio, and thus fully discharge their responsibility to the decision maker.

The initial question of verification in LAAP, whether it is addressed by primary-phase agencies or by counter-analysts, should therefore consider the linguistic evidence in light of both the hypothesis that the applicant's story is true and an alternative hypothesis that the claimant's story is false. Even a positive finding at this stage, i.e. a report supporting the claimant, should show that it has evaluated the observations in favour of/against the first hypothesis in *explicit* relation to those in favour of/against the second, in the form of a verbal (and not necessarily a quantitative) likelihood ratio (cf. Aitken et al. 2011).

The expert might, for instance, observe a set of vowel and consonant pronunciations that largely conform to known patterns of Syrian Arabic. On this basis, it may be concluded that the evidence is consistent with the hypothesis that the speaker is Syrian. But could those patterns also be found in other dialects of Arabic? Without explicitly assessing this alternative, the expert cannot judge to what extent the observed patterns support the claimant's story. To what degree are they consistent? Do they contain relatively unremarkable and non-distinctive features, or do they contain a set of shibboleths? In short, without comparing the strength of the evidence for hypothesis 1 against that for hypothesis 2, the expert cannot determine the likelihood that the applicant's claim is true.

As far as can be ascertained, however, formally weighing the evidence for *two* hypotheses in turn—and in so doing generating a likelihood ratio—is absent from the approaches of all known LAAP practitioners at the initial (i.e. verification) stage. Given the limited conclusions that can

logically be drawn from exclusively examining the evidence for a single hypothesis in verification cases, an obvious risk in the LAAP context is that such a procedure could result in a large number of false acceptances. It is entirely possible, for example, that the high proportion of claims to origin substantiated by Verified (c. 95% in a one-year period; see Chapter 2) includes more false acceptances than would have been occasioned by an approach considering the strength of the evidence for two hypotheses in turn. A similar caution applies to counter-analyses. Militating against this possibility, however, is the relatively large proportion of claims rejected by LINGUA over a three-year period (c. 35 to 40% over three years; see again Chapter 2). Whatever the truth of the matter, LAAP practitioners might better guard against the risk of false acceptances by incorporating the weighing of the evidence for two hypotheses, expressed as verbal likelihood ratio, into their respective approaches.

3.6. The expression of conclusions

The fourth and final methodological difference between the two approaches to LAAP has to do with the expression of conclusions. This was mentioned in Chapter 2 but warrants further coverage here.

Three of the four primary-phase agencies (Verified, the OCILA and LINGUA) are known, through their invariant use of fixed conclusion scales, to be in harmony with standard practice on expert testimony, at least in the United Kingdom (Criminal Evidence Rule 19.4(f)(i); Ministry of Justice 2015), as well as established forensic practice generally (Hoskin et al. 2020). See Chapter 2 for specification of the fixed conclusion scales employed by these three agencies; at Verified, for example, the available range is from +3 to -3.

However, neither Matras not De Taalstudio express conclusions in line with such practice. Their conclusions are instead freely worded. Hoskin et al. (2020, pp. 270-71) list a range of freely-worded conclusions submitted by one counter-analyst (not Matras) over the course of several years. They are reproduced below.

- (…) overwhelmingly likely she is a [nationality] [ethnicity].

- I conclude that the most plausible explanation for the combination of his language performance and his considerable local knowledge of [island name] is that the claim that he lived on [island name] for most of his life is correct.

- I believe she is from [country X].

- The applicant is a native [ethnicity] and [ethnicity] speaker from [place name], [country X]. I am absolutely sure.

- Most likely [country X].

- [country X].

- I would say with certainty he is from [island name].

- I am sure she is from [island name], southern [country X].

- I would say he is surely a [ethnicity]. His local geographical knowledge suggests strongly he is from [place name].

- A native [ethnicity] from [island name], southern [country X].

- Some limited non-linguistic data about [place name], which I would say give some support to her claim to be from [place name].

- I would say with considerable confidence that he is a native [ethnicity] speaker, and with some confidence that he is from [place name], southern [country X].

- I'd say with confidence that he is [nationality] [ethnicity] from [place name].

- [place name], without a doubt.

- [island name], southern [country X], possibly [place name] village. High probability.

- I have no hesitation in saying he is from [place name].

Hoskin et al. (2020, p. 270) note that some of the above statements, in expressing conclusions of high certainty, mention specific villages—presumably the very location in which the applicant in question claimed origin. The wording of these conclusions, without reference to other fixed scalar alternatives, may simply reflect lack of forensic awareness. However, barring extraordinary expertise on the part of the counter-analyst concerned, and/or an especially distinctive local dialect, the *confidence* expressed is difficult to account for without suspecting the intrusion of bias.

At the very least, fixed conclusion scales have the advantage that they allow the decision-maker in forensic contexts to see where the strength of the conclusion lies relative to the range of conclusions that the expert is prepared to draw from the available data. There is no obvious reason that a dominantly 'sociolinguistic' view of the LAAP question should prevent conclusions being expressed in a fashion consistent with established forensic practice. Yet no counter-analyst is known to do so.

Matras (2018) nevertheless makes some valid points in his critique of Verified's approach. One of these concerns the opacity of the weighting of features cited in Verified's reports, and how their (non-)occurrence contributes to the overall conclusion expressed on its scale. Matras also correctly observes that there is no apparent quantitative relationship between the individual features cited and the conclusion reached. Indeed, the ready impression is that the analysis of individual features is in a sense retrofitted to the overall judgement. Furthermore, considering the caution generally associated with conclusions expressed in a legal context, one might add that Verified's scale is forensically flawed in assigning an excessive degree of certainty to conclusions at the extremes ("the analysis shows with certainty..."; see Chapter 2 for a verbatim reproduction of the scale).

These observations do not bear on the reliability of the work underlying Verified's conclusions. They do, however, prompt the question of whether the format of the agency's reports is optimally designed. I return to this subject in Chapter 4.

3.7. Summary and discussion

In this chapter I have investigated opposing conceptions of the central LAAP task and the associated forensic considerations. I have shown that there are two broad versions of the LAAP question, termed for shorthand's sake 'sociolinguistic' and 'forensic'. The former is adopted by counter-analysts and the latter by primary-phase LAAP agencies. I have established that LAAP occurs in a fundamentally forensic context and, further, that the practitioner's first duty is to assist the decision-maker in establishing the truth of the applicant's claim. I have taken the view that the LAAP question, and the presentation of conclusions reached in response, must be formulated with this overriding duty in mind. I have demonstrated that the following are inconsistent with sound forensic practice:

- giving undue weight to inferred episodes of secondary socialisation rather than attempting to discern the applicant's substratal linguistic variety;

- taking the applicant's story as *a priori* truthful, instead of acknowledging that it is the very question at issue;

- insisting on the possession of academic credentials of vague and dubious value to the task at hand;

- expressing freely-worded conclusions;

- during the verification phase of LAAP, considering the linguistic evidence against a single hypothesis, as opposed to two hypotheses in turn, thereby rendering impossible the generation of a likelihood ratio.

I have noted that most of these forensically unsound practices are observable in counter-analyses, though the last is likewise present in reports produced by Verified, a primary-phase agency. What, then, can be done to improve the forensic fitness of LAAP?

The obvious solution would be for practitioners to bring their practices into closer alignment with established forensic procedure and opinion. However, it is plain that primary-phase agencies, on one side, and counter-analysts, on the other, are thoroughly entrenched in mutually hostile positions. Over the course of more than 25 years the discourse on LAAP has

been, for all its spirit and intensity, overwhelmingly static: little change is detectable, in either rhetoric or approach, on either side. There appears to be correspondingly little hope that practitioners might modify their existing methods—all the more so because they have been sanctioned, in spite of their forensic shortcomings, by the efficacy with which they persuade decision-makers in one direction or the other. In this sense, they are already 'fit for (forensic) purpose'.

However, there is reason to believe that at least some practitioners would be amenable to the adoption of tests augmenting, but not modifying, existing practice. In the proposed perception test, the asylum applicant would not even be required to speak, while the production test would demand only that the applicant accurately repeat audio-recorded language samples. The designs of the tests forestall the manifestation in speech of secondary-socialisation phenomena such as code-switching or accommodation. Thus the tests acknowledge and bypass longstanding critique of the potentially distorted nature of speech data emerging from one-shot LAAP interviews, and of the role of NSNL analysts in their evaluation. In a general sense, the envisaged tests would also enhance the general forensic rigour and transparency of LAAP by their potential to (in)validate analysts' assessment of the primary, interview-derived, LAAP data.

There are three further ways that the tests might assist practitioners in meeting their fundamental duty to decision-makers. First, there is the issue of the burden of proof in asylum cases. As specified in Chapter 2, Cambier-Langeveld & van den Boogert (2008) vouch that shifting the burden of proof away from LAAP practitioners and onto the applicant is entirely in line with Dutch asylum law. In substatively identical terms, the UNHCR Handbook (2019, p. 49) states of asylum cases:

"It is a general legal principle that the burden of proof lies on the person submitting a claim."

In accordance with the spirit of the UNHCR's declaration, the tests envisaged would allow asylum applicants to assume a greater share of the burden of proof in LAAP. In this sense, administration of the tests would be forensically sound, and any country which is signatory to the relevant UN conventions could have no reason to reject them.Neither, given the general

acknowledgement that LAAP is situated in a forensic context, should the adoption of such tests provoke objections from either side of the LAAP debate.

Second, in offering an additional layer of validation to existing procedure, uptake of auxiliary tests might moderate the risk of false acceptances associated with single-hypothesis testing. As suggested in Section 3.5, this may be a partial explanation for the large proportion of claims confirmed by Verified, as well as by counter-analysts (but see also the same subsection for evidence to the contrary, apparent in LINGUA's statistics). Supplementary tests may assist practitioners in clarifying whether their rates of acceptances are indeed artificially high or low. Either might be inferred from mismatches between results of an original, confirmatory analysis and the supplementary test(s). The eventual result of this type of retrospection might be that practioners institute the formal weighing of evidence against two hypotheses,, thereby enhancing the forensic fitness of LAAP. In any case, whether or not their administration were to result in changes to practice, supplementary tests would have considerable value to LAAP generally as a mechanism for enhanced cross-checking of conclusions.

The third and final manner in which adoption of supplementary tests might improve LAAP practice has to do with the widely accepted view that native speakers and linguists perceive language differently. In section 3.6 I expressed the view that NSNL judgements appear to be somewhat retrofitted to Verified's reports, the format of which—demanding as it does analysis at the level of the segment—may be poorly suited to accurately reflecting their perceptions. Here, again, supplementary perception tests could enhance LAAP's forensic utility, in the sense that they would present to the decison maker the unmediated perceptions of NSNLs—albeit those of the applicant rather than the analyst. The debate in the LAAP-related literature over the nature of NSNL perceptions is the subject matter of Chapter 4.

Chapter 4: NSNL perceptions

4.1. Overview

In this chapter I first evaluate persistent claims that NSNLs are incompetent to act as analysts in LAAP because they harbour a misleading suite of linguistic perceptions, imputed synonomously to 'folk views/knowledge' or an 'ideology of homogeneism' (cf. e.g. Eades & Arends 2004, Eades 2005, Fraser 2009, Patrick 2010; see below, section 4.3.3., for definitions of these terms). I trace the basis of these assertions to a combination of discourse-analytical, perceptual dialectological and variationist sociolinguistic theory. I argue that this theoretical foundation is flawed by the limitations and inapplicability of the source material.

I next submit a range of counter-arguments derived from an alternative reading of insights from variationist sociolinguistics, as well as material from sociocultural evolution, social psychology and anti-reductionist philosophy. I conclude on this footing that NSNLs as a whole do not subscribe to an intellectually crippling 'ideology of homogeneism'. Instead, much of the literature reviewed explains and demonstrates their likely capabilities as judges of language in LAAP.

I end this chapter by proposing three principles of native speaker perception, which will inform the hypotheses tested in the experiments described later in this thesis. In short, I propose that, possibly owing to evolutionary factors, native speakers have tacit (i.e. non-explicit) knowledge of their own language variety which is mostly resistant to description by the reductive techniques availed by formal education in linguistics.

4.2. A note on terminology

I now arrive at a brief terminological halt. In previous chapters I have referred on many occasions to the 'competence' of NSNLs. In the following two chapters, however, I turn to consider assertions in the LAAP literature—and beyond—as to the nature of the three levels

comprising NSNL competence: perception, judgement and confidence (see e.g. Bargh & Chartrand 1999, pp. 465-7, for a similar hierarchical model).

For present purposes, it is enough to define these three levels as follows. Perception is the primary cognitive act. It is often discussed in theoretical terms but can be apprehended empirically in the form, *inter alia*, of judgements of acoustic stimuli, which may be accurate or inaccurate. In turn, confidence may be expressed as to the accuracy of judgements.

In this chapter, I deal mainly (but not only) with a selection of literature on NSNL perceptions. In Chapter 5 I look at NSNL judgements and expressions of confidence in them. Because of the theoretical slant of much of the literature on perception, the focus of this chapter is largely (but not entirely) theoretical. In accordance with the conceptualisation of judgements as the observable output of perception—and confidence, in turn, of judgements—the focus of Chapter 5 is mainly empirical.

4.3. An 'ideology of homogeneism'?

4.3.1. Background

The first (English-language) literature critical of LAAP, as carried out at the time on behalf of the Swedish immigration authority, made its appearance in 1998, in an online bulletin called FORTRESS EUROPE? - Circular Letter 53. This document is now unavailable online but is avaliable as a digital copy saved by me in April 2018. The FECL, quoting academic linguists and an immigration lawyer, criticised the fact that LAAP in Sweden was then being performed by staff of the agency Eqvator, which had grown out of the immigration authority's language section.

According to the FECL, Eqvator's staff lacked academic linguistic qualifications; their sole authority derived from the fact that they were, in broad terms at least, native speakers of the various languages on which Eqvator performed analyses. It was alleged, however, that some of

Eqvator's language analysts did not have origins in the same country as the asylum seekers whose language they were engaged to assess.

The fact that Eqvator refused to release details of its staff's qualifications was criticised in the FECL on the grounds that it was impossible to judge whether and to what extent NSNLs and/or linguists were involved in composing Eqvator's reports, and so to judge the merit of their conclusions. The assumption, however, appears to have been that NSNLs lacking any kind of linguistic or forensic training were responsible for the entirety of Eqvator's LAAP.

Concerns were also expressed in the FECL about whether Eqvator was submitting its reports to cross-checking by experts outside the company. No evaluation, however, was made of the accuracy of the conclusions of the reports themselves—with the exception of a single report, which was alleged to be of a very low standard, on an applicant who claimed to be from Afghanistan.

Owing to their (putative) production by non-linguists, general doubt was cast on the reliability, consistency and scientific rigour of Eqvator's reports. Related misgivings centred on the alleged failure to take into account—especially in the context of Africa and Afghanistan—sociolinguistic complexity and the frequent occurrence of cross-border language varieties in the repertoire of asylum applicants.

Dr Ruth Schmidt, then of the University of Oslo, is quoted in the FECL making the recommendation that "scientifically acceptable language tests must be carried out by trained linguists, specialising in the language concerned." Schmidt's recommendation expresses an unqualified preference for the competence of linguists over that of NSNLs. The substance of this recommendation, as well as the rest of the criticisms made in the FECL, have been reiterated and amplified multiple times by other parties in the succeeding quarter-century. Criticisms have since extended beyond the practices of the now defunct Eqvator to those of currently operating agencies which use the NSNL + linguist team approach: the OCILA, Verified and Sprakab.

Reath (2004) presents a thorough and fairly balanced attempt at a summary of the then-currently known practices and issues, with reference to Eqvator, Sprakab, the Belgian government's LAAP agency and cases from Australia and New Zealand. In general, though, the thread of continuity between the 1998 FECL and more recent critical literature on LAAP is notable in the rather minimal evidential foundation of most criticisms made. This is partly attributable to the fact that prior to 2010 no LAAP practitioner had yet published any details of their approach; neither, in most cases, were the qualifications known of those responsible for the conclusions reached. This appears to have prompted some scholars to resort to statements of contestable first principles (e.g. 'only linguists should do linguistic analysis') and highly questionable evidence. Many also sought authority for their criticisms in theories of doubtful validity.

4.3.2. The critique of NSNLs in LAAP: theory and evidence

The most durable of these theories is the notion of 'ideology of homogeneism', derived proximally from a 1998 publication by Blommaert & Verschueren and ultimately from the sociolinguist Dell Hymes's critique of the early nationalist ideas of the 18th century German Enlightenment philosopher Johann Gottfried von Herder (Eades 2009, p. 35). In the LAAP literature, the concept of an insidious 'ideology of homogeneism', supposedly at work only among non-linguists, has come to be more or less synonymous with 'folk views/knowledge'—a separate but related postulate current in the field of perceptual dialectology (Preston 2019).

Eades, Fraser, Siegel, McNamara & Baker (2003), Eades & Arends (2004), Eades (2005, 2009) and Patrick (2010) each criticise, in similar terms to the FECL, the practice of LAAP for its deployment of NSNL competence. They catalogue a range of LAAP cases dating from 1999 to 2009 involving asylum applicants claiming origins in Afghanistan (Eades et al. 2003; Eades 2005, 2009), Sierra Leone (Eades & Arends 2004) and Somalia (Patrick 2010). It is claimed—though in no instance known for certain—that in each of these cases NSNLs were wholly responsible for the conclusions of the resulting report, in which the authors locate serious shortcomings. These are alleged to consist mainly of citations in the reports of supposedly aberrant pronunciations

and erroneous claims that certain languages are not spoken in certain countries; see Eades (2005) and Eades et al. (2003) for a number of *prima facie* egregious cases. The authors each invoke NSNLs' 'ideology of homogeneism'/'folk views' as an explanation for the (presumed) errors in the reports and, on the basis of this evidence, urge that NSNLs not be decisively involved in LAAP. However, as Wilson (2016) asserts, the correctness of the reports' conclusions remains unknown; therefore it is also unknown whether those responsible for the reports were incorrect in their interpretation of the features cited, regardless of the frequently inexpert manner of their expression.

It is worth noting here that Arends, Eades, Fraser, Siegel, McNamara and Patrick were six of the 19 scholars responsible for the *Guidelines* (LNOG 2004)—though, as Wilson (2016) points out, none of these six signatories (nor most of the other signatories) had direct experience of LAAP at the time the *Guidelines* were written. Although the *Guidelines* wholly avoid mention of an 'ideology of homogeneism' or 'folk views', points 3 and 7 of the document, in common with the individual contributions of these six authors, also effectively recommend the exclusion of non-linguists from either making "judgements" or providing "expertise" in LAAP cases.

It is also the case, as mentioned in Chapter 1 and detailed in Chapter 2, that NSNL analysts from all three 'team-approach' LAAP agencies are to varying degrees trained—both before and after commencing work on real-life cases—in (socio)linguistic and forensic precepts germane to the field. The extent to which it is legitimate to impute an ideology of homogeneism/folk views to trained professionals, even if they are not formally qualified linguists, is highly questionable.

4.3.3. Definitions

Leaving aside the equivocal nature and scanty quantity of the evidence adduced in the above critiques, the implictly exclusionary nature of the *Guidelines*, and the dubiousness of imputing 'folk views' to NSNL LAAP analysts who are in fact trained in (socio)linguistics, the natural first step in interrogating the validity of claims invoking theories of 'folk views' and 'ideologies of

homogeneism' is simply to define the two terms. What do the authors concerned mean by their use?

According to Eades et al. (2003), folk views are exemplified in the conviction among non-linguists that they are capable of determining a speaker's place of origin by the words or pronunciations they use, as well as the belief that language, ethnicity and nationality are co-determinative categories. For his part, Patrick (2010, p. 77) avers that folk views are,

> "...grounded in prescriptive biases (especially educated speakers), and based upon underlying constructs significantly at odds with the facts described by linguistics."

The associated 'ideology of homogeneism' is evident, according to Eades (2005, p. 511), in the commonly-held belief that societies and the people who comprise them are in essence monolingual and monocultural. Moreover, this ideology is said to be "rampant in many societies around the world".

Eades & Arends (2004, p. 180) assert that the ideology of homogeneism renders NSNLs ignorant of "... the realities of language variation and bilingual speech ...". Similarly, Fraser (2009, p. 114) alleges that it is "well known" that non-linguists, because of their folk knowledge, are "ignorant of many aspects of language" as well as ignorant of "their own ignorance". In the judgement of these commentators, folk knowledge and ignorance thus apparently amount to the same thing.

The "scientific" judgements of linguists (Eades et al., 2003, p. 186) are said to be free of the spurious assumptions engendered by the ideology of homogeneism. Eades et al. argue that linguists alone are aware of phenomena such as language spread, language change and above- or below-consciousness pronunciation differences. Apparently, too, only linguists know that language, ethnicity and nationality are *not* co-determinative categories and that a speaker's place of origin cannot *always* be decided solely on the basis of the words or pronunciations they use (Eades et al. 2003, pp. 183-186). According to Eades, NSNLs' lack of knowledge of the International Phonetic Alphabet renders them further incapable of reporting their conclusions reliably (2005, pp. 509-10).

In essence, those scholars who posit theories of 'folk views'/'folk beliefs' and an '(ideology of) homogeneism' to characterise NSNLs' supposedly false consciousness appear to be referring to what the structuralist linguist Leonard Bloomfield called 'stankos': "ignorant or stupid remarks about language" by non-linguists (McGregor 2001). The theory of homogeneism thus characterised, the second step in an examination of the validity of the critique of NSNLs in LAAP is to examine the congruence of the theory with the work of the (claimed) originators of its foundational constructs.

4.3.4. Origins of the theory: Blommaert & Verschueren, Preston, Labov

Homogeneism: Blommaert & Verschueren

The first appearance of the term 'homogeneism' I have been able to detect is in Blommaert & Verschueren (1991, p. 528), as an apparent neologism coined by the authors. It occurs again in Blommaert & Verschueren (1992, p. 362), where it is defined as follows.

> "...the ideal model of society is mono-lingual, mono-ethnic, mono-religious, mono-ideological. Nationalism, interpreted as the struggle to keep groups as 'pure' and homogeneous as possible, is considered to be a positive attitude within the dogma of homogeneism. Pluri-ethnic or pluri-lingual societies are seen as problem-prone, because they require forms of state organization that run counter to the 'natural' characteristics of groupings of people."

Both Blommaert and Verschueren are linguists: Verschueren specialises in pragmatics and Blommaert in African languages and sociolinguistics, particularly critical discourse analysis. Both of the works cited above (Blommaert & Verschueren 1991, 1992) seek to interpret conceptions of ethnic and national identity in the media of various European states from a discourse-analytical perspective—a theoretical vantage point with an open agenda of social and political reform (Johnson & McLean 2020).

Blommaert & Verschueren manage to find copious evidence of 'homogeneism' in the material they survey, most of it dating from the late 1980s and early 1990s. They do not, however, discuss the views of non-linguists as a whole; their focus is largely upon metapolitical commentary in newspapers. Notably, too, their critique extends only as far as European states. Plainly, this does not amount to direct evidence of the linguistic opinions held by non-linguists in Europe, let alone by non-Westerners, who constitute by far the majority of the world's peoples. Nonetheless, Blommaert & Verschueren's concept of homogeneism, if not its focus as originally conceived, has been widely adopted in the LAAP literature.

Perceptual dialectology/folk views: Preston

As we have seen, the concept of 'folk views/beliefs' is drawn from perceptual dialectology, especially the work of Dennis Preston. Preston is frequently cited in the scholarly critique of NSNLs in LAAP (see e.g. Fraser 2009, Patrick 2010), yet the relevance of such citations is rarely made clear.

Patrick (2010), for instance, cites two authorities in support of the assertion that NSNLs harbour 'folk views'. The first is referenced as 'Preston & Long 2002', which is an anthology, *The Handbook of Speech Perception*. Since Patrick does not cite any particular article or page number in this 700-page work, it is not possible to say exactly which evidence he relies upon. The other authority is Preston (2002). This chapter in *The Handbook of Language Variation and Change* is mainly devoted to the influence of linguistic value judgements ('attitudes') on dialect perception in the United States. For example, Preston summarises a study of his own in which participants from Michigan rated the accents associated with various parts of the United States for affective constructs such as 'pleasantness' and 'correctness' (2002, pp. 54-61). Another of the studies Preston cites, by Niedzielski (1999), is on Michigan informants' ability/willingness to identify a particular vowel with their own variety of English. Participants were presented with an answer sheet with either 'Michigan' or 'Canadian' written on it. Preston attributes their decision to identify the vowel with either their own variety or Canadian English largely to priming—i.e. which answer sheet they received (Preston 2002, pp. 47-8).

It is difficult to tell what such findings are supposed to mean for LAAP. On the basis of the first of the above experiments, Patrick may believe that non-linguists could misattribute accent features which they perceive negatively to members of a linguistic community other than their own or that, conversely, they may misidentify positively-evaluated features as diagnostic of one of their own members. The consequence for LAAP would presumably be that authentic claimants may be incorrectly excluded due to misguidedly negative (or positive) 'folk' evaluations of the variety they speak. It is noteworthy, however, that Niedzielski's experiment was conducted under conditions of such heavy manipulation and priming that it is difficult to see what relevance it may have to LAAP, from which such conditions are completely absent (for further commentary on this study, see Chapter 5).

Other possibly relevant studies are discussed in Preston (2019). This recent contribution cites much the same range of empirical studies as does Preston (2002). Helpfully, it is divided into two sections, which summarise research demonstrating, respectively, 'good' and 'bad' folk judgements. In the first part, Preston summarises the results of multiple studies over a 15-year period in which listeners were able to correctly assign language samples to one of three categories—Standard American, African-American and Chicano English—at a rate of approximately 80 percent (2019, pp. 133-4). He also cites a 1996 study of his own in which NSNLs were asked to identify the location in the United States of nine speakers and concludes, "We must … be impressed by nonlinguists' abilities to detect linguistic differences" (2019, p. 139).

As an example of a 'bad' folk view, Preston (2019, pp. 140-42) cites a study conducted on language judgements by NSNLs living in the midst of one of Europe's best known and densest dialect continua: the Dutch-German border. The task involved respondents drawing arrows on a map to indicate where people speak the same on both sides of the frontier. Preston reports that few respondents drew arrows crossing the border. He interprets this to mean that "...the folk are empowered by many factors, not least the ideological notion of "one people, one language", itself enhanced by the notion "one nation, one people" (2019, p. 150). However, Preston also reproduces IPA transcriptions of a sentence as it is spoken in two of the locations

in question, one on either side of the border. These show that the dialects do indeed differ from each other, albeit in a single, relatively subtle feature (perhaps tellingly, no indication is made as to prosody). Were the respondents wrong in this instance? Or did they make accurate judgements based on their perception of the sorts of fine-grained detail to which Nolan (2012) and Klein (1988) allude?

In the LAAP context, it is certainly conceivable that misjudgements may occur when NSNLs attempt to determine on which side of a national border a person originates. Yet it is difficult to see how this applies exclusively to the performance of LAAP by NSNLs, since in LAAP-like tasks the judgements of even highly trained phoneticians are demonstrably imperfect (see e.g. Wilson 2009; Muhammad 2021).

One especially notable (and LAAP-like) real-world speaker-profiling case on which linguists had great difficulty in agreeing was that of 'Lord Buckingham' (French, Foulkes & Wilson 2019, pp. 98-9). Here, a large number of experienced academic phoneticians (as well as NSNLs) from four separate continents was consulted in attempting to identify 'Buckingham's' likely substratal linguistic background. Descriptions of language varieties local to the four continents concerned are extremely thorough. Yet only with extreme difficulty and much contention was the decision eventually reached that he was likely to be a speaker of North American English—a conclusion that was subsequently confirmed by other evidence.

It is noteworthy that Preston never compares the judgements of NSNLS with those of linguists on the same task. Neither does he clarify how 'folk' *evaluations* of a particular language variety might influence an NSNLs judgement as to the likely provenance of one of its (claimed) speakers. An NSNL may, for example, think of his/her own variety as 'correct' and a related and neighbouring variety as 'incorrect', or that one is a dialect of the other. But this does not mean that he/she would deny the geographical proximity of each variety to the other or that a speaker of a neighbouring dialect is (or may be) a co-citizen of the same nation-state.

Overall, the kinds of studies Preston discusses are mostly irrelevant to LAAP. Much is made of how linguists and NSNLs allegedly differ in their judgements—yet this is unknown, since the cited experiments feature only NSNL participants. Moreover, little attention is given to an issue

of central importance to LAAP: the *accuracy* of NSNL judgements. Taking steps towards the resoltuion of this matter is central to the experimental work of this thesis (Chapters 7 to 11). It is also the subject of the remainder of this chapter, as well as a large part of Chapter 5.

Variationism: Labov

A further theoretical influence on the critique of NSNLs in LAAP appears to come from mainstream ('Labovian' or 'variationist') sociolinguistics (see, e.g., frequent references to Labov in Patrick 2010). However,as is the case with his references to Preston's work, Patrick never makes clear *how* variationism buttresses the theory that NSNLs are captive to an ideology of homogeneism. In any case, the extent to which it is valid to draw upon Labov's work in this connection is questionable.

Essentially, variationist sociolinguistics seeks to describe individuals' actual use of language in the context of the speech communities to which they belong. This natural language use is governed by variable forms arranged in "internally structured heterogeneity" (Weinreich, Labov & Herzog 1968, p. 101). It is axiomatic among variationists that language use varies across parameters beyond those conventionally described in classical dialectological surveys. Differences in the frequency of individual variants (e.g. rhotic vs. non-rhotic pronunciations of certain words) often occurs on a spectrum rather than in conformity with clear-cut isoglossic boundaries, and their use may be statistically correlated with the speaker's social class, sex, age group or ethnicity, or by the speech style adopted in a particular context (Foulkes & Hughes in press). This is what is meant by "internally structured heterogeneity", which variationists consider a universal fact of natural language.

Patrick (2010) seems to accept the universal applicability of these variationist precepts, awareness of which (at least in their explicit form) is solely available to (socio)linguists, regardless of whether they are native speakers of a given variety. But what he does not appear to acknowledge is the *particularistic* implications of variationist theory: speech communities are particular social formations whose members—and no-one else—ultimately control and systematically comprehend the particular (albeit socially and temporally dynamic) occurrence and arrangement of features in the variety they speak. This is so regardless of heterogeneity

among individual speech community members (e.g. their age or sex) and the greater or lesser associated statistical likelihood that they will deploy certain variants in a given context. The point is that the available range of variants is always *structured*—i.e. constrained by the linguistic norms of a particular speech community.

Applying this reading of variationist theory to LAAP, it would seem to follow that speakers attempting to imitate a variety other than their own cannot competently deploy 'indexical' (i.e. socially or regionally marked) variants used by authentic members of the target speech community. Moreover, if one of the characteristics of natural language is 'structured heterogeneity', and if native speakers of a given variety are those who are fully conversant in it, it follows that they are most capable of recognising the *unstructured* heterogeneity characteristic of language imitation—albeit perhaps not with 100 per cent accuracy.

Seen from this perspective, the knowledge characteristic of native speakers is not governed by an 'ideology of homogeneism' at all. Rather, it consists in part of the ability to distinguish, in the act of speaking and by ear, between native and non-native types of variability: structured versus unstructured heterogeneity; authentically native speech versus imitation. This ability to distinguish is crucial for the validity of the existing practice of LAAP, the team approach in particular. It also implies that authentic native speakers of a particular language variety would be able to perform competently in the production test initially outlined in Chapter 1 of this thesis.

4.4. Further theoretical and empirical work countering the 'ideology of homogeneism'

I have traced the theory of homogeneism, which underpins much of the critique of NSNLs in LAAP, to three sources: Blommaert & Verschueren's account of the influence of normative ideologies in European states; Preston's work on folk views; and an incomplete reading of the variationist axiom that language use is internally heterogeneous. On this rather superficial theoretical foundation, a case has been constructed by various scholars that NSNLs, absent

training and supervision by linguists, are incapable of reliable judgements of who is and who is not a member of their own speech community.

In the following subsections I present a range of theoretical and empirical work which also runs counter to the theory of homogeneism. This material is drawn from diverse fields: sociolinguistics, the practice of LAAP itself, sociocultural evolution, social psychology and anti-reductionist philosophy. In examining it I arrive at three alternative theoretical propositions as to the nature of native speaker perception. These propositions, along with empirical findings reviewed in Chapter 5, undergird the hypotheses tested in the experimental work of this thesis.

4.4.1. The sociolinguistic context: Western vs. non-Western societies

Khattab & Foulkes (in press) point out that most sociolinguistic studies (notably, such as Labov's) have been conducted on monolingual Western subjects, in Western societies. As I have shown, the same is true of Blommaert & Verschueren's theory of 'homogeneism', which relates solely to Europe. Khattab & Foulkes contend that the existing state of sociolinguistic theory may be of limited applicability to non-Westerners, whose lifeways and patterns of language use are in many cases quite unlike those typical of Westerners. This argument is particularly relevant to LAAP, since most asylum applicants originate in non-Western countries.

The Arab world and the West

Khattab & Foulkes discuss the Arab world, which is well-known for the prevalence of diglossia—or, perhaps more accurately, multiglossia. This sociolinguistic situation is unlike that of most English-speaking countries. Particular language varieties might be spoken only among members of certain ethno-religious groups (e.g. Greek Orthodox Christians, Druze, Jews and even Sunnite and Shi'ite Muslims). A level above are regional varieties (e.g. Horani Arabic in Syria). National standard varieties (Syrian, Egyptian etc.) are often used in communication beyond the regional level, or interpolated piecemeal into the speaker's regional dialect. In addition, a 'levelled' variety, known as Modern Standard Arabic, is often spoken in the media, and 'borrowed from'

by many vernacular varieties. Classical Arabic, meanwhile, is restricted to religious contexts, largely in recitations from the Quran or other religious texts.

A further difference between Arabic speakers and most Westerners is that the former may be less maritally, socially and occupationally mobile—but more geographically mobile—than the Western norm. This is likely to be the case especially among (semi-)settled rural people who maintain tribal and/or strong ethno-religious affiliations and, to some extent, traditional occupations, such as the herding of cattle, camels, sheep or goats (Dukhan 2014; Jamestown Foundation 2013). However, at least where marriage patterns are concerned, low mobility is not confined to rural areas. In their survey of marital trends in Syria, for example, Othman & Saadat (2009) show that the frequency of consanguinity between partners is c. 30% in urban areas and c. 40% in rural districts (c. 36% overall), the most common type of consanguineous union being between first cousins (c. 21% of the total). Bener & Mohammad (2017, p. 316) review even higher rates of consanguineous marriage in other parts of the Arabic-speaking world: 51 to 58% in Jordan; 54% in Kuwait; 49% in Tunisia; 33% in Morocco; 58% in Saudi Arabia; 50% in the United Arab Emirates; 52% in Qatar; 40 to 47% in Yemen; 50% in Oman; and 68% in the Egyptian city of Alexandria.

Individuals entering into consanguineous marriages (a practice associated with belonging to a tribe or 'segmentary lineage') are more socially static than the Western norm. As Milroy (1980) would put it, they are situated within relatively 'closed social networks'. Yet they may be more geographically mobile—e.g. some Bedouin—than is typical in the West, albeit often within bounds circumscribed by tradition and, more latterly, by political factors. Syrians with tribal affiliations, for instance, often maintain close transnational connections to fellow tribe members in certain parts of the Levant, the Arabian Peninsula and Iraq; however, owing to the erection of national borders across the Arab world in the 20th century, they are less able than previously to move freely, *en masse*, between contemporary states (Dukhan 2014; Jamestown Foundation 2012). In subsection 4.4.3. I return to consider briefly the possible implications of consanguineous marriage within 'segmentary lineages' for modes of linguistic perception and reasoning.

Non-Anglophone and Anglophone countries

The LAAP-focused literature is replete with discussions of the generally multilingual nature of non-Western and/or non-Anglophone societies—even by proponents of the theory of homogeneism. Muysken (2010) stresses the prevalence of multidialectalism in West Africa, in Morocco and in European countries such as Belgium and Germany. Blommaert (2001) and Maryns (2010) also chart multilingualism in Africa. Nurse (2019) treats the complexity of language use among the Bajuni of East Africa. On a world scale, Fraser (2009, p. 128) contends that

"Asylum seekers ... often come from regions characterised by complex multilingualism, creolisation or diglossia, which have had little attention from linguistic science."

Maryns & Blommaert (2001, p. 61), meanwhile, suggest:

"Perhaps one should take multilingualism and mixedness, rather than monolingualism and linguistic purity, as defaults in language use."

Fitzmaurice (2019, p. 206),with a focus on Zimbabwe, similarly argues:

"The traditional Eurocentric assumption that ethnicity is synonymous with language identity is increasingly challenged and now shown to be highly problematic for most communities in present-day Africa."

The *Guidelines* (LNOG 2004, p. 264), too, point out that:

"Sociolinguistic research shows that multilingualism is the norm in many societies throughout the world.

To these assertions may be added a simple *prima facie* observation: it is exceedingly unlikely that the world's population of nearly 8 billion individuals (excluding linguists) all perceive language in the same errant way implied by an ideology of homogeneism, whatever the reach of such an ideology in the present day. It is equally unlikely that different *groups*—Moroccans and Belgians, Zimbabweans and Bajunis—all share the same set of perceptions.

It is especially difficult to believe that people from places in which multilingualism is the norm would subscribe to an ideology of homogeneism, "that sees societies as 'normally' monolingual and a member of one culture" (Eades 2005, p. 511). Moreover, acknowledging that there is very limited data on some geographical areas and communities of vital interest to LAAP, one wonders how it would be possible to evaluate the extent to which their inhabitants harbour 'homogeneistic' language ideologies. All this together sits uneasily with the attribution of an ideology of homogeneism to non-Anglophone, non-monolingual and/or non-Western NSNLs.

With reference to the revealed linguistic perceptions of Arabic speakers, Hoskin (2018, pp. 32-3) catalogues evidence that Arabic-speaking NSNLs working as analysts for Verified, far from being victims of an ideology of homogeneism, are adept at differentiating internally structured from unstructured heterogeneity (i.e. authentic from inauthentic language mixing) in the speech of asylum applicants. The first three examples were associated with reports confirming the applicant's account:

"He speaks [X] dialect but his dialect is mixed between [X] and [Y] dialect. He is imitating [Y] dialect all the time. And hiding [X dialect]. Maybe he is hiding is [X] dialect because it is not desirable by the people in [an area of Syria]. People there used to imitate [Y] dialect because their [X] dialect is subject to discrimination."

"...the person doesn't speak pure [X] dialect as people of [X] speak it. He speaks ... [X] mixed with [Y] dialect... He is not trying to imitate any other dialects."

"He tends to imitate [X] dialect; this is usual case in [Y] speakers' areas and [city X]. He is not trying to hide his dialect."

The final three comments were made in connection with a non-confirmatory conclusion. In the case at issue, Verifed's hypothesis was that the applicant's Arabic was consistent with the variety spoken in a Syrian city near the Turkish border. In his remarks, the NSNL analyst shows that he is explicitly aware that the variety at issue is also spoken in adjacent parts of Turkey. He is, in other words, cognisant that ethnicity, language and nationality are not mutually determinative categories.

> "Nevertheless, one cannot neglect the existing of this dialect in some Turkish cities, which are located along the Turkish-Syrian border..."

The analyst exercises further caution (and an awareness of naturally-occurring language mixing) in pointing out that the applicant's account of a period of residence in Lebanon is consistent with certain patterns observed in his speech:

> "... his dialect was, to some extent, influenced by the Lebanese dialect. He mentioned that he lived in Lebanon for four years."

However, the analyst still finds himself obliged to recommend a '-1' (qualified negative) result for the hypothesis that the speaker's claim of residence in Syria is supported (bolding in original).

> "The applicant's way of speaking, his **intonation** at the end of each answer, **repeating** and **unintelligible** words led me to a conclusion that the applicant's dialect is more likely deviating than consistent with the tested hypothesis."

Note that the NSNL emphasises the applicant's "intonation" in justifying his (cautiously rated) non-confirmatory conclusion. This is precisely the sort of suprasegmental information that Nolan (2012) theorises is characteristic of native speaker language perception but remains inadequately captured in conventional linguistic descriptions.

This evidence shows that (some) NSNLs are able to separate 'natural' from 'deceptive' language mixing—and to give their assessments proper weight in arriving at both positive and negative conclusions in LAAP. Assertions as to the universal and benighting influence of an ideology of homogeneism are thus, at the very least, demonstrably overgeneralised.

It is worth briefly considering whether homogeneism is indeed more prevalent in the relatively monolingual Anglophone countries than elsewhere. Milroy & Milroy's book *Authority in Language* (1985) deals at length withthe tradition of prescriptivism and standardisation in the English-speaking world, from both diachronic and synchronic perspectives. The authors discuss multiple instances of pressures shaping English towards 'official' norms (i.e. essentially

'homogeneism') while also according space to the endurance—and in some cases the emergence, for instance in Singapore—of multiple non-standard or dialectal varieties of English. Milroy & Milroy characterise these non-standard varieties as assertive, robust and commanding strong allegiance from their speakers: the opposite of what one would expect among people in thrall to the normative assumptions of an ideology of homogeneism.

In this connection it is instructive again to recall the successful use of NSNL informants in the Yorkshire Ripper hoax case (Foulkes et al. 2019). This suggests that British (or, more specifically, English) NSNLs remain highly alert to subtle—and, as it turned out, very real—differences in speech in extremely narrowly-defined geographical areas. It would seem, then, that the alleged Western prescriptive tendency, in favour always of official norms, does not necessarily translate to an all-pervasive 'homogeneistic' ideology, even in the Anglosphere.

4.4.2. Sociocultural evolution: the ultimate origin of language perceptions?

A growing body of theoretical and empirical work in sociocultural evolution suggests that the ability to distinguish authentic from inauthentic fellow speakers of a particular language variety may be ultimately evolutionary in origin. This evidence appears to reinforce that adduced previously, from sociolinguistics, that native speakers of particular language varieties are uniquely able to recognise fellow speakers. The basic propositions of the relevant strand of sociocultural evolutionary thought are that language variation emerged as a signal of group membership, and that the ability to perceive shared patterns of variation assisted individuals in making decisions as to the trustworthiness of potential cooperators (Nettle & Dunbar 1997; Richerson & Boyd 2010; Haidt 2012; Nolan 2012; Cohen 2012).

Sociocultural evolution and language: theoretical accounts

For the majority of human (pre)history, ethnolinguistic groups are thought—and, via genetic research, increasingly known (Reich 2018)—to have consisted of relatively closely-related individuals. With environmental and technological change, however, human groups experienced a generally accelerating pattern of geographical, genetic and cultural fission (Foley

& Mirazón Lahr 2011, p. 1080). Under the resulting conditions of ever-growing social complexity and concomitantly decreased personal acquaintance, shared packages of language variation may have functioned as a proxy for genetic relatedness among speakers of specific language varieties (Foley & Mirazón Lahr 2011, pp. 1086-7). It was on this footing that trusting reciprocal relations (or preferential cooperation) could most securely to be established and maintained in pre-modern societies (Nettle & Dunbar 1997; Richerson & Boyd 2010).

Distinctive and systematic linguistic variation has a vital advantage over other social markers in that it is especially difficult to fake; therefore, the ability to detect a shared system of linguistic variants, or even single shibboleths, is especially useful in identifying fellow 'ingroup' members (Nettle & Dunbar 1997; Cohen 2012). Exercise of this 'gatekeeping' ability tends to hinder infiltration by members of 'outgroups', who are more likely to 'free-ride' on ingroup resources or to propagate lies inimical to ingroup interests (Richerson & Boyd 2010). Groups which developed systematic linguistic variations of their own—and which were able to distinguish variations typical of the ingroup from those typical of outgroups—were thus at a considerable evolutionary fitness advantage relative to those less capable in this domain (Nettle & Dunbar 1997).

The contribution of individual- and group-level, and of cultural and genetic, factors in the evolution of language variation is hotly disputed (cf. Pinker 2012). However, a simplified yet plausible 'multi-level' synthesis, incorporating all four factors, might be as follows (cf. Haidt 2012). Greater willingness on the part of individuals to cooperate preferentially with more trustworthy ingroup members, identified by their use of shared linguistic features, resulted in a greater number of mutually-profitable interactions. The relative material prosperity derived from these interactions conferred a fitness dividend—i.e. the production of relatively large number of offspring—to the individuals concerned. These individual-level processes were in turn highly consequential for selection at the group level. Transmission among individuals of genes coded for 'groupish' cooperation, and/or cultural traits which engender 'groupish' behaviour, increased groups' propensity to cooperate preferentially with fellow ingroup members. These genes co-evolved with the ability, honed by intensive cultural selection, to

identify shared linguistic variants as a reliable indicator of individual trustworthiness, which through a feedback loop with individual- and group-level selection became more pronounced and prevalent.

According to this account, the tendency of human groups to maintain systematic, and often highly subtle, language variation—and to perceive it as an essential marker of group identity—has deep roots in prehistory. It is not surprising, then, that group-level language variation, and the ability of individuals to detect it, has endured into the present.

Theories of language perception founded on evolutionary premises are not without support among linguists, including forensic phoneticians. Hollien (2002, p. 17) suggests that the ability to distinguish members of the in-group from those of out-groups by their speech alone may be as old as spoken language itself. Nolan (2012, p. 273) conjectures that language variation may have evolved through natural selection as a mechanism to exclude free riders.

Sociocultural evolution and language: empirical evidence

It is readily apparent that multi-level theories emphasising bottom-up, group-oriented evolutionary processes stand in opposition to the basically social constructionist theory of homogeneism, which proposes top-down transmission of ideas (including formal linguistic education) to individuals as the fundamental determinant of language perception. Previously in this chapter I have attempted to demonstrate that the empirical basis for the latter theory is dubious. But what of the merits of the empirical evidence for the various evolutionary accounts?

Nettle & Dunbar (1997) conducted a computer simulation of reciprocal exchange between 100 'organisms' of four different types. These were each programmed with a 'dialect' and varying permutations of exchange/linguistic strategies. In the modelled exchange scenarios, the first type of organism, named COOP, always gives unless its giving has been previously unreciprocated by another organism; the second, CHEAT, never gives; the third, POLYGLOT, gives only to 'speakers' of the same dialect and is programmed to introduce innovative

linguistic variations to other organisms; the fourth, MIMIC, always cheats while imitating the dialect of its exchange partner.

Results showed that, initially, populations of CHEATS were collectively impoverished, while those of COOPs were collectively wealthy. Unsurprisingly, incursions of CHEATS into COOP populations had a marked impoverishing effect. However, as the 'cognitive' ability—i.e. memory span—of COOPs was increased, CHEATS and MIMICS (i.e. 'free riders') were eventually rendered ineffectual or died out entirely. Crucially, stable and distinctive dialects emerged along well-established paths of mutually successful exchange involving COOPs and POLYGLOTs. This is taken by the authors as support for the inference that, in the real world, stable groups speaking shared language varieties emerge from patterns of healthy reciprocity, thus confirming the association made in evolutionary accounts between linguistic gatekeeping and group-level fitness.

Several empirical studies (e.g. Wilson 2009, Hedegard 2015, Muhammad 2021) have located highly accurate performance by NSNLs in LAAP-like accent identification tasks, though none interpret their findings through an evolutionary prism. These experiments are discussed in detail in Chapter 5.

A study that *does* explicitly attempt to investigate evolutionary theories of group patterns in accent identification is by Goodman, Crema, Nolan, Cohen & Foley (2021). Initial results were drawn from 50 participants from the British Isles who were recorded reading sentences, once in their own dialect and once in imitation of another variety. They then listened to 12 recordings, six of imitators and six of authentic speakers. The task was to identify which were imitated and which were genuine. All listeners, regardless of local identity, had a combined 66.7% probability of correctly rejecting imitators and accepting authentic native speakers of their own local variety. The authors interpret this finding as support for evolutionary accounts that emphasise linguistic gatekeeping as a mechanism assuring the evolutionary fitness of groups in the face of potential infiltration by free riders. Preliminary results of a much larger study of c. 1,000 participants engaged in the same tasks indicate support for local listeners' superior ability to

identify imitators of their own variety (70% to 75% probability), with non-locals at 55% to 60% probability on the same stimuli (J. Goodman, pers. comm.)

In its attempt to investigate theories of evolutionarily-rooted group patterns in same/different accent identification, the study by Goodman et al. must be regarded as a pioneering work. One obvious limitation, however, concerns the relatively small difference it reveals in local and non-local listener performance. Other reservations relate to the interpretation of such results: With what degree of security can they be ascribed to temporally-remote evolutionary processes? Can explanations instead be sought in more recent cultural history (e.g. strong public interest in and exposure to regional accent variation in the British Isles)? How do such findings clarify the respective roles of, and relationships between, genetic and cultural change in (pre)history? Overall, though, the theoretical and experimental work reviewed in this section may be interpreted as reinforcing the argument, made out in previous sections on sociolinguistic grounds, that native speakers (including NSNLs) are peculiarly able to perceive the norms defining membership of their own speech community.

4.4.3. Social psychology: evidence for group-level differences in reasoning and perception

In the preceding subsection I examined evidence for the likely influence of evolutionary pressures on the *universal* human perception of language. There is, however, evidence that human *groups* differ in degree (though not in kind) in their dominant modes of perception and reasoning. The corollary—that there may be intergroup differences in language perception—is compatible with observations made in subsection 4.4.1. about society and sociolinguistics in the non-Western world. In further investigating this notion, I now turn to consider social-psychological studies conducted on cognitive styles: holistic vs. analytic reasoning and the related construct of field-independent vs. field-dependent perception, both of which are in turn linked to dependent vs. independent self-perception.

Westerners and non-Westerners: differences in cognitive styles and self-perception

According to Henrich, Heine & Norenzayan (2010), holistic reasoning is essentially the inclination to predict and explain on the foundation of perceived relationships between a focal entity and the context in which it is situated. Analytical reasoning refers to the tendency to perceive entities separately from their background, and to the use of categorical rules to predict and explain them. These modes of reasoning are closely tied to cognitive styles: field dependence on the one hand and field independence on the other.

Zebian & Denny (2001, p. 59) explain the gradient nature of these cognitive styles. They are:

"...used to describe domain-general modes of thinking that emphasize some kinds of processing over others; they are not abilities that some have and others lack."

Henrich et al. (2010, p. 12) make a similar point about the gradience and universal availability of cognitive strategies. They also explain that the holistic/field-dependent and analytical/field-independent cognitive styles are closely associated with differing modes of self-perception, the first individualistic and the second group-oriented.

"Do people conceive of themselves primarily as self-contained individuals, understanding themselves as autonomous agents who consist largely of component parts, such as attitudes, personality traits, and abilities? Or do they conceive of themselves as interpersonal beings intertwined with one another in social webs, with incumbent role-based obligations towards others within those networks? The extent to which people perceive themselves in ways similar to these independent or interdependent poles has significant consequences for a variety of emotions, cognitions, and motivations."

(2010, p. 70)

Henrich et al. (2010, pp. 71-2) show that various populations of Westerners (Australians, Americans, Canadians, and Swedes, for instance) show a greater degree of independent self-perception than do non-Westerners (e.g. Native Americans, Cook Islanders, East African pastoralists, Malaysians and East Asians). These differences have been measured by various

means, including the 'Twenty Statements Test', on which participants make statements about themselves and rate them on a scale designed to measure whether self-evaluations are more linked to individual qualities than to group roles.

Henrich et al. then review a wide range of literature on the relative inclinations of diverse populations towards holistic vs. analytic reasoning and field independence vs. field dependence. This consistently demonstrates that Westerners evidence a greater inclination towards analytical reasoning and field independence than do non-Westerners. Within the non-Western category, East Asians, Malays, Asian-Australians, Chinese-Malaysians, (Indian) Hindus, Filipinos, Japanese and Mexicans have been found to be more holistic/field-dependent than people of Western European background, from various countries. Even within Italy, significant differences have been observed between northerners (more analytical/field-independent) and southerners (more holistic/field-dependent).

One example of the experimental methods used in making assessments of dominant cognitive style is eye-tracking, measuring in this case differences in time spent looking at either an object situated in a field or at the field itself. A second is the 'Embedded Figure Test', which evaluates participants' ability to distinguish a figure from its background. A third is responses to questionnaires designed to ascertain the extent to which participants see personality traits as, on the one hand, individually stable and fixed or, on the other, determined by roles and duties performed in the interests of the group to which the individual belongs. In the latter case, behaviour is commonly conceived as varying with the group identity of the other people involved in a given interaction (Heinrich et al. p. 72). The latter technique makes particularly clear how closely cognitive style is related to self-perception.

In summary, Henrich et al. (2010, p. 72) conclude:

"…compared to diverse populations of non-Westerners, Westerners (1) attend more to objects than fields; (2) explain behavior in more decontextualized terms; and (3) rely more on rules over similarity relations to classify objects."

Another non-Western population that Henrich et al. mention is Arabs. In the study concerned (Zebian & Denny 2001), Arab-background participants generally—i.e. regardless of education and upbringing—were found to be significantly more field-dependent (i.e. more 'integrative') than Euro-Canadians (more field-independent, or 'differentiative'), measured according to their respective performance in sorting various objects into groups. The sorting strategies involved participants' perception of unidimensional vs. multidimensional and major vs. minor object properties. In each pair of oppositions, dominant use of the first strategy was taken to evince a differentiative and the second an integrative cognitive style. Arab participants tended to sort objects using an integrative—i.e. field-dependent—strategy. However, there was also a significant effect of education/culture on degree of field-dependence, with the 'traditional' group proving less field-independent than the 'modern' (i.e. more Western-educated) group.

Cognitive style: implications for experiments

The results of Zebian & Denny (2001) have a number of possible implications for the perception experiments described in this thesis, testing as they do (though not directly comparing) the linguistic judgements of Arab and Western listeners, with NSNLs and linguists present among both populations. Overall, however, the number of complicating factors makes firm conclusions difficult to reach.

Firstly, consider the implications of Zebian & Denny's findings for possible group-level differences in performing the principal task of the experiments in the thesis: identifying speakers of a specific dialect. One expectation might be that Arab and Western listeners will perform *in some way* differently from each other. Another may be that, among Arab listeners, differences in accuracy will co-vary with presence or absence of linguistic training. However, beyond these vague predictions, it is difficult to imagine in which direction differences in reasoning/field perception might influence linguistic judgements. Would field independence confer an advantage in accuracy, and field independence engender a disadvantage? Would an academic linguistic education, combined with relatively high field independence, be a help or a hindrance? Owing to the lack of direct evidence on how the two cognitive styles influence

language judgements, no informed predictions can be made of the likely answers to questions such as these.

This inquiry could be extended to observations made above, in subsection 4.3.1., of the endurance among some Arabic speakers of consanguineous marriage and tribal affiliation. Would partaking in 'viscous' social bonds of this type accentuate the prevalence of the integrative cognitive style and thus the ability to recognise fellow members of the ethnolinguistic 'ingroup'? Accepting the claim of Henrich et al. that holistic reasoning/field dependence is a correlate of a generally group-oriented self-perception—such as might be expected among people who, as statistics show, are inclined to marry close relatives—this would be a *prima facie* plausible prediction. Again, though, an absence of directly language-related evidence precludes the making of secure predictions on this account.

For the secondary task of the experiments in this thesis—i.e. citing the linguistic cues that informed decisions about the provenance of accents—the implications of Zebian & Denny's findings are no more obvious. Since education showed a significant within-group effect on field perception among Arab listeners, a clear-cut difference in the types and/or quantity of features cited might be expected to obtain between linguistically-educated Arab listeners and their NSNL counterparts. This expectation, however, is complicated by two factors.

First is Zebian & Denny's finding that, *regardless of education*, Arabs inclined significantly to the holistic/field-dependent/integrative end of the spectrum. This was so to a greater degree, according to Henrich et al., than similarly designed studies have demonstrated for strongly field-dependent Chinese participants. Second is the fact that the experiments in this thesis test accent judgements at the national level—i.e. those of Arabs of widely divergent dialectal backgrounds—with (Damascene) Syrian Arabic as the target. It is here that convergent theories and empirical findings in variationism and sociocultural evolution again become vital. If membership of a particular linguistic community (i.e. being a native speaker of a specific dialect) is the primary factor in the ability to perceive fellow speakers, linguistic education must play a comparatively minor role. Overall, then, predictions founded on (degree of) native-

speakerhood would appear to be more securely supported than those implicating differences in education or field perception.

A further implication of Zebian & Denny's experiment concerns the format of reports used by LAAP agencies, a matter covered also in Chapters 2 and 3. These reports, demanding as they do the close description of individual phonological and mophosyntactic segments, are oriented towards the Western linguistic method—an analytical and field-independent tradition *par excellence*. However, if Arab NSNLs are indeed pronouncedly holistic/field-dependent in their cognitive and perceptual style, a format mandating adherence to the Western linguistic method must be considered a questionable match for the exercise of their capabilities. An obvious difficulty with proposing change in this area is that the rules of evidence in the Western nations where LAAP is conducted likewise require a distinctly analytical approach, and few would argue that it is wrong in this regard. This deep-rooted epistemological incongruity, as with many aspects of existing LAAP practice, seems irresolvable.

Cognitive style: implications for the theory of homogeneism

There are, then, a number of difficulties in making solid predictions for the present experimental series based on Zebian & Denny's findings of population-level differences in cognitive style. Their research is perhaps more relevant to strengthening refutations of the theory of homogeneism attempted elsewhere in this chapter. For example, it seems reasonable to suppose that non-Westerners' greater degree of field-dependence, and Westerners' relative field-independence, *somehow* influences their respective perception of language. If this is so, the universalistic assumptions of the theory of homogeneism appear still more dubious, and positive support is lent to previous arguments that there may be group-level differences in perceptions of language variation—e.g. between various ethnolinguistic groups or between native and non-native speakers of particular varieties.

In addition, the fact that Arabs generally (regardless of education) were found by Zebian & Denny to be more field-dependent than relatively uneducated Euro-Canadians casts doubt on the assertion that individuals' level of *linguistic* education fundamentally determines perceptions of language variation. Instead, again as suggested by theories in sociocultural

evolution and Labovian sociolinguistics, individual factors (e.g. level of linguistic education) are of secondary importance to those operative at the level of the speech community. This is so even though neither the individual nor the group can exist without the other, and the two are in a fundamentally interactive relationship both in the present and (pre)historically.

In short, given the evidence adduced so far, it seems plausible that a *group* of native speakers would perform considerably better in perceiving their own variety than would a *group* of non-natives, whatever their respective educational backgrounds. I stress, however, that no denial of wide variation in the strength of language perceptions among individuals is implied here (cf. Wilson & Foulkes 2014).

4.4.3. Anti-reductionist philosophy: the irreducibility of language perception

Anti-reductionist philosophers adhere in common to five premises bearing on the present discussion—and likely more. A specialist in the field of philosophy would probably be aware of other relevant anti-reductionist postulates. However, owing to my limited background in philosophy, I rely here chiefly on Polanyi (1966, 1968, 1970) and Schumacher (1977). The five common premises I have identified are as follows.

First, perception is thoroughly unified with the rest of the human organism. Contra the mind-body dualism dominant in the Western scientific tradition since Descartes (Schumacher 1977, pp. 17-18), anti-reductionists regard perception as inseparable from—indeed, a "sentient extension" of—the human body (Polanyi 1966, pp. 16-17).

Second, human perception is a primary mental category; it is neither contingent upon, nor capable of demonstration by, the reductive experimental methods commonly employed by modern descriptive science to investigate non-human phenomena (Schumacher 1977, pp. 134-7; Polanyi 1970, p. 92). Reductive investigations of "atomic or molecular accidents" (Polanyi 1968, p. 1311) are capable only of inadequate—at best partial—explanations of the fundamentally primary nature of perception.

Third, while anti-reductionists do not deny either the objective nature or the discoverability of reality, they assert that the exercise of 'objective' scientific methods does not efface the fundamentally personal nature of perception. There is an unavoidably subjective background to the work of all scientists. Knowledge of the lower levels of reality can be apprehended by the tools available to science, but interpretation of the whole depends on subjective perception. In short, there is no 'view from nowhere' (Polanyi 1966; Schumacher 1977).

Fourth, nature—including natural language—is hierarchical and 'emergent' in structure (Lähteenmäki 2004; Harrison & Raimy 2007; Polanyi 1966, 1970). Natural phenomena owe their capacity to function to their interaction as a system, and this interaction gives rise to new functions (Polanyi 1968). However, the lower-level processes (e.g. chemical reactions or molecular change; cf. Schumacher 1977) commonly investigated by science cannot account for the unified human perception of natural entities. In terms more pertinent to the present discussion, Polanyi (1970, pp. 89-90) uses the example of language:

"The lowest level is voice production, sounds which leave open all kinds of uses to which the voice may be put. The next lowest levels are vocabulary and phonetics which restrict the manner in which the voice is used while leaving open the many forms of order which are supplied by the next level, the rules of grammar and syntax. Grammar and syntax restrict the use of vocabulary by making sentences while leaving the content of sentences open. The highest level in this hierarchy is the level of content or meaning. Meaning or content exercises control over the construction of sentences and the relations among them."

Fifth, "we know more than we can say": many facets of human knowledge are "tacit" and thus not susceptible of articulation or reduction to rules, 'even' by science (Polanyi 1966, 1970; Schumacher 1977). As Polanyi (1970, p. 92) explains:

"I am not able to specify with distinctness the particulars of a comprehensive entity. In this sense I know more than I can tell."

124

Augusto (2018) offers a valuable terminological distinction, referring to tacit knowledge as "procedural" and its counterpart, explicit knowledge, as "declarative". With the same distinction in mind, Stanley & Williamson (2001), following the 20th century British philosopher Gilbert Ryle, employ the terms "knowledge-how" and "knowledge-that".

Augusto (2018) also states that an example of procedural knowledge, or knowledge-how, is the ability to speak and understand a particular language variety as a native. The Polanyi quotes above, however, suggest that this ability does not in itself permit description of either the nature or the interaction of the hierarchically-ordered levels governing the use of language.

The anti-reductionist conception of the irreducible and 'emergent' nature of language essentially parallels conclusions drawn earlier in this chapter from variationist sociolinguistic theory, sociocultural evolution and social psychology, all of which have been interpreted as countering the theory of homogeneism. How and in what ways this is so can be explained by asking and then answering two questions. First, to what extent might explicit knowledge of a language variety, such as that availed by an education in the reductive discipline of linguistics, be sufficient as a substitute for the tacit, unified knowledge of even an uneducated native speaker? Second, would a linguistic education improve the ability of a native speaker of a particular variety in recognising fellow natives?

Both questions can be answered simply. At the very least, explicit knowledge is different in kind from implicit (or tacit) knowledge: the first is declarative and the second procedural. Linguistic education might therefore assist a native speaker in reductively *describing* the discrete parts of his/her variety—its phonology, morphosyntax and so on—but it could never act as a *substitute* for, or fundamentally alter, the holistic perceptions of language made available by tacit knowledge (Polanyi 1966, pp. 16-17).

As we have seen, the theory of homogeneism derides NSNL perceptions as fundamentally conditioned by top-down institutional pressure, while it positions the knowledge of linguists as 'scientific' and wholly objective. To this, anti-reductionists would retort that there is no such thing as a perfectly objective perspective and that, furthermore, the use of such reductive methodological tools as the IPA cannot reveal the unified nature of any particular language

variety (cf. also Nolan 2012). Social psychologists might enter the substantively similar response that, given the gradient nature of cognitive styles, no completely field-independent (i.e. perfectly objective) stance is possible.

As we have seen, even while stressing that cognitive style is a scalar and non-categorical construct, social psychologists have located significant group-level differences in its manifestation. Arabs more educated in the Western tradition evince a more field-independent mode of perception than those more educated, yet all maintain a decidedly field-dependent perceptual strategy when compared to, for instance, Westerners and Chinese. This evidence undergirds the anti-reductionist position that human perception is a primary category, which retains its primacy even in the face of top-down factors such as formal education.

Sociocultural evolution and variationist sociolinguistics, meanwhile, are in accord with anti-reductionism in acknowledging the inimitible character of native speaker knowledge. Only native speakers of a particular variety possess the unified, deeply-embedded, tacit knowledge necessary to transmit subtle phonetic signals of speech community membership, and to fully comprehend their social meaning when transmitted by other members.

4.5.Summary and discussion: alternative principles of native speaker perception

In this chapter I first outlined the mainstream critique of NSNL perceptions in the LAAP literature (the theory of 'homogeneism'). This posits the dissemination of normative language ideologies as an important factor conditioning the linguistic perceptions of individuals. According to proponents of the theory, the influence of such ideologies can be effectively countered only by advanced education in linguistics, which apparently grants its recipients intellectual independence from the normative pressures exerted on other individuals—a perfectly objective 'scientific' standpoint, effectively a 'view from nowhere'.

I then examined theoretical accounts and empirical studies of native speaker perception from the fields of perceptual dialectology, variationist sociolinguistics, sociocultural evolution, social psychology and anti-reductionist philosophy. This examination revealed an alternative view of

native speaker competence, the three most important insights of which are set out below in the form of general principles (to which there may always be individual exceptions).

(1) Perhaps due to hardwiring by evolutionary selection pressures, native speakers of particular linguistic varieties harbor unified, deeply-embedded, tacit, variety-specific knowledge. Possession of this kind of knowledge makes them uniquely able to perceive authentic fellow native speakers of the same variety.

(2) Owing to their tacit, holistic perceptions of language NSNLs are unlikely to be able to express, in the reductive and partial terms customary in linguistic analysis, precisely how they are able to distinguish fellow native speakers from non-natives.

(3) Expertise in particular language varieties, acquired by either native or non-native speakers through formal education, cannot functionally replace native speaker perception of language as an integrated system. However, depending on its nature, such expertise may assist experts in the reductive description of the elements of language.

These three principles inform the hypotheses, formulated in Chapter 6, guiding the experiments conducted in this thesis. In the following chapter I critically review empirical work on NSNL judgements and confidence in tasks more or less analogous to LAAP. Findings emerging from this review further undergird the hypotheses stated in Chapter 6.

Chapter 5: NSNL judgements & confidence

5.1. Overview

In the previous chapter I evaluated opposing theories and empirical studies of NSNL perception. In this chapter I survey empirical work concerning NSNL judgements, and the confidence expressed in such judgements.

I first explore the empirical basis of claims by Fraser (2009) that NSNLs are on the whole unacceptably unreliable in the basic LAAP task: judging speakers' origins. I find that this work, conducted in non LAAP-like experimental settings, bears questionable relevance to real-world LAAP.

I then examine the varying influence of ecologically-valid prior studies on the methodology employed and the predictions made in this thesis. I show that some of these studies have produced evidence that NSNLs are as accurate as (or more accurate than) non-native speaker linguists in addressing the central LAAP question.

I next evaluate findings as to the relationship between listener accuracy and the confidence expressed in these judgements. I find that on this score the literature is equivocal, and that no secure overall predictions on the relationship between the two variables can be reached.

I reach two conclusions. First, the relevant research indicates that NSNL analysts in primary-phase LAAP can legitimately act as judges of their own variety. Second, this fact is sufficient to demonstrate that NSNL asylum applicants may be able to do likewise in the context of the perception tests developed via the experimental work of this thesis.

5.2. Non LAAP-like empirical studies

Fraser 2009 (pp. 119-123) provides a comprehensive (at that time) overview of research on the capacity of native speakers to judge accents. She grants, however, that "…unfortunately, a thorough literature review revealed there is very little research that bears directly on the LADO situation". This is because "…none of the studies…called for the kind of analysis and justification required in a LADO judgment" (2009, p. 124). By her own admission, then, the studies Fraser reviews are, at best, of limited relevance to LAAP. Yet their inapplicability runs much deeper than Fraser acknowledges.

One of the design shortcomings of many of the studies Fraser cites is their use of stimuli featuring individual sounds in isolation, or fragments of speech of various kinds. As demonstrated in Chapter 2, this is quite unlike the situation in LAAP, which is conducted on the basis of at least 15 minutes of spontaneous speech.

Another dissimilarity to LAAP is that the listeners in the experiments Fraser summarises apparently had little or no time for review of conclusions. As also described in Chapter 2, this is inapplicable to primary-phase LAAP, where there is considerable scope, via NSNL-to-NSNL and NSNL-to-linguist discussion and cross-checking, for the reconsideration of initial judgements. This is so even at Verified, by whom (non-binding) preliminary conclusions are supplied to the UK Home Office within 30 minutes of an interview. The full report, however, is due in three to five working days, meaning that there is time for review of preliminary findings.

Moreover, as Nolan (2012, pp. 274-5) points out, some of the studies Fraser reviews call on listeners to identify genuine or imitated varieties other than their own (e.g. Markham 1999, Neuhauser & Simpson 2007, Munro, Derwing & Burgess 2003). This task is seldom if ever required in LAAP. Put in Nolan's terms, the design of these experiments demands a 'one of them?' judgement, as opposed to the 'one of us?' decision required in both the verification and the classification tasks that arise in LAAP (cf. Chapters 2 and 3).

Furthermore, the design of two studies Fraser cites makes their applicability especially suspect. Niedzielski (1999) subjected listeners to fairly intense explicit priming, and her (single-sound)

stimuli involved synthetically modified stimuli (i.e. they were not natural speech tokens). This general experimental environment, though evidently common in studies of perceptual dialectology (cf. Chapter 4), is incomparable to LAAP. Munro et al. (2003) actually *reversed* the experimental stimuli. Fraser does not mention these severe limitations on the relevance of the two studies.

Additionally, there is the problem of incompatibility between the respective educational backgrounds of experimental subjects and NSNLs involved in primary-phase LAAP. Fraser, one of the authors of the *Guidelines*, was expressly motivated to undertake her review by doubts about the judgements of analysts in LAAP. This position is evident in her declaration (2009, p. 118) that

> "...judgments about language based on 'folk knowledge', while sometimes accurate, are not reliable enough to be the basis of important decisions ..."

Yet many of the experiments Fraser reviews do not test the judgements of NSNLs—i.e. listeners who are allegedly guided by errant 'folk knowledge'. Instead, they commonly involve linguists trained to a fairly advanced level (generally in phonetics). Indeed, this is evident in the title of Fraser's paper: *The role of 'educated native speakers' in providing language analysis for the determination of the origin of asylum seekers*. According to the precepts of the theory of folk knowledge/homogeneism, the judgements of linguistically-trained listeners cannot reasonably be inferred as representative of the judgements of NSNLs.

A further and related problem is that Fraser does not survey comparable experiments performed on the judgements of *linguist* non-native speakers, a category to which most counter-analysts belong (cf. Chapter 2). None of the studies mentioned can therefore be taken to show that NSNL judgments in primary-phase LAAP are unreliable in comparison to those that would be made by linguist non-native speakers.

In summarising these dubiously relevant findings, Fraser (2009, p. 128) concludes:

"...the literature review...supports the assumption that people are generally better at recognising their own accent than identifying other accents, [yet] it makes very clear that even in this task, even with European languages, they are far from generally reliable..."

Here Fraser concedes that listeners are more accurate in 'one of us?' than in 'one of them?' tasks. In her commentary on Neuhauser & Simpson (2007), she even goes as far as to agree with the authors that listeners' performance was "generally good" (Fraser 2009, p. 121). Yet elsewhere she repeatedly suggests that this is not sufficient. It is difficult to know what degree of accuracy Fraser would find acceptable. What can be said is that any expectation of total accuracy is unrealistic in both principle and practice. As Cambier-Langeveld (2018a) explains, the quality of data and the judgements of human beings are variable and rarely perfect. Such limitations are the main reason that forensic conclusions—including those arising from LAAP— are, or should be, expressed probabilistically (see Chapter 3). As a forensic practitioner herself, Fraser presumably acknowledges the imperfection of human judgements; indeed, she has stated her support for the probabilistic statement of conclusions in LAAP (e.g. as a signatory to the *Guidelines*). This being the case, it is difficult to imagine that Fraser would counsel against the use of experts in other forensic contexts on the grounds that their judgements are unlikely to be accurate in all cases.

Fraser (2009, p. 124) makes the point that judgements of accents are:

"...based on many aspects of context, background knowledge and prior expectations, as well as on speech characteristics. However, people attribute their judgments to characteristics of speech itself, ignoring the contribution of context ... "

In response to this assertion we may ask: Where is the listener who is *not* influenced by "context, background knowledge and prior expectations" of one kind or another? As I argued in Chapter 4, such factors, whether conceptualised as 'tacit' knowledge or (some degree of) field-dependence, may be integral to the ability of native speakers to detect fellow natives. It is further asserted in the same chapter that a perfectly objective, field-independent 'view from

nowhere' is impossible. However, Fraser's own theoretical standpoint, with its negative conception of 'folk knowledge', would seem not to admit arguments such as these.

Nonetheless, in spite of its flaws, Fraser's paper is worthy of recognition as the first attempt to summarise the then-extant literature on native speaker judgements. The observation that Fraser failed to attend to many of its most serious limitations should be tempered by the understanding that prior to 2009 no empirical work in LAAP-like contexts had yet been conducted.

I turn next to consider the range of more or less ecologically-valid experimental work that has appeared since the publication of Fraser's survey. Some of these studies were influential on the design of the experiments in the present thesis, as well as on the composition of the hypotheses they test.

5.3. LAAP-like empirical studies

As we have seen, at the time Fraser's paper was published, no LAAP-centred empirical work had yet been conducted. This is no longer the case: there is now a small but growing body of relevant experimental studies. In the following subsections I summarise the findings of these studies from two perspectives: firstly, evidence of the relative accuracy of judgements by various listener groups; secondly, evidence of the relationship between accuracy and confidence. I also examine the respective studies' influence on the design and hypotheses adopted in the experiments undertaken for this thesis.

5.3.1. Cambier-Langeveld (2010b)

Cambier-Langeveld (2010b) compared the respective real-life conclusions of the OCILA (i.e. primary-phase LAAP conducted via the team approach) and counter-analysts in eight cases, of a total of 124, where asylum applicants had revealed their true linguistic identity under an amnesty offered by the Dutch government. One counter-analyst was a native-speaker linguist,

and five were non-native speaker linguists. One of the latter, who was responsible for two of the eight counter-analyses, was also a signatory to the *Guidelines* (Cambier-Langeveld 2012).

The OCILA's assessment of the applicant's background proved accurate in all eight cases. Seven of the eight counter-analyses were inaccurate. The only accurate conclusion was reached by the single counter-analyst who was a native-speaker linguist. Unlike in the studies summarised by Fraser (2009), in all cases the relevant question (at least for OCILA's NSNL analysts) was of the 'one of us?' type.

Despite generally marginal criticisms (e.g. by Verrips 2011 and Fraser 2011), this study is consequential as a defense of the judgements of NSNLs—and thus of primary-phase LAAP, at least as conducted by the OCILA. But it is not a direct comparison between NSNL judgements and those of non-native speaker linguists, since the former may have been mediated to some degree by the mechanism of the team approach and the latter possibly influenced by the (undeclared) input of NSNLs. Moreover, it is a *post hoc* evaluation of results rather than a true experimental study.

For these reasons, Cambier-Langeveld's study had no direct influence on the methodology of the series of experiments conducted for this thesis. It was, however, naturally influential on the fundamental prediction of NSNLs' accuracy in judging their own variety—i.e. answering the 'one of us?' question.

Matras (2018) merits brief mention here. This paper, a critique of a subset of Verified's conclusions, is valuable in the sense that, like Cambier-Langeveld 2010b, it examines the validity of conclusions drawn from real-world LAAP. It is nonetheless left aside here, because the true origins of the applicants in the cases discussed remain unknown.

5.3.2. Wilson (2009), Shen & Watt (2015) and Hedegard (2015)
The broad conceptual direction, hypotheses and/or design of the experiments in this thesis draw primarily on three studies of accent judgements by various types of listeners in simulated LAAP contexts: Wilson (2009), Shen & Watt (2015) and Hedegard (2015). Each of these studies

produces valuable findings in this connection. Nevertheless, methodological decisions made in all three studies about listener or stimulus selection—or both—impose some limitations on their relevance to the experiments undertaken for this thesis. The limitations relate to two criteria. The first is the extent to which the designs of the three prior experiments result in a proper 'one of us?' question being posed to listeners. The second is whether their design serves the central purpose of my experiments: developing new tests of the judgements of asylum applicants.

Wilson compared the ability of four groups to correctly identify Ghanaian English targets and reject Nigerian English foils. Four different types of listeners were sorted into a corresponding number of groups: non-linguist native speakers of Ghanaian English, LAAP professionals, academic phoneticians, and undergraduate students of linguistics. Wilson used a mixture of recounts of the *Cinderella* story and spontaneous speech in two-minute stimuli. Listeners were briefed on the diagnostic phonetic and phonological markers of Ghanaian English and asked the LAAP-like (speaker verification) question, 'Do you believe this person is speaking Ghanaian English?' Wilson's central finding was that Ghanaian NSNLs—for whom the task was of the 'one of us?' kind—were more accurate, at 86%, than any of the other three groups. Academic phoneticians were the next most accurate, at c. 60% (but see 5.3.3. for discussion of the assumptions made in Wilson's inferential statistical analysis).

It is common at the primary stage of LAAP that linguists work on languages in which they have little or no competence. In these cases, they operate in tandem with NSNLs who are speakers of the variety being assessed (the 'team approach'). Wilson's Ghanaian English speakers/listeners were somewhat analogous to NSNL analysts, with the important difference that Wilson's participants had undergone no forensic or linguistic training whatsoever. Her 'LADO professionals', meanwhile, were analysts working in real-world LAAP, without knowledge of the language variety at issue.

Foulkes & Wilson (2011, p. 692) state that, in Wilson's experiment,

"...few of the academic or LADO participants had prior experience of work with [Ghanaian English], hence the task was in this sense deemed to be equally challenging for all groups."

However, a limitation associated with the design of Wilson's experiment is that, while many (not all) of Wilson's academic phoneticians and students were native speakers of English—though not of Ghanaian English—none of the LAAP practitioners was a native speaker of *any* variety of English. Neither were any of the non-Ghanaian listeners speakers of any *African* variety of English. Further, while the phonetician and undergraduate student groups each outperformed the LAAP experts, there is no consideration of whether this might have been due to disparities in the groups' overall English-language proficiency—i.e. the extent to which it was a 'one of us?' task for all listeners. It is possible, of course, that being a native speaker of a non-African variety of English is of little help in identifying Ghanaian English; it may even be the case that it would serve to mislead. But the fact that this variable was not accounted for lessens somewhat the applicability of Wilson's experiment design, and consequently her results, to the thesis.

Even with these caveats entered, Wilson's results considerably influenced the hypotheses tested in the present experimental series. Of particular interest is her finding that Ghanaian NSNLs (and, in fact, all her listener groups) were much more likely to falsely accept non-Ghanaians than to falsely reject Ghanaians. One possible explanation for this result may be the 'other-accent effect' (Atkinson 2015), which refers to the tendency of listeners to find it relatively difficult to distinguish among individual speakers with an accent other than their own. As used in the existing literature, the concept refers to listeners' relative success in recognising individual speakers in forensic voice line-ups (Atkinson 2015). For example, the prediction would be that a New Zealander would find it more difficult to recognise an individual British speaker in a line-up composed of all British speakers than an individual New Zealand speaker in an all-NZ line-up. In this context, the operation of the effect is supported by the results of multiple studies (Adank et al. 2009; Atkinson 2015; Stevenage et al. 2012) However, the term 'other-accent effect', and its associated concepts, can also be applied to the relative ability—with the same predicted bias—of listeners to identify speakers with a given location in LAAP-like environments. As further specified in Chapter 6, the prediction that the other-accent effect would affect the pattern of listener responses in the present series of experiments was therefore incorporated into hypotheses (i) of the thesis.

Another way in which Wilson's study was influential on the hypotheses constructed for this thesis was her inference that, despite similar overall accuracy in the tasks, Ghanaian NSNLs and academic phoneticians appeared to respond to different cues:

"One sample yielded a 44% error rate for Ghanaians, with only 10% for the academics, whereas another sample gave an 11% error rate for the native speakers and 40% for academics."

(Wilson 2016, p. 82)

Hypothesis (ii) of the thesis is built mainly on the theory and experimental results detailed in Chapter 4. However, my investigation of this work was in part prompted by Wilson's tentative conclusion that NSNLs and linguists rely on different cues in identifying specific language varieties.

Shen & Watt (2015) offer their research as a possible first step towards developing a new LAAP test for applicants, not analysts. Indeed, the series of experiments in this thesis follows in the broad conceptual direction they prefigure. Shen & Watt compared the accent recognition performance of L1 English listeners and Chinese (L2 English) listeners in two tasks: (1) distinguishing among stimuli featuring Chinese, Korean and Japanese L2 speakers of English; and (2), when a Chinese L1 speaker was identified, determining where in China (south, north or west) the speaker was from. Unlike in Wilson or Hedegard, none of Shen & Watt's listeners was furnished with a brief on Chinese phonetics. A combination of read texts and spontaneous speech in 15-second stimuli was employed. Chinese (English L2) listeners were found to be significantly more accurate than the L1 English listeners in distinguishing Chinese from Japanese and Korean L2 English speakers.

The difficulty with Shen & Watt's design is that no listener group was really called upon to make a 'one of us?' judgement. It is a central fact of LAAP practice that the question at issue almost always involves distinguishing between a native and a non-native speaker of the linguistic variety hypothesised on the basis of the applicant's claimed residential history. The core LAAP task does not usually involve detecting a foreign accent in a language genetically unrelated to

the substrate language. Shen & Watt's experiment, though, compared the ability of two types of listener to detect an East Asian L1 substrate in a series of English L2 stimuli. Under these circumstances neither of the listener groups was (or could have been) composed of true native speakers of the varieties of English—really inter-languages—found in the stimuli. Nor, naturally, was this true of the speakers in the stimuli, since all were L2 English speakers. The consequent absence of the 'one of us?' question in Shen & Watt's study—a consequence of their selection of both stimuli and listeners—limits the relevance of their results to present series of experiments.

Hedegard (2015) compared the performance of three groups: linguistically untrained Japanese native speakers, Japanese native speakers trained in linguistics, and non-native linguists who were also competent speakers of Japanese. The groups' performance was assessed under four conditions: one group comprised untrained native speakers paired with non-native linguists; those assigned to one of the remaining three conditions worked alone.

Similarly to the procedure of Wilson (2009), all listeners who were included in one of the three unpaired conditions were sent an 'intonation information sheet'. Hedegard's stimuli were of spontaneous speech recorded in a business setting, edited into one-minute stimuli. While Wilson and Shen & Watt had focused exclusively on segmental phonetics, Hedegard concentrated on intonation. The question Hedegard posed to participants was, 'Where do you think this speaker is from?'). In combination with the multiple-choice alternatives provided ('Tokyo', 'Kansai', 'Tohoku' or 'I don't know'), this question suggests a LAAP-like verification task.

Hedegard's listener categories and assigned conditions are a closer match than either Wilson's or Shen & Watt's to the various types/combinations of analysts found in LAAP, and her stimuli featured speakers of native, not L2, varieties. Listeners in Hedegard's two native speaker categories were native speakers of just one of the target dialects (i.e. Tokyo; the other stimuli featured the dialects of Tohoku and Kansai). In short, her design decisions resulted in an authentic 'one of us?' task.

Similarly to the results of Wilson's experiment, Hedegard's native speaker listeners—of both types, and in both of the conditions to which they were assigned—were more accurate than others in accepting Tokyo speakers but demonstrated a lesser advantage in rejecting non-Tokyo speakers, especially those from Tohoku. This again appears to validate the other-accent affect. As discussed in Chapters 7 to 11, similar findings emerged from experiments in the present series, on both English and Arabic. Local listeners were more accurate than non-locals in accepting speakers of their local dialect but not in rejecting non-local speakers. Hedegard's NSNL + non-native linguist pairing proved the most accurate of the four conditions in the task assigned, lending support to the validity of the team approach in LAAP.

The design of the experiments in this thesis departs somewhat from those employed by Wilson and Hedegard, each of whom sought to investigate the accent-recognition accuracy of different kinds of analyst in a task more or less closely related to the real-life LAAP question as currently posed. Here, however, as presaged by Shen & Watt, the idea is to explore the relative accuracy of different types of NSNL listener, primarily in order to develop new tests of perception. Analysts would have at most a minor role in this test, in which the asylum applicant him/herself would make the judgements. This means that comparison of listener categories strictly analogous to the categories of *analyst* found in LAAP (as particularly well exemplified by Hedegard's work) is beside the point. Rather, to the maximum practicable extent, it is vital to investigate the judgements of listeners selected to mirror the types of *applicants* commonly encountered in LAAP.

Overall, leaving aside matters of design, the results of Wilson (2009) and Hedegard (2015) suggest two reasonable inferences of particular relevance to the experiments conducted in this thesis. The first, reinforcing Cambier-Langeveld (2010b), is that NSNLs working in primary-phase LAAP are relatively reliable judges of their own linguistic variety—especially, in Hedegard's experiment, when paired with a (non-native) linguist. The second is that authentic NSNL asylum applicants will perform more accurately than inauthentic applicants if confronted with an unambiguous 'one of us?' task. These empirically-founded conclusions amplify the validity of the three principles of native speaker perceptions specified at the conclusion of Chapter 4.

Two other relevant studies—Muhammad (2021) and Kudera et al. (2022)—appeared after the hypotheses tested in this thesis had been composed and its experiment design determined. Muhammad (2021), in a replication of Wilson 2009, found that native speakers of Nigerian languages were 71 to 92% accurate in identifying the accent of fellow speakers in English-language stimuli. Kudera et al. (2022) compared the accuracy of NSNLs of four Slavic languages in identifying speakers their own and other Slavic languages exclusively on the basis of pseudo-words designed to encode contrasts in lexical stress. Their finding was that NSNLs of three of the four languages identified their own language most accurately, the exception being native Bulgarian NSNLs.

5.4. Relationship between confidence and accuracy

Fraser (2009, 2011, 2019) and Patrick (2012) assert that NSNL listeners show greater confidence than linguists in their judgements of speaker origin and that confidence and accuracy are poorly correlated. As observed above, Fraser's (2009) review overwhelmingly concerns the accuracy in accent identification of native-speaker linguists, not NSNLs. In discussing the relationship between accuracy and confidence, she cites from one in-progress PhD study that apparently found a poor correlation between accuracy and confidence. She concludes (2009, p. 128):

"…it is a mark of the non-expert to place unjustified confidence in one's own judgments. Such non-experts might well be quite willing to give evidence in LADO cases."

Fraser later states (2009, p. 132) that trained phoneticians, too, evidence undue confidence in their judgements:

"Both native speakers and trained phoneticians have been shown to be far less accurate than their confidence suggests."

Fraser's point seems to be that investigating the relationship between accuracy of and confidence in judgements is vital to establishing the validity of LAAP:

"It would be extremely useful to have findings from research which clearly differentiated identification, recognition and discrimination abilities in relation to a range of different kinds of task and different types of data (especially including different kinds of in-person and third party interviews), and correlated these abilities with confidence ratings."

(2009, p. 133)

Wilson's pioneering work (2009) fulfills Fraser's desiderata of being both conducted under LAAP-like conditions and examining listeners' confidence in their own judgements. Hedegard (2015) and Muhammad (2021), too, are ecologically-valid experiments which assess correlations between confidence and accuracy; Shen & Watt (2015) do likewise.

Fraser (2011) finds in Wilson's study a "poor correlation between accuracy and confidence" in NSNLs' "individual judgements" (p. 126); Patrick (2012) agrees in this interpretation. However, this characterisation ignores the effect of a number of somewhat questionable assumptions underlying data selection for Wilson's statistical analyses.

In all four of her listener groups (i.e. not only among NSNLs), Wilson found"...no significant correlation between confidence and accuracy of response" (Foulkes & Wilson 2011, p. 692). Further, as measured on a five-point Likert scale,

"...responses by the NS[NL]s in particular tended to be polarized, with highly unlikely or highly probable the most frequent responses"

(Foulkes & Wilson 2011, p. 692).

This resulted in the unsurprising finding that NSNLs, whether correct or incorrect, were significantly (p = < .0001) more confident than all other groups (Wilson 2009, p. 16). Because Wilson's NSNLs left no 'unsure' responses, a greater proportion of their incorrect responses than those of other groups was inevitably associated with judgements of undue confidence.

Fraser also reports (2011, p. 126) that the academic phoneticians in Wilson's experiment:

> "...had lower correct scores than native speakers but, crucially, showed better evaluation of their conclusions (via 'unsure' responses) – to the extent their outright error rate was similar to that of the native speakers."

This claim is true as far as it goes: in the final analysis, Wilson found no significant difference in accuracy between NSNLs and academic phoneticians (Foulkes & Wilson 2011, p. 693). But this result is likely attributable to Wilson's decision to exclude 'unsure' responses from her inferential statistical analysis on the grounds that, among her academic phonetician group:

> "...these did not simply equate to 'don't know'. Rather, academics indicated that a decision could not be reached with adequate confidence. They typically offered a full explanation of the decision, outlining observed features that matched the training materials, and other features that did not."

(Foulkes & Wilson 2011, p. 693)

However, before 'unsure' responses were excluded, Wilson's NSNLs achieved 86% accuracy, with academic phoneticians next best at 60% (Foulkes & Wilson 2011). On the combined basis of these raw results a positive correlation may well have been located between overall accuracy and confidence (with the latter measured by number of 'unsure' responses), at least among Ghanaian NSNLs. Yet the discovery of any relationship between the two variables was forestalled by Wilson's decision to remove 'unsure' responses. Whatever the merits of this treatment of 'unsure' responses in the case of academic phoneticians, it is arguably less valid when dealing with some or all of the other listener groups, among whom such a response might have been plausibly considered as amounting to a simple 'don't know'.

Neither Wilson (2009) nor Foulkes & Wilson (2011) specifies whether other listener groups showed similar sophistication to academic phoneticians in their comments on stimuli to which they gave an 'uncertain' response. However, examination of a document in which Wilson compiled verbatim the comments of all listeners in her experiment shows that those belonging to the academic phonetician group (and to a lesser extent the undergraduate phonetician

group) were indeed more inclined than others to express reasoned doubt in their conclusions. This, combined with the fact that listeners from the two phonetician groups returned the largest percentage of 'unsure' responses, suggests that Wilson's decision not to consider the interaction between 'unsure' and 'decision' responses may have unbalanced results to some extent. In any case, the indisputable effect of excluding 'unsure' responses was to elevate the accuracy of all three non-NSNL groups—notably of academics, by c. 20%—thereby effectively 'rewarding' their indecision, regardless of whether degree of uncertainty it was motivated by proper caution (as appears to have been generally the case with academics) or other, less apparent, factors.

In sum, Wilson was able to reasonably infer a link between a comparatively large number of 'unsure' responses and expertise in the performance of academic phoneticians owing to their ability/willingness to explain their decisions. But, precisely because such detailed explanations were absent among the other groups in her study, nothing whatsoever could reasonably be said about the relationship between 'unsure' responses and these groups' respective expertise in the tasks. It seems at least possible, though, that at least in the case of the Ghanaian NSNL group Wilson may have found a relationship between *lack of* 'unsure' ratings and expertise (i.e. accuracy) had the distinction between 'unsure' and decision responses not been effaced by the decision to exclude the former response type from her analysis.

Wilson thus did not undertake an analysis of a potentially useful measure of confidence: groups' comparative tendency towards 'unsure' vs. 'decision' responses. This is understandable given that her experiments were oriented towards evaluating the relative aptitude of various types of listeners for LAAP as currently practiced. However, a comparison of this type is of considerable utility to the present series of experiments, with their ultimate aim of developing a perception test for asylum applicants. This is because it offers additional potential—beyond comparisons involving accuracy—for differentiating the response patterns of various types of listener. I return to this point in Chapter 12, when considering the range of response types that might be permitted in a supplementary perception test for LAAP.

The findings of three other previous studies designed to replicate LAAP-like conditions, Hedegard (2015), Shen & Watt (2015) and Muhammad (2021), contradict each other as to the interaction of confidence and accuracy among different types of listeners. Muhammad (2021),

in his study of the ability of speakers of Nigerian languages to identify Nigerian-language accents in English-language stimuli, found that native speaker *linguists* were "highly confident even when inaccurate" (2021, p. 247). His academic phonetician group, meanwhile, was overall considerably more circumspect in its judgements.

The latter finding on the relationship between accuracy and confidence replicates that of Wilson (2009). The former, however, shows that advanced linguistic training does not make listeners *in general* more cautious in their responses. In other words, linguistic training may engender circumspection only among non-native speakers. In this connection, recall the second of the three prospositions advanced in Chapter 4: linguistic education does not decisively alter native speaker perceptions of language. From this arises the implication that caution may be better attributed not to linguistic training but to non-native speakerhood, or perhaps to some combination of the two factors. Strengthening the plausibility of this inference, Muhammad's finding on the confidence of native-speaker linguists is strongly replicated in Chapter 11, where it is observed that Syrians trained to postgraduate level provided only 'decision' responses to Syrian stimuli—and these overwhelmingly at the maximum degree of (Likert scale) certainty.

Hedegard's study on Japanese resulted in the finding, largely contrary to Wilson (2009), that NSNLs placed in an unpaired condition were the least confident (as measured by 'uncertain' vs. decision responses) and the second most accurate, after the NSNL + non-native linguist condition, of her four groups. In addition, while Wilson's academic phoneticians performed as accurately as NSNLs (after exclusion of 'unsure' responses) but with a lower general degree of caution in their 'decision' responses, Hedegard's non-native linguists working alone both expressed lower confidence and achieved lower accuracy than all other conditions.

Shen & Watt, in their experiments on East Asian accented English, located a significant statistical correlation between accuracy and confidence in the correct responses of NSNL Chinese listeners (modelled as 'native speakers'). However, as discussed above, it is difficult to determine exactly who qualifies as a 'native speaker' when assessing stimuli spoken in a second language. This makes Shen & Watt's finding difficult to interpret.

As it is difficult to reconcile the results of these four studies, no predictions are made in this thesis as to the relative confidence of listener groups or the interaction between confidence and accuracy. However, the decision to avoid making formal predictions on this score does not preclude a statistical investigation of confidence, which is attempted in Chapters 8 and 11 of the thesis.

5.5. The 'familiarity effect'

The 'familiarity effect' refers to the postulate that listeners with a history of secondary residence in a given area tend to show a greater ability, compared to those without such a residential history, to identify speakers originating in the area of secondary residence (see e.g. Baker et al. 2009, Clopper & Pisoni 2006, Sumner & Samuel 2009).

Baker et al. (2009) found that the familiarity effect was especially significant in listeners who had spent at least five years in their area of secondary residence. Interest in investigating the effect informed the initial composition of listener groups, detailed in Chapter 7, wherein the question 'What is a local?' is investigated. There, in line with the finding of Baker et al., a cut-off of five years' residence in Yorkshire was adopted in composing the Yorkshire resident ('Yorks Res') group.

However—as with the relationship between accuracy and confidence—no predictions are made in the thesis as to the effect of familiarity on listener responses, for two reasons. Firstly and most importantly, studies investigating the effect are inconsistent in their findings. Baker (2009), focusing on the accent of Utah, found a significant advantage associated with secondary residence in Utah of more than one but under five years as well as a more significant advantage associated with more than five years residence. Sumner & Samuel (2009), too, found that exposure to NYC and Georgia speech resulted in enhanced speed in identifying stimuli featuring these varieties. Clopper & Pisoni (2006), in an experiment similar to that of Clopper (2004), located no significant effect of residential history on accuracy; however, they review several further studies of their own whose findings contradict this.

Secondly, the studies above tend to focus on explaining their findings via sociolinguistic constructs such as 'perceptual categories' and 'linguistic prestige'/'linguistic security'/'stigmatisation'. However, these purely explanatory categories are not of interest to this experimental series. The principal interest is instead in the 'how' and the 'what': *How* do different types of listener vary in their response patterns and on *what* parameters?

5.6. Summary and conclusion

In this chapter I have argued that, largely for reasons of methodology, the studies reviewed by Fraser (2009) are essentially irrelevant to assessing the accuracy of the 'one of us?' judgements essential to LAAP. Relevant research in real-world LAAP and LAAP-like experimental environments, however, demonstrates that NSNL analysts in primary-phase LAAP are likely to be capable judges when confronted with some kind of 'one of us?' task. This is taken as sufficient evidence that NSNL asylum applicants may be able to do likewise in the perception tests developed via the experimental work of this thesis.

In the next chapter I specify the three guiding hypotheses of the thesis. I combine in them the three general principles of native speaker perception laid out at the conclusion of Chapter 4 with corroborative findings as to the accuracy of NSNLs judgements reviewed in this chapter.

Chapter 6: Hypotheses

6.1. Introduction

In the two preceding chapters I conducted a critical review of the theoretical and empirical literature on native speaker perceptions, judgements and self-rated confidence. From this arise three hypotheses, which are tested via the experiments detailed in Chapters 7 to 10. Below I formally state the three hypotheses in turn. I then briefly describe their relation to the experiments and summarise the justification for each in the reviewed literature.

6.2. Specification and justification of hypotheses

Hypothesis (i)

'Local' speaker-listeners will show greater accuracy in recognising 'local' voice samples in comparison to speaker-listeners of other ('non-local') varieties.

Justification

Theory and empirical work in sociolinguistics, sociocultural evolution, social psychology and anti-reductionist philosophy converge on the conclusion that native speakers possess holistic, tacit knowledge of their own language variety that is unavailable to non-natives. This knowledge lends native speakers inimitable competence in identifying fellow native speakers by their use of language alone. The findings of specifically LAAP-related experimental work, as well as studies from forensic phonetics (the 'other-accent effect'), strongly suggest the same.

The primary task of all five experiments conducted for this thesis involves listeners judging whether or not audio samples feature a speaker of a specified language variety. The intention is to use the resulting response patterns to develop supplementary perception tests for LAAP.

There are two series of experiments, and five experiments in total. The first series, comprising two experiments, features English-language samples; the second series, composed of three

experiments, features Arabic. The term 'local' refers to differing levels of native-speakerhood in the two series. In the experiments on English it signifies speaker-listeners born and raised in the historical English county of Yorkshire, with samples featuring Yorkshire English as the corresponding 'target'. In the experiments concerning Arabic, it denotes those born and raised in Syria, with samples featuring Damascene Arabic as the 'target'.

Hypothesis (ii)

The predicted differences in accuracy will hold irrespective of whether speaker-listeners are trained in any branch of linguistics.

Justification

The justification here is predominantly negative. There is no unequivocal evidence in the reviewed literature to support the belief that linguistic training provides an advantage, either to native speakers of a particular variety or to non-natives, in identifying same-variety speakers. There is evidence that *phonetic* training assists listeners in describing language, especially at the segmental level (cf. Wilson 2009), but there is no evidence that it can function as a substitute for the holistic, tacit knowledge of native speaker-listeners.

Hypothesis (iii)

Analysis of feature citations will show that speaker-listeners vary by region of origin in the number and kinds of cues they recognise.

Justification

In the Arabic experiments only, listeners are also asked to specify the cues that informed their judgements. The ultimate aim is to classify the resulting responses so as to assist in the development of a supplementary production test for LAAP.

The tacit, holistic nature of native speaker knowledge, theorised in Chapter 4, suggests that Syrians will acknowledge fewer individual features than non-Syrians, who will be conversely more dependent on recognition of discrete features in attempting to identify speakers of a variety other than their own. Another influential factor here is Wilson's inference that Ghanaian NSNLs and non-Ghanaian academic phoneticians relied on different cues in making judgements (cf. Wilson 2009, Chapter 5). It thus seems plausible that differences between NSNLs and non-linguists may be reflected in listener comments in the Arabic experiments conducted for this thesis.

Having defined and justified the three hypotheses guiding the empirical component of the thesis, I turn in the next chapter to a description of the methodology and results of the first experiment in the English series.

Chapter 7: English experiment 1

7.1. Introduction

Chapters 7 to 11 describe the empirical work conducted for this thesis. Chapters 7 and 8 are on English, with Yorkshire English as the target variety, while Chapters 9 to 11 concern Arabic, with Syrian Arabic as the target. The experiments use foils from various places in, respectively, northern England and Arabic-speaking countries outside Syria. I note here that none of the surveys conducted for these experiments elicited listeners' age. This variable is therefore not explored in any of the following chapters.

As detailed below, in section 7.3, the overarching objective of both English experiments is to examine differences in group response patterns to the question, 'Is this a local (in this case Yorkshire) accent?' The basic wording of this question—retained, *mutatis mutandis*, in all experiments conducted for this thesis—was motivated by two considerations. The first is the fact that LAAP is in the first instance a verification task (cf. Chapters 1 to 3), in which a question very much like this has to be answered. Experimental results emerging from responses to the question are therefore of maximum applicability to assessing the validity of current primary-phase LAAP practice. The second consideration was simplicity: the question is brief and unambiguous, so that the risk of misinterpretation by listeners is minimal. Initial vetting of the question among academic staff and postgraduate students in forensic speech science at the University of York demonstrated satisfactorily that this was so.

This initial experiment (henceforth Eng-1) was a pilot study conducted in an exploratory spirit, and largely in the interests of developing a suitable design for the following experiment on English (herein Eng-2; see Chapter 8). Seven design features were under investigation.

The first four were stimulus-related: selection (i.e. the geographical origin of the speakers featured), duration, quantity, and use of a reference sample. The fifth was optimal use of a web-based platform. The sixth was the number of alternatives available for (Likert scale) rating. The seventh was the validity of the four listener groups essayed in Eng-1, these having been

devised with place of birth/current residence as determinative parameters. At issue here was whether, in the task of identifying Yorkshire accents, non Yorkshire-born but Yorkshire-resident listeners would cluster more closely with local or non-local listeners. As discussed in subsection 7.5.1, this question is closely linked to both the 'other-accent' effect and the 'familiarity effect'.

7.2. Eng-1: hypotheses

As set out in chapter 6, three hypotheses guide the experimental work of this thesis:

(i) *'Local' speaker-listeners will show greater accuracy in recognising 'local' voice samples in comparison to speaker-listeners of other ('non-local') varieties.*

In the English experiments 'local' is defined as people born and raised in Yorkshire. This hypothesis is the only one of the three examined in Eng-1.

(ii) This difference in accuracy will hold irrespective of whether speaker-listeners are trained in any branch of linguistics.

As explained in section 7.3, this hypothesis is not examined in Eng-1.

(iii) Analysis of feature citations will show that speaker-listeners vary by region of origin in the number and kinds of cues they recognise.

In the Arabic experimental series *only*, listeners are asked to specify the cues that informed their judgements of each stimulus as either Syrian or non-Syrian. For this reason, hypothesis (iii) is not examined in Eng-1.

7.3. Eng-1: research objectives

Hypotheses (i) and (ii) relate to what I refer from now on as research objective 1 of the thesis: examining differences in group response patterns to the question, 'Is this a local (in this case Yorkshire) accent?' Research objective 1 in turn bears directly on the ultimate aim of the thesis: developing novel supplementary tests of perception for asylum applicants. The kind of test

envisaged eventually to materialise from research objective 1 is an accent identification (perception) task to be performed by Arabic-speaking asylum applicants. However, only non-linguists were recruited as listeners for Eng-1, rendering exploration of hypothesis (ii) irrelevant (as mentioned above, in section 7.2). Research objective 1 is approached in this chapter largely as a means of refining the design of subsequent experiments, in line with the parameters set out in section 7.1, above. Results relevant to research objective 1 are discussed in connection with matters of design (see section 7.8).

What I refer to herein as research objective 2 involves investigating Wilson's (2009) conclusion that non-local linguists and local NSNL respondents appeared, to some extent at least, to rely on different cues in identifying stimuli with a given location (cf. Chapter 5). Research objective 2 thus emerges naturally from hypothesis (iii) and relates directly to the ultimate aim of developing a new test of production for asylum applicants. However, feature citations, i.e. the type of data required to investigate research objective 2, were considered germane only where Arabic speakers are concerned. In accordance with this assumption, feature citations are considered only in connection with the Arabic experimental series.

7.4. Methodology

7.4.1. Task design

Listeners in Eng-1 were presented with short voice samples and asked whether they were able to identify them as coming from Yorkshire. In Eng-1, these stimuli were presented in fixed order. This order of stimuli presentation obtained owing to my unfamiliarity at the time with the design features of the survey interface used (Qualtrics). This was remedied in subsequent experiments. Qualtrics was selected because it would enable responses by listeners from all over the world (some of whom reside in Australia).

As specified in Section 7.1., the question asked of listeners was, 'Is this a Yorkshire accent?' This question is substantively similar to Wilson's, 'Do you believe this person is speaking Ghanaian English?' (Wilson 2009). In Eng-1, three forced-choice alternatives were provided for responding

to this question: 'yes', 'no' and 'don't know'. For ease and speed of data analysis it was decided not to use a more graded Likert scale (i.e. five points or more) in Eng-1. This decision was reviewed for Eng-2 (see chapter 8).

There was some hesitation about including the 'Don't know' alternative, since it was initially thought that it would reduce the ecological validity of the experiment: in a real-life test asylum applicants could simply respond 'don't know' to each stimulus, thus rendering the prospective test useless. The 'Don't know' alternative was nevertheless included in Eng-1 for three reasons. All had to do with experiment design rather than ecological validity.

First, as stated in Ssection 7.1, was the fact that Eng-1 was conducted in order refine the composition of listener groups in subsequent experiments; the intention was not to use it as a prototype of the envisaged supplementary test. Second there was the imperative of reducing the operation of chance effects that may have resulted from having only 'yes' and 'no' as possible responses. Third was the attempt to gauge how decisively listeners would attribute '(non-)Yorkshireness' to each stimulus and how groups varied among each other in this respect. If groups did in fact vary in the number of 'Don't know' responses, confidence (in the form of differing percentages of 'Don't know' responses) may have proved useful as an additional measure—combined with ratio of correct and incorrect responses—to separate listener groups.

If listeners answered 'no' to the main question in each task, they were invited to speculate on the provenance of the stimulus, in a free-text box associated with the following question and placed after each stimulus:

If your answer above was 'No', what kind of accent do you think the speaker has?

Because they were of no direct interest to hypothesis (iii) and by extension research objective 2 (see Section 7.3), comments received in response to this question are not reviewed in detail in this chapter. They were, however, of interest in directing the design of Eng-2. Of particular import here are comments apparently referring to the General Northern English (GNE) character of many of the stimuli, a confounding factor which seems likely to have influenced results. For discussion of more remotely historical factors which may have affected listener

perceptions, see Appendix B.

GNE is spoken mainly in mid-northern England, including in Yorkshire and Lancashire. It is thought to have emerged from the relatively recent operation of dialect-levelling effects in the region (Trudgill, Hughes & Watt 2012). A characteristic of GNE is that it evidences few readily-identifiable local pronunciations (i.e. distinctive of any particular location in northern England) beyond the regional level. Several listeners made comments expressing difficulty in deciding whether some of the stimuli were of Yorkshire, a typical remark of this type being "It could be from Yorkshire but I can't be sure". The majority of stimuli subject to commentary of this type were omitted from Eng-2—from which, even so, a fairly large number of similar comments emerged.

Two further task design decisions were made, following Wilson (2009). Firstly, listeners were instructed that they were able to listen to each stimulus as many times as they wished. This was permitted because it was envisaged that in a real-life test multiple playbacks will be allowed, likely with some sort of limit imposed. However, it proved impossible to limit the number of repetitions in Qualtrics.

Secondly, listeners were allowed to leave and return to the survey at will. This condition was set in an attempt to maximise the completion rate, even though it is unlikely to be workable in a real-life LAAP test.

7.5. Listeners

7.5.1. Recruitment and categorisation

Listeners were recruited through contacts at the University of York and via personal networks. All self-declared as both non-linguists and native speakers of English (NSNLs). These two conditions—consistent with characteristics shared by the vast majority of asylum seekers—were set because the present series of experiments seeks primarily to investigate the perceptions of NSNLs, not those of listeners with training in linguistics.

There were 28 listeners whose responses could be used; 40 more were excluded on the grounds of incompleteness (see section 7.8. for further discussion) or provision of insufficient demographic information (e.g. county of origin/residence). Four categories of listener were composed according to nativeness to or history of residence in Yorkshire. Table 7.1 summarises these categories and the number of listeners assigned to each.

Table 7.1*:* Listener categories and number

Listener category	Yorks B&R	Yorks Res	Brit non-Yorks	Non-Brit	Total
Number	13	6	5	4	28

The following are the grounds on which the categories were composed.

(1) Thirteen listeners declared having been born and raised in Yorkshire (Yorks B&R). This type of listener was classified as equivalent to asylum applicants who claim to have been born and socialised without interruption in a particular area (the majority of cases). It was also intended to cover two further types of applicant: First, those who claim origins in a given area but who claim to have lived elsewhere subsequent to acquisition of their native dialect(s); second, those who claim to have been brought up by parents with origins in the given area.

It is known that, in the course of being interviewed by immigration agencies (e.g. Home Office 2018, p. 8), applicants are asked to nominate their native dialect(s). In this context it seems reasonable to assume that even applicants who have lived for a long time away from their claimed locus of socialisation would nominate at least one of the varieties they first learned to speak. In line with this assumption, any listener who claimed birth and upbringing in Yorkshire (or, in one instance, an upbringing elsewhere by Yorkshire-born parents) was included in this category, regardless of any intervening residential history they declared. Listeners were not, however, asked to nominate their own native dialect. The other-accent effect (see Chapter 5), as well as the theorised tacit knowledge unique to native speakers of a given variety (cf. Chapter 4), together led to the prediction that Yorks B&R listeners would perform more accurately than others in identifying fellow Yorkshire speakers.

(2) Six listeners reported that they had been resident in Yorkshire for at least five years but had been born outside Yorkshire (Yorks Res). Grouping this type of listener into a discrete category was motivated by the fact that some asylum seekers claim to have spent significant periods in locations other than that with which they identify their native variety (e.g. born and raised in Damascus until the age of four, and a speaker of the Damascus dialect, but then spent 15 years in Aleppo, or born and raised in Egypt but lived in Syria since the age of eight; many declared life-history variations are encountered in casework). In the LAAP context, administering a test to applicants who claim not to have been resident in their declared area of origin for an extended period may be unfair. A test centred on the variety spoken in their area of secondary residence might in such cases be the most reasonable alternative—thus the composition of this category of listeners, with a history of residence in, but not nativeness to, a secondary area. Considering that the 'familiarity effect' (see Chapter 5) has been found to obtain principally among speakers with a minimum of five years in their area of secondary residence (Baker 2009), a period of at least five years' residence in Yorkshire was selected as the minimum required for a listener to be included in the Yorks Res group.

(3) Five listeners identified as British citizens without any history of residence in Yorkshire (Brit non-Yorks). These were sorted into a category to represent genuine citizens of an asylum-eligible country who for one reason or another misreport their origins, who speak a closely related variety and who are relatively well exposed to the 'target' variety (e.g. the Jordan-Syria relationship, in the Arabic context).

(4) Four listeners reported being native English speakers and citizens of countries other than the United Kingdom, without any history of residence in the country (Non-Brit). This category was intended to mirror asylum seekers who falsely claim to be from an asylum-eligible country—for example, Syria—and who are in fact from a country elsewhere in the Arabic-speaking world, e.g. Morocco or Egypt.

The 'other-accent effect' (see Chapter 5) was expected to influence the accuracy of each of these listener groups, such that Yorks B&R would find it easiest and Non-Brit most difficult to identify Yorkshire stimuli. However, as discussed in section 7.7, the results of Eng-1 violated this

expectation.

7.5.2. Method used to allocate listeners to groups

To determine into which of the four groups they should be placed, each listener was first asked whether they had ever received formal academic training in linguistics. Since NSNLs were the survey's target group, Qualtrics' skip-logic function prevented progression further into the survey by listeners who answered 'yes' to this question.

Each listener was then asked the following five preliminary questions. All except one were multiple-choice questions (the third was open-text).

What is your residential background in the United Kingdom?

Were you born in Yorkshire and/or have you at some stage lived in Yorkshire for more than five years consecutively?

If you answered 'Yes' above and are not originally from the place in Yorkshire where you live now, how long have you lived in your present area of residence? (e.g. 'I was born in London but have lived in Leeds since I was five.')

Which part of Yorkshire are you from/have you mainly lived in?

If you are from the United Kingdom, which part are you originally from?

After answering these questions, listeners were asked to listen in turn to a reference sample and the 13 stimuli. After completing the tasks, listeners were invited to comment further on the tasks via an open text box. The question was phrased as follows.

In the box below, please enter any additional comments you wish to make about the tasks and/or your responses.

Screenshots of the layout of these questions can be found in Appendix A.

7.6. Stimuli

7.6.1. Stimuli: selection

Stimuli selection for Eng-1 this thesis was in many respects handled differently than in the three immediately relevant prior studies, Wilson (2009), Shen & Watt (2015) and Hedegard (2015). Whereas all three featured at least some spontaneous speech, stimuli used in Eng-1 featured read texts only. The disadvantages associated with read texts (e.g. risk of shifting to a style closer to the standard, no or virtually no occurrence of regionally marked grammatical or lexical items) were considered reasonable trade-offs for the time and effort spared in composing stimuli from scratch. In addition, read texts permit the exercise of greater control over the occurrence of features. Stimuli were c. 30 seconds in duration (i.e. longer than Shen & Watt's but shorter than Hedegard's or Wilson's); subsection 7.6.3. explains the motivations behind this decision.

Also in contrast to Wilson's and Hedegard's experiments, the decision was made not to brief listeners in advance or to direct listeners' attention towards an intentionally limited class of features (phonetic and intonational, respectively). The use of briefing materials in both cases fitted the broad aim of Wilson's and Hedegard's experiments: to compare the responses of linguists with those of non-linguists. The present series of experiments, however, sought primarily to compare the responses of locals with those of non-locals—all of them NSNLs—with the development of a new perception test founded on the local/non-local distinction as the ultimate aim.

Furthermore, Wilson was naturally unable to exhaustively list in her briefing materials all the phonetic features of Ghanaian English; comments by some of the academic phoneticians in Wilson's experiment alluded to this fact when expressing doubt about whether speakers matched completely the phonetic profile indicated in the materials (Foulkes & Wilson 2011). This may have had the effect of adding to the number of uncertain responses by listeners of this type. The desire to avoid a similar situation was another reason that briefing materials were not used in the present experimental series.

Finally, entirely different languages (or varieties of the same language) were featured in each series of experiments. In Eng-1 and Eng-2, British English-language stimuli were used (cf.

Ghanaian English in Wilson 2009), owing partly to the relative ease with which both speakers and listeners could be obtained and partly to my own status as a mother-tongue speaker. As specified in Section 7.1., Yorkshire stimuli were the 'targets' and Lancashire stimuli the 'foils'. In Ar-1, Ar-2 and Ar-3, Arabic—chosen because of its relevance to LAAP—was the language of featured in the stimuli. Syrian (Damascus) Arabic was the target, and various non-Syrian varieties acted as foils.

7.6.1.2. Motivation for selection of Yorkshire/Lancashire stimuli

Yorkshire stimuli were selected because it was thought that listeners familiar with them would be easy to recruit among personal contacts at the University of York and in the city itself. As stated above, the ultimate objective of the thesis is to develop a supplementary LAAP test of the applicant's accent recognition (i.e. perception) skills. For the purposes of such a test, group tendencies towards certain response types must be distinguishable from one another. Thus a vital principle informing selection of stimuli (including the duration specified in subsection 7.6.1.), and by extension their retention in or exclusion from later experiments, was maximising differences in performance among the listener groups.

Historical Lancashire stimuli were selected over the available alternatives because one of the aims of the experiment was to test how well NSNL listeners of various types are able to distinguish between closely related varieties—that is, not between northern English and southern English accents, for example, which seemed likely to be too easy a task. Lancashire and (especially west) Yorkshire English are historically closely related (Wells 1982). Yet a range of differences between the Yorkshire and Lancashire stimuli selected for this experiment, while in many cases subtle, seemed to my ear (and with the assistance of authorities on English dialects such as Wells 1982) to be sufficient to permit distinction by listeners, at least in some cases.

7.6.2. Stimuli: source

Thirteen stimuli, of six female and seven male speakers, reading either *Comma Gets a Cure* or *The Rainbow Passage*, were downloaded from [IDEA: International Dialects of English Archive](#).

Six stimuli were of Yorkshire speakers, and seven of Lancashire speakers. The 13 stimuli—a sufficient number to produce differences in group accuracy but not so many that the task would be too long—and the reference sample were posted on Soundcloud, in .wav format. These were then linked from Qualtrics, the platform on which listeners took the survey.

All stimuli were volume-normalised and noise reduced, with Audacity editing software, to standardise them as much as possible. Recordings were not of uniform quality. However, initial vetting among academic linguists raised no objections on this score, and they were thus considered adequate for use in the English experimental series.

The geographical origins and names of stimuli, with reference data from the IDEA database, are specified in Table 7.2. All of the IDEA stimuli used here were recorded in the early 2000s. Stimuli names refer to the location in which each speaker has spent the majority of his/her life. The corresponding geographical locations are shown in Figure 7.1, a map of (historical) western Yorkshire and eastern Lancashire.

Table 7.2: Geographical origins and naming of stimuli

Geographical origin/name of stimulus	Description in IDEA
Greetland (Yorkshire)	England 55 female, 91, 1912, Greetland and Calderdale (West Yorkshire)
Leeds (Yorkshire)	England 81 male, teens, 1989, white, Dewsbury and Leeds (West Yorkshire)
Stainland (Yorkshire)	England 56 male, 48, 1955, white, Stainland (West Yorkshire)
Halifax (Yorkshire)	England 57 male, 52, 1951, white, Leeds and Halifax (West Yorkshire)
Skipton (Yorkshire)	England 83 male, 27, 1982, white, Skipton (North Yorkshire)
Harrogate (Yorkshire)	England 74 female, 30, 1976, white, Harrogate (North Yorkshire)
Manchester1 (Lancashire)	England 10 female, 20s, 1980s, white, Manchester
Manchester2 (Lancashire)	England 92 female, 78, 1933, white, Manchester
Manchester3 (Lancashire)	England 71 male, 39, 1968, white, Manchester
Bolton (Lancashire)	England 9 female, 30s, 1968, white, Edgeworth and Bolton (Lancashire)
Salford (Lancashire)	England 54 female, late 40s, 1950s, white, Salford (Lancashire)
Wigan 1 (Lancashire)	England 15 male, 34, 1966, white, Wigan (Lancashire)
Wigan 2 (Lancashire)	England 60 male, 59, 1946, white, Wigan (Lancashire)

Figure 7.1*: Map of stimuli locations (Yorkshire in blue, Lancashire in purple; approximate line of historical Yorkshire-Lancashire border in orange)
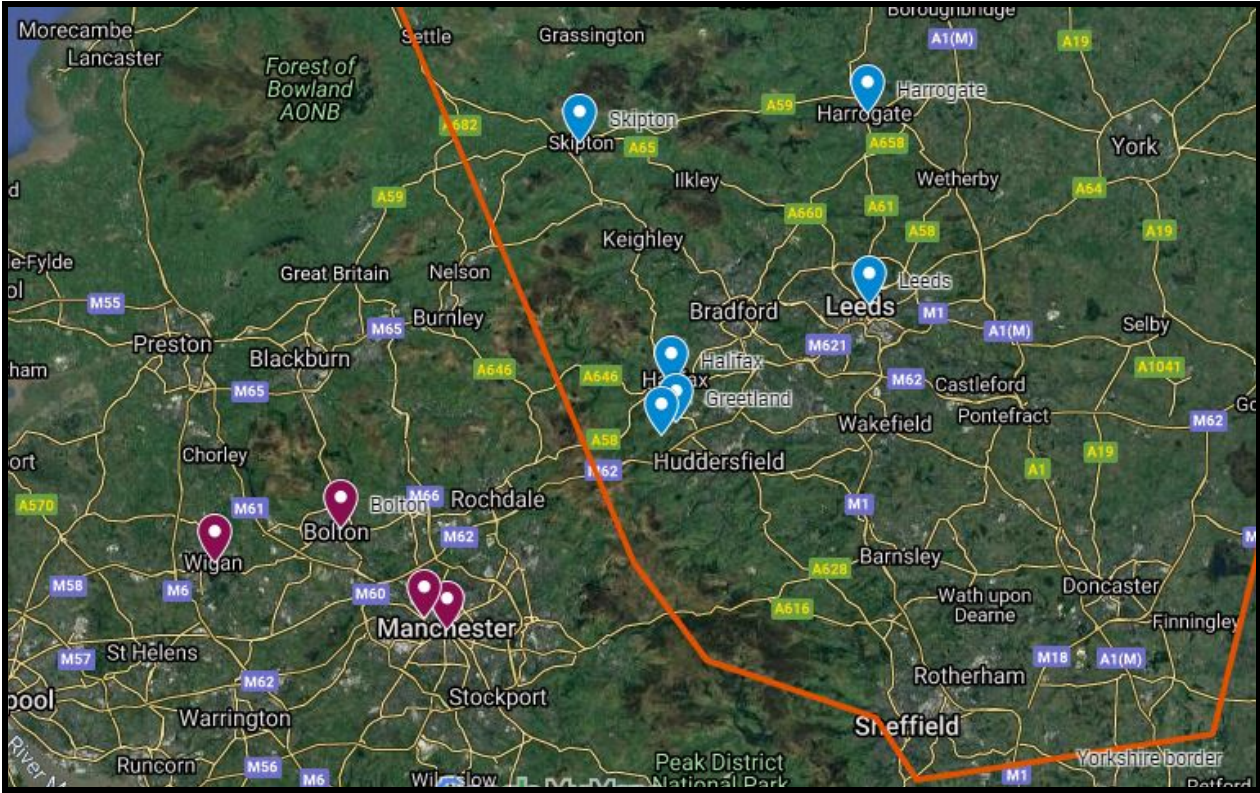
7.6.3. Stimuli: duration

Stimuli were edited to a duration of between 28 and 31 seconds. As mentioned in 7.6.1, Wilson (2009) and Hedegard (2015) used considerably longer and Shen & Watt (2015) rather shorter stimuli than 30 seconds, yet the stimulus duration in none of these studies approached those which are common in LAAP. Although it is true that a duration of c. 30 seconds is likewise incommensurate with that encountered in LAAP, which is commonly conducted on stimuli of at least 15 minutes (Hoskin 2018; Hubbuch 2019), stimuli of this duration were used for four reasons.

The first is derived from Yarmey (2012), who states that an exposure time as brief as two seconds has been found sufficient for the identification of individual speakers. He explains, however, that longer samples generally increase accuracy—but only up to a point. Stimuli of 18

seconds to eight minutes have been found to result in a greater number of false alarms. In Eng-1, a duration of c. 30 seconds was considered adequate to expose listeners to a range of diagnostic features without markedly increasing the likelihood of false alarms.

Second, as discussed in Section 7.2., the ultimate objective of the present project is to develop new, supplementary tests for LAAP, not to redesign extant tests. This being the case, there was no necessity to approximate current LAAP practice.

Third was my recognition of the fact that in an anonymous online survey it is vital to secure as many completed responses as possible while providing listeners with a reasonable range of potentially diagnostic features and avoiding what Hedegard (2015) calls "listener fatigue".

Fourth was the inference that the tacit, holistic knowledge of a NSNL local listener should make accent recognition possible from exposure to a stimulus appreciably briefer than those common in LAAP. As Nolan (2012) suggests, locals may have an enhanced ability to perform accent recognition on the basis of rapid apprehension of suprasegmental information which is not apparent to others.

7.6.4. Stimuli: text

The section of text included in the stimuli, for *Comma Gets a Cure* and *The Rainbow Passage* respectively, was as below:

> *Well, here's a story for you: Sarah Perry was a veterinary nurse who had been working daily at an old zoo in a deserted district of the territory, so she was very happy to start a new job at a superb private practice in north square near the Duke Street Tower. That area was much nearer for her and more to her liking. Even so, on her first morning, she felt stressed. She ate a bowl of porridge, checked herself in the mirror and washed her face in a hurry.*

(The last sentence was omitted from one of the Yorkshire stimuli, Greetland, due to the speaker's slower reading rate).

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colours. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no-one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Two Lancashire stimuli (Manchester 1 and Bolton) featured readings of *The Rainbow Passage*; the remainder were readings of *Comma Gets a Cure*. Stimuli featuring *The Rainbow Passage* were used because there was, for the purposes of Yorkshire-Lancashire balance, an insufficient number of Lancashire stimuli featuring readings of *Comma Gets a Cure*.

7.6.5. Stimuli: discussion of features included

7.6.5.1. Yorkshire stimuli

As shown in Table 7.2 and Figure 7.1, six stimuli were of Yorkshire English. These featured speakers from Halifax, Leeds, Skipton, Harrogate, Stainland and Greetland. The first four are urban centres; the final two are (semi-)rural areas in Calderdale, between Halifax and Huddersfield.

Only the Halifax, Greetland and Stainland stimuli feature typical, though not in all instances exclusively, Yorkshire features. All of these three stimuli include—in the words *happy*, *daily, territory* and *hurry*—several tokens of what Wells (1982) calls 'untensed happY', a feature which is typical of Yorkshire speakers (outside Hull) but occurs also in historical Lancashire (Beal 2010, p. 18; Baranowski & Turton 2015, p. 296). In these three stimuli, tokens of this final vowel show intra- and inter-speaker variability but are in all instances monophthongal and with phonetic realisations around [e] or [ɛ].

The speakers in the Halifax, Greetland and Stainland stimuli also use the monophthongal FACE vowel in their pronunciations of *face, plain* and *daily*. This is infrequent in Greater Manchester, where it is usually a narrow diphthong, [ei] (Hughes, Trudgill & Watt 2012, p. 117; Baranowski

&Turton 2015, p. 295), though in the rest of historical Lancashire it is predominantly monophthongal (Hughes, Trudgill & Watt, p. 150). Tokens of this vowel were considered partially diagnostic within the total range of stimuli because none of the seven historical Lancashire stimuli was from an area outside the administrative district of Greater Manchester, and only three were from outside it as defined linguistically by Baranowski & Turton (2015) (Figure 7.2).

These three stimuli also feature monophthongal GOAT (in *so*), which is a diphthong in the Lancashire stimuli. There is in addition the pronunciation of the GOOSE vowel in *zoo* (and *you*, in the Greetland and Stainland stimuli) as, approximately, [ɛʊ]. This is not present in any of the Lancashire stimuli, where it is monophthongal. Finally, the Halifax stimulus featured a traditional northernism (commonly associated by laypeople with Yorkshire, though also present in Lancashire) in the domain of grammar: 2nd person singular *were*. It was thought that the constellation of these phonetic and grammatical features might be sufficient to mark the 'target' stimuli as of (West) Yorkshire—readily so for local listeners and less so for others.

The remaining three Yorkshire stimuli (Leeds, Harrogate and Skipton) feature younger, urban speakers, whose accents incline more towards GNE than do those of the Halifax, Greetland and Stainland speakers. However, rather than attempting to prejudge how listeners would perform, these less distinctively Yorkshire stimuli were worth including as 'targets' in order to gauge the range of listeners' discriminatory abilities. Especially relevant in this connection is Nolan's (2012) suggestion that there are likely regionally diagnostic aspects of speech above the segment which are resistant to linguistic description. In addition, the Harrogate and Skipton stimuli were the only samples of non West Yorkshire speech available in the *IDEA* database. They were used even though they included few (stereo)typically northern features, because it was anticipated that some listeners would be local to North Yorkshire.

As mentioned above, the Leeds stimulus includes tokens of lax happY and monophthongal FACE: variably lax happY (in *happy*, though not in *daily*) and FACE (in *face* and *plain*). In this stimulus monophthongal GOAT also occurs (in *so*). The Leeds and Harrogate stimuli exemplify several features which occur in the accents of both Yorkshire and Lancashire, as well as GNE:

'unsplit' STRUT-FOOT (for example in *much*) and front short [a] realisation of BATH (Wells 1982, p. 353).

7.6.5.2. Lancashire stimuli

Seven stimuli, three female and four male, were of (historical) Lancashire speakers, including present-day Greater Manchester but excluding Merseyside. Figure 7.2 shows how Baranowski & Turton (2015, p. 295) define 'linguistic' Greater Manchester. As shown in Figure 7.1, the Lancashire speakers originate in Manchester, Salford (linguistically Greater Manchester), Wigan and Bolton (both administratively Greater Manchester).

Figure 7.2: Greater Manchester according to Baranowski & Turton (2015); the linguistic area is bounded by the M60.

Three segmental features were relied upon to help listeners identify (i.e. reject) Lancashire speakers. The first was initial /t-/. Foulkes & Hughes (in press) and Turton (pers. comm.) note that affricated /t-/ is common in Manchester; descriptions of similar realisations elsewhere in

historical Lancashire are encountered in dialectological accounts—for example, in Bolton (Shorrocks 1998). Each of the Lancashire stimuli (excluding the two readings of the *Rainbow Passage*) contained two tokens of this realisation, one each in *territory* and *tower*. By contrast, non-affricated, possibly laminal realisations of initial /t-/ are impressionistically common in traditional Yorkshire speech (Turton, pers. comm.). I was unable, however, to locate any published literature on the subject.

The second feature was diphthongal GOAT, which occurs in all seven Lancashire stimuli but is pronounced as a monophthong in four of the six Yorkshire stimuli. Baranowski (2017) reports that (diphthongal) GOAT-fronting is a common though relatively new phenomenon in Greater Manchester.

The third feature was the SQUARE-NURSE merger, present (in *square*) in Manchester 2 as well as in Wigan 1 and 2. This realisation, as [ə:], is reported as typical of parts of NW England but not of most of Yorkshire (Barras 2006; Barras, Honeybone & Trousdale 2007; Watson & Clark 2013; Wells 1982).

In five of the Lancashire stimuli there were features found both in northern dialects generally and in GNE—for example, unsplit STRUT-FOOT in *much* and unsplit TRAP-BATH in *trap.* However, the lax happY vowel, which occurred in three of the Yorkshire stimuli, was missing from all of the Lancashire stimuli (where the vowel was either a diphthong, [ei] or a tense monophthong, [i]). According to Wells, tense happY separates Manchester from much of Yorkshire (1982, p. 362).

The accents featured in four of the seven Lancashire stimuli (Manchester 1 & 2, Bolton and Salford) were to my ear close to GNE. One of the stimuli, Manchester 3, evidenced no distinctively northern features other than unsplit STRUT-FOOT in *much*. It too was therefore included as a presumptively obvious foil. However, it was impossible to know whether listeners would hear them the same way unless they were included in this (pilot) experiment.

As in the case of the Yorkshire stimuli, the majority of the Lancashire stimuli available in the *IDEA* database were included in the survey. This was done with a view to measuring how

accurately listeners were able to identify each of them and so to establish a range of difficulty informing stimulus selection for Eng-2. At the very least, the seven Lancashire stimuli were thought to be clearly northern but not (necessarily) of Yorkshire

7.6.5.3. Reference sample: description and motivation for inclusion

It was decided to include a reference sample in order to help inform the judgements of those, especially among the Non-Brit group, who may have had limited exposure to Yorkshire accents. The reference sample was downloaded from the [IViE Corpus](#) (Grabe, Post & Nolan 2001). It was 54 seconds in duration and normalised/noise reduced using the same editing software and parameters applied to the stimuli proper. It featured a young, male speaker from Leeds reading the *Cinderella* story. This particular sample was chosen for two reasons: (1) its ready accessibility; (2) the fact that as a reading of a third text it might serve not to further accentuate possible biasing effects resulting from the use of two different texts in the stimuli. The reference sample featured an accent which was criticised by specialists in British sociophonetics, after the commencement of response collection, on the grounds that it was either too close to GNE or redolent of a particular sociolect referred to colloquially by one informant as 'Leeds posh boy'.

7.7. Results

7.7.1. Organisation

This section provides an account of the results of Eng-1, indicating how and to what extent the response patterns of the various groups differed when confronted with the question, 'Is this a Yorkshire accent?' Owing to the small number of respondents and the fact that Eng-1 was a pilot, results are presented only in descriptive (and not inferential) form.

Group response patterns to Yorkshire and Lancashire stimuli, respectively, are compared in subsection 7.7.2. This is followed, in 7.7.3., by a comparison of group response patterns to all stimuli combined. In 7.7.4., response patterns by group to all individual stimuli are detailed in

turn. Finally, in Section 7.8., I discuss the design decisions, arrived at on the basis of these results, for Eng-2.

A brief note is required here to assist interpretation of the results below. Where Yorkshire stimuli are concerned a correct response means 'correctly accepting' as from Yorkshire the stimulus in question—that is, the response must be 'Yes'. The term 'false rejection' is used herein to describe an incorrect response ('No') to Yorkshire stimuli.

In connection with Lancashire stimuli, a correct response means 'correctly rejecting' as from Yorkshire the stimulus in question, meaning that the response must be 'No'. The term 'false acceptance' is used below to describe an incorrect response ('Yes') to Lancashire stimuli.

This scheme is shown in Table 7.3.

Table 7.3: Designation of 'yes' and 'no' responses by stimulus type

	'Yes' response (i.e. this is a Yorkshire accent)	**'No' response (i.e. this is not a Yorkshire accent)**
Yorkshire	Correct acceptance	False rejection
Lancashire	False acceptance	Correct rejection

7.7.2. Results by stimulus type: Yorkshire vs. Lancashire

Figure 7.3 shows the percentage of correct, incorrect and uncertain responses per group to Yorkshire and Lancashire stimuli, respectively.

Figure 7.3: Percentage per group of correct, incorrect and uncertain responses: Yorkshire (YORKS) vs. Lancashire (LANCS) stimuli
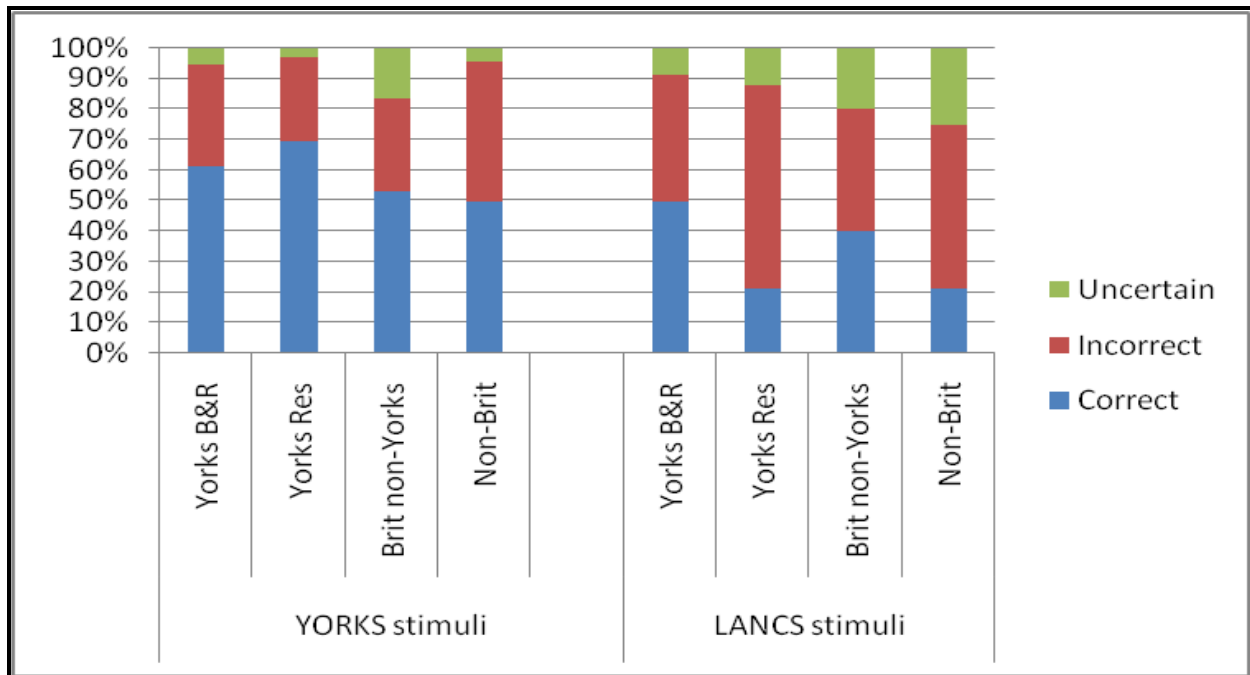
On Yorkshire stimuli, Yorks Res achieved the highest percentage of both correct acceptances (69%) and the lowest of false rejections (28%), making it the most accurate group. Yorks B&R achieved the next-highest percentage of correct acceptances (62%) but the second highest of false rejections (33%). Brit non-Yorks and Non-Brit performed at around chance on correct acceptances, with the former considerably more accurate as measured by false rejections (30% vs. 46%).

On Lancashire stimuli, the performance of Yorks Res on Yorkshire stimuli was almost a mirror image of its performance on Non-Yorkshire stimuli: 66% false acceptances and 21% correct rejections, making it the least accurate group on stimuli of this type. This is reflected in the group's overall performance, as shown below in Figure 7.4.

Yorks B&R performed best of the four groups on correct rejections, but at only just above chance level (51%), while entering 42% false acceptances. Brit non-Yorks entered 40% of both correct rejections and false acceptances. Non-Brit achieved the equal lowest percentage of correct rejections (21%) and the second highest number of false acceptances (54%). next-highest percentage of correct acceptances (62%) and the second-lowest percentage of false acceptances (51%).

All four groups entered a larger proportion of 'Don't know' (herein referred to as 'uncertain') responses to Lancashire than to Yorkshire stimuli: Yorks B&R, 9% vs. 5%; Yorks Res 12% vs. 3%; Brit non-Yorks 20% vs. 17%; Non-Brit 25% vs. 4%. Notably, the disparity between uncertain/other response types to Lancashire vs. Yorkshire stimuli was greater for Non-Brit than for any of the British groups. Both Yorkshire-affiliated groups returned fewer uncertain responses to both Lancashire and (especially) Yorkshire stimuli than either of the other two groups.

7.7.3. All stimuli combined

Figure 7.4 shows the total percentage of correct (acceptance/rejection), incorrect (acceptance/rejection) and uncertain responses, by each listener group, to all stimuli combined.

Figure 7.4: Percentage per group of correct, incorrect and uncertain responses to all stimuli
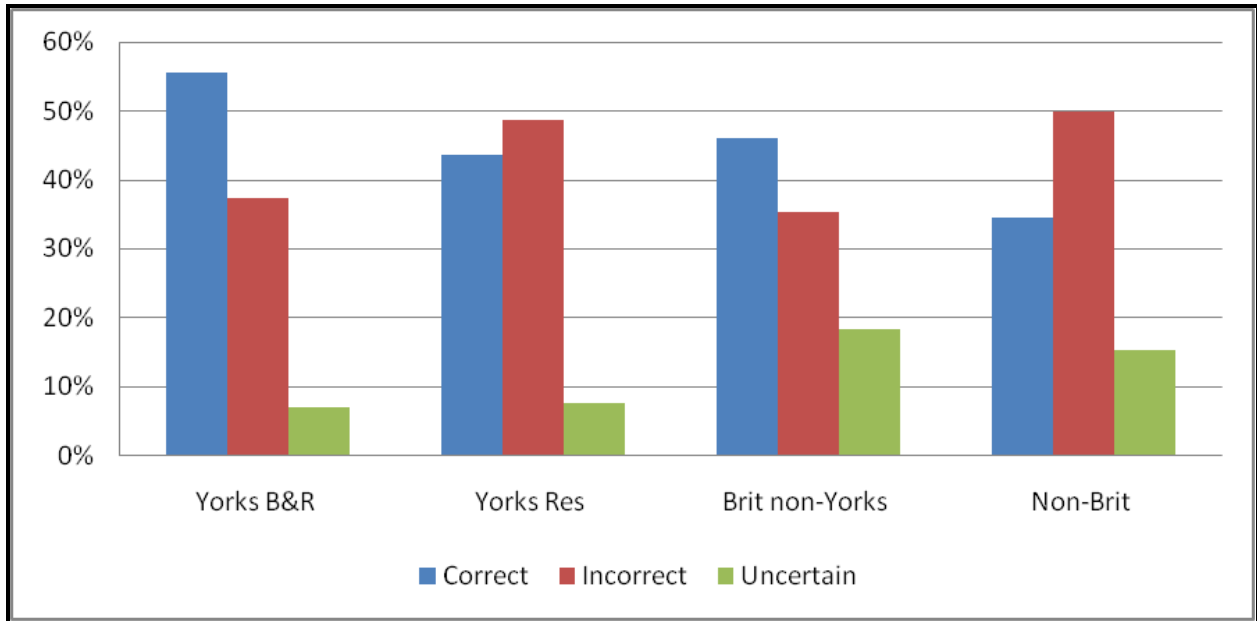
In total proportion of correct responses Yorks B&R was most accurate (56%). Brit non-Yorks (46%) was slightly more accurate than Yorks Res (44%). Non-Brit was the least accurate group (35%).

Yorks Res showed the second highest number of incorrect responses (49%) while Brit non-Yorks recorded the lowest (35%), thus outperforming Yorks B&R (37%). Non-Brit scored highest on this measure (50%).

Brit non-Yorks was the group most inclined to select the uncertain response, at 18%, and Non-Brit the second most (15%). The two Yorkshire-affiliated groups were relatively more certain in their responses (Yorks B&R, 7% uncertain, Yorks Res, 8% uncertain).

7.7.4. Group response patterns to individual stimuli

As noted in section 7.1 and elsewhere, Eng-1 was conducted partly to determine retention, exclusion or replacement of stimuli for Eng-2. Examination of group response patterns to individual stimuli was essential to making these decisions. Figures 7.5 to 7.8 show the proportion of correct, incorrect and unsure responses to individual stimuli, according to listener group.

Figure 7.5 shows that Yorks B&R achieved a large percentage of correct acceptances on four Yorkshire stimuli (Leeds, Halifax, Greetland and Stainland), while two (Skipton and Harrogate) were overwhelmingly subject to false rejections. Of the Lancashire stimuli, Manchester 3 and Wigan 1 attracted a proportion of correct rejections greater than 60%. Salford and Wigan 2 prompted correct rejections at slightly higher than chance level.

Figure 7.5: Yorks B&R: correct, incorrect and uncertain responses to individual stimuli
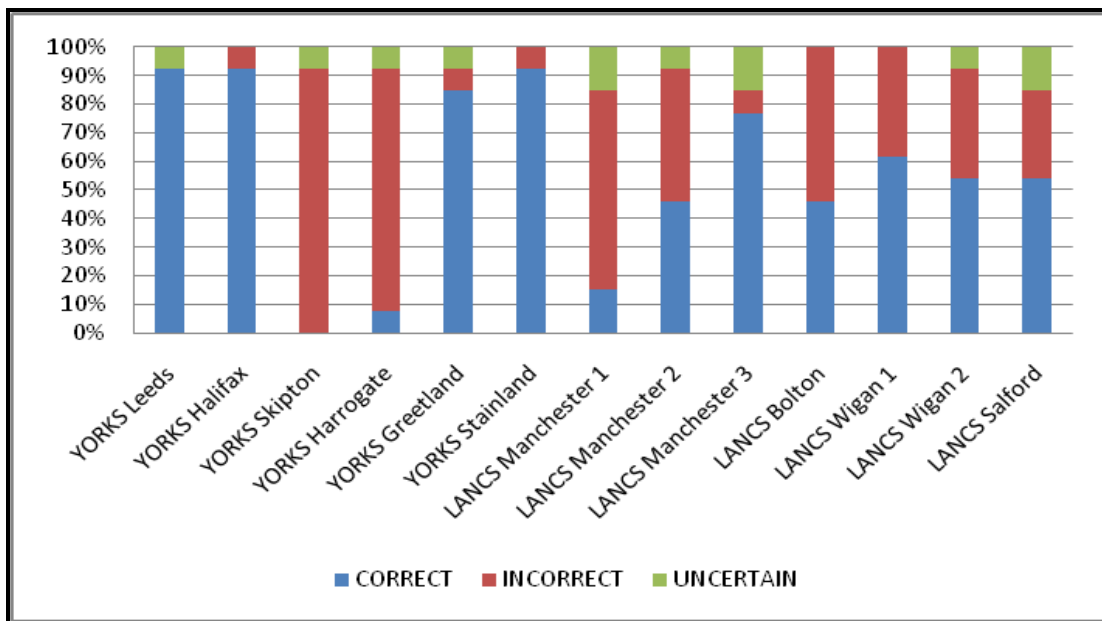
Figure 7.6 shows for the Yorks Res group a similar (though still more accurate) pattern to that of Yorks B&R—but only on Yorkshire stimuli. Leeds, Halifax, Greetland and Stainland each occasioned no false rejections; accuracy on Harrogate and Skipton was similar to that of Yorks B&R. Contrary to the Yorks B&R pattern, no Lancashire stimulus was correctly rejected by Yorks Res at above chance level. Differences to the Yorks B&R pattern on Lancashire stimuli are particularly noticeable in responses to Wigan 1 (no correct rejections vs. 62%) and Salford (33% vs. 54%).

Figure 7.6: Yorks Res: correct, incorrect and uncertain responses to individual stimuli
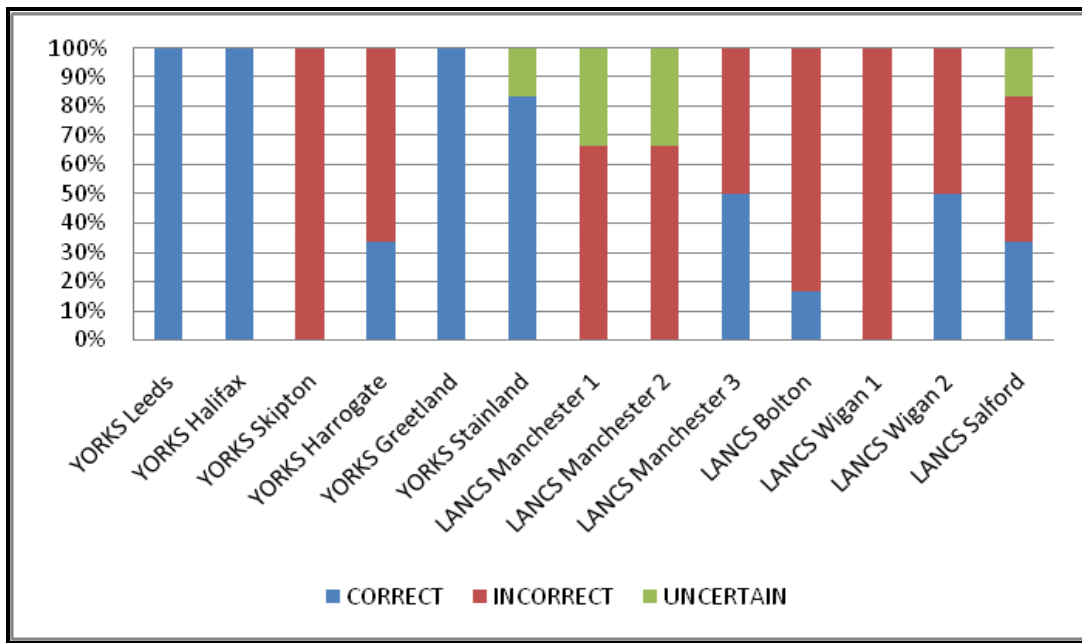
As Figure 7.7 indicates, the Brit-Non Yorks group achieved a large number of correct acceptances of three of the same Yorkshire stimuli (100% on Greetland and Stainland, 80% on Halifax) as did other two British groups. On the Leeds stimulus, however, Brit non-Yorks were much less accurate (40%) than either of the Yorkshire-affiliated groups. The Harrogate and Skipton stimuli attracted only false rejections and uncertain responses.

Brit non-Yorks correctly rejected three Lancashire stimuli (Manchester 2 & 3 and Wigan 2) at 60%. On two further Lancashire stimuli (Salford and Bolton) the proportion was only 40%. Manchester 1 proved similarly difficult for Brit non-Yorks (80% false acceptances) as it had for the Yorkshire-affiliated groups. In contrast with Yorks B&R but similar to Yorks Res, accuracy on Wigan 1 and Salford was well below chance level.

Figure 7.7: Brit non-Yorks: correct, incorrect and uncertain responses to individual stimuli
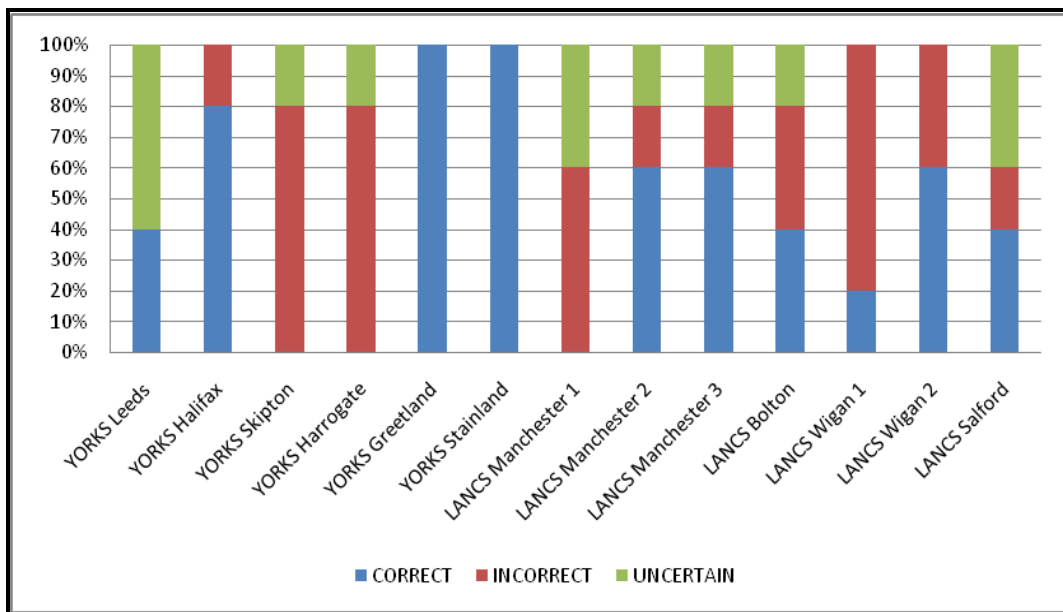
Figure 7.8 demonstrates that the Non-Brit group achieved a hit-rate of over 50% only on Halifax and Greetland. Interestingly, the percentage of correct acceptances of Skipton (50%) was much greater than it was for any other group. Non-Brit also reached 50% (correct acceptances and correct rejections respectively) on both Stainland and Manchester 2. Responses to Manchester 1 pattern broadly with those of the other three groups.

Figure 7.8: Non-Brit: correct, incorrect and uncertain responses to individual stimuli
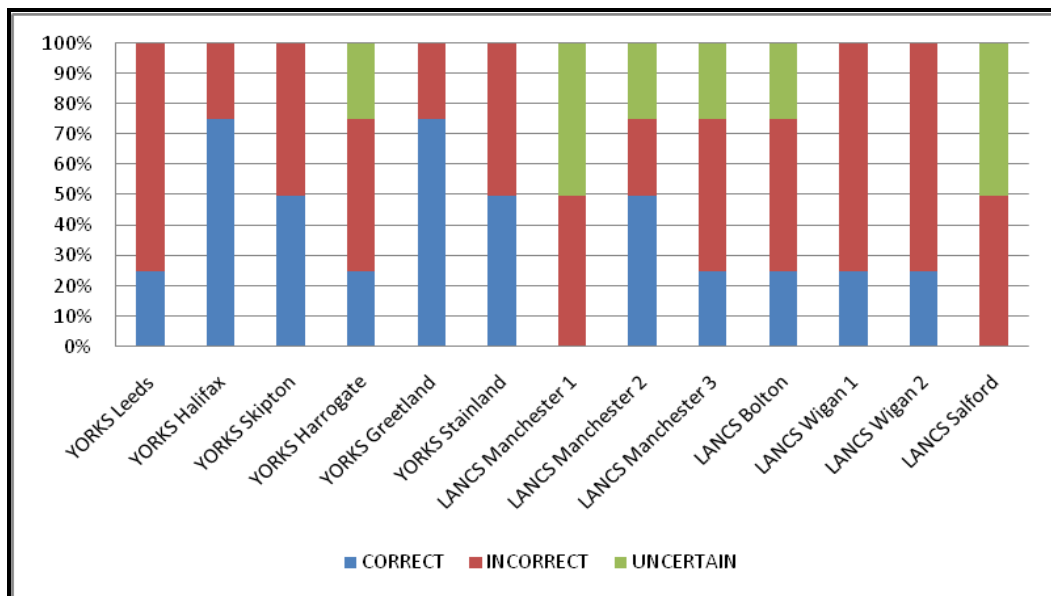
7.8. Summary and discussion: design features trialled for use in Eng-2

7.8.1. Composition of listener groups

As stated in Section 7.1., one of the design features investigated in Eng-1 was the composition of listener groups for Eng-2. The following discussion of results illuminates how decisions were reached on this score.

The raw data in Figure 7.3 shows Yorks Res to be the most accurate of the four groups on Yorkshire stimuli. This result suggests that familiarity with Yorkshire accents, achieved through sustained residency in Yorkshire, has a positive effect on listener accuracy, as the familiarity effect would suggest (see subsection 7.5.1). This appears to violate the supposition of hypothesis (i), that the tacit knowledge of local listeners and the other-accent effect—rather

than the familiarity effect, about which no predictions were made—would condition accuracy on Yorkshire stimuli.

However, the two non-Yorkshire-affiliated groups, Brit Non-Yorks and non-Brit, each performed broadly in the direction suggested by the other-accent effect. For this reason they were retained as discrete listener categories. In further (partial) confirmation of the other-accent effect, Yorks B&R performed more accurately on Yorkshire stimuli than either of the non-British groups and for this reason was likewise retained as a distinct listener category for Eng-2.

The performance of Yorks Res suggested that the operation of the familiarity effect would be worthy of further investigation in Eng-2. With this aim in mind, Yorks Res was retained as a separate listener category for Eng-2 in order to further explore, via a larger data set, whether Yorks B&R and Yorks Res listeners would differ in their response patterns.

7.8.2. Further design features: web interface and stimuli properties

Five further design features were trialled in Eng-1. The first, use of the web-based platform Qualtrics, can be dealt with briefly: no difficulties were identified with the Qualtrics interface, and it was retained for Eng-2.

Two other design features had to do with quantitative properties of the stimuli: their duration and overall number. As mentioned in subsection 7.5.1., a relatively small proportion of responses (28 of 68) turned out to be usable in Eng-1. Some were excluded because they did not specify the listener's geographical/residential background, others because the listener had failed to complete the task. On the basis of listener comments, I was able to attribute the low completion rate partly to the duration of the stimuli and partly to the number of stimuli used. On average, listeners had taken 17 minutes to complete the survey, with a range between 50 and 6 minutes. Stimuli duration and number were therefore both reduced for Eng-2, in an attempt to increase the completion rate.

Another design element under evaluation was the number of alternatives available for Likert-scale rating. In order to obtain a more detailed view of listener judgements, the number of points on the scale was increased, from three in Eng-1 to five in Eng-2 (cf. discussion in Chapter

8). A further feature—the reference sample—was also discarded for Eng-2, because of its dubious representativeness (cf. subsection 7.6.5.3).

The final factor of design for Eng-2, the origins of speakers featured in the stimuli, was determined via examination of differences in group responses to individual stimuli (cf. 7.7.4). There were two main criteria determining retention or omission of stimuli.

First was the discriminability of a given stimuli: if it occasioned chance-level or below accuracy across all groups, it was deemed too difficult for the task and thus omitted (Harrogate, Skipton, Manchester 1, Bolton). The Harrogate and Skipton stimuli had proven especially difficult for British listeners. Comments revealed that many British listeners had identified them with southern England ('London' and 'south(ern)' were frequent guesses). Non-British listeners left no comments in connection with these two stimuli to illuminate their relatively more accurate performance in identifying them (cf. chance-level for Skipton in Figure 7.8.). Conversely, stimuli which occasioned accuracy across groups at chance level or above were retained (Leeds, Halifax, Greetland, Stainland, Manchester 3). Manchester 1 and Bolton (the only readings of *The Rainbow Passage*) were discarded for the additional reason that doing so would achieve uniformity of text.

The second criterion had to do with the objective of accentuating differences in performance among the four groups. The retention of Wigan 1 and Salford was motivated by the relatively accurate performance on them by Yorks B&R, even though each had provoked accuracy among other groups at well below chance level.

Two stimuli were omitted for reasons somewhat outside the above criteria. Manchester 2, in spite of its having occasioned above-chance level accuracy by both Yorks B&R and Brit non-Yorks, was omitted because of its overwhelmingly GNE character. Two other stimuli retained (Leeds and Salford) also seemed to evince a strong GNE influence, and it was decided that featuring too many speakers of this type would unbalance the range of stimuli, perhaps making some too easy and some too difficult to identify. Although all three British groups had performed at chance-level or above on Wigan 2, it was omitted because the range of intergroup performance was relatively low (50% to 60%) and in all cases just above chance.

The above process of retention/omission left four Yorkshire (Greetland, Stainland, Leeds and Halifax) and three Lancashire stimuli (Manchester 3, Wigan 1 and Salford (all within Greater Manchester) for use in Eng-2. In the following chapter I detail the motivations for inclusion in Eng-2 of three additional stimuli.

A point worth noting beyond the aim of design refinement is the demonstration in Figure 7.3 that all groups entered an appreciably larger percentage of false acceptances than false rejections. This is especially noticeable in the case of Yorks Res, whose percentage of the former type of error was exactly double its percentage of the latter. This trend might be ascribed to the number of more or less GNE-type stimuli used here, which are difficult to identify with any particular county. In any case, this result accords with the finding of Wilson (2009): all listeners in her experiment were also more inclined toward false acceptances than to false rejections. As we shall see, group responses in Eng-2 (cf. Chapter 8) generally conform to this pattern; however, the trend is violated in the response patterns of Arabic speaker-listeners in Ar-1 and Ar-3, where only local (i.e. Syrian) listeners show a similar tendency.

Chapter 8: English experiment 2

8.1. Introduction

Following Eng-1, the plan had been to run four further pilot studies, varying mainly in the duration of stimuli. The onset in March 2020 of COVID-19 and the associated restrictions, which continued in full force for approximately 14 months, forced a reconsideration of this plan. In order to best make use of the time remaining, the decision was made to run a single further experiment on English and to proceed from it directly to the final series of experiments, on Arabic.

The design selected for the present experiment, Eng-2, was the most challenging for listeners of the four originally planned. It deploys stimuli of c. 10 seconds (cf. 30 seconds in Eng-1). This stimulus duration was selected partly to increase task completion (cf. only 28 of 68 in Eng-1) and partly to maximise differences in the response patterns of the various categories of listener.

8.2. Eng-2: hypotheses

As discussed in Chapter 7, section 7.1., in Eng-1, only hypothesis (i) was relevant. In Eng-2, the validity of hypotheses (i) and (ii) are examined. I explain why this is so in Section 8.3. Hypotheses (i) and (ii) are as follows.

(i) 'Local' speaker-listeners will show greater accuracy in recognising 'local' voice samples in comparison to speaker-listeners of other ('non-local') varieties.

(ii) This difference in accuracy will hold irrespective of whether speaker-listeners are trained in any branch of linguistics.

8.3. Eng-2: Research objectives

Research objective 1 of Eng-2 is to investigate the validity of hypotheses (i) and (ii), above. This involves presenting audio stimuli featuring Yorkshire and non-Yorkshire English speakers to seven groups of listeners and analysing their respective response patterns to the question, 'Is this a Yorkshire accent?'

In Eng-1 (Chapter 7) there were four groups of NSNL listeners: Yorks B&R, Yorks Res, Brit non-Yorks and Non-Brit. In Eng-2, however, each of the first three groups is additionally stratified into a phonetically-trained and non-trained group. This allows investigation of the effect of education—and therefore hypothesis (ii)—where such had been impracticable in Eng-1. The Non-Brit group is not divided according to education, since only one Non-Brit listener declared any sort of training in phonetics.

As was the case in Eng-1, the ultimate aim of research objective 1, the development of a new perception test for asylum applicants, is not directly approached in Eng-2. This is because Eng-2 is an experiment on British English, a set of dialects of no interest to LAAP. Steps towards development of a new test of perception are made only in Ar-3, the last in the Arabic experimental series (Chapter 11).

The findings of Eng-2 are of nonetheless of interest in their own right. The much larger pool of listeners in Eng-2 (N = 197), compared to that of Eng-1 (N = 28), allowed for an inferential statistical analysis of research question 1 and thus the degree of support in the data for hypotheses (i) and (ii).

Research objective 2 of the thesis is detailed in Chapter 7, section 7.3. For the same reasons explained there, research objective 2 is not relevant to Eng-2.

8.4. Methodology

8.4.1. Task design

The task design of Eng-2 was broadly the same as that of Eng-1 (Chapter 7, section 7.4). There were differences, however, in the criteria used to sort listeners into groups, in the number and geographical origins of stimuli and in stimuli duration, as well as in the use of an expanded Likert scale (see Chapter 7, section 7.8, for the motivations behind these design modifications).

As in Eng-1, the question listeners were asked to answer in response to each stimulus was 'Is this a Yorkshire accent?' Likewise, too, listeners were asked to speculate (optionally, in a free-text box) on the provenance of each stimulus if they responded to this question in the negative. Unlike in Eng-1, however, listeners were prompted to respond on a five-point Likert scale: *Highly Likely – Likely – Uncertain – Unlikely – Highly Unlikely* (cf. Wilson 2009).

Figure 8.1 is an example of the task display, including five-point Likert scale, on Qualtrics. Figure 8.2 shows the free-text box for optional entry of additional comments.

Figure 8.1: Example of task display, including five-point Likert scale, in Qualtrics
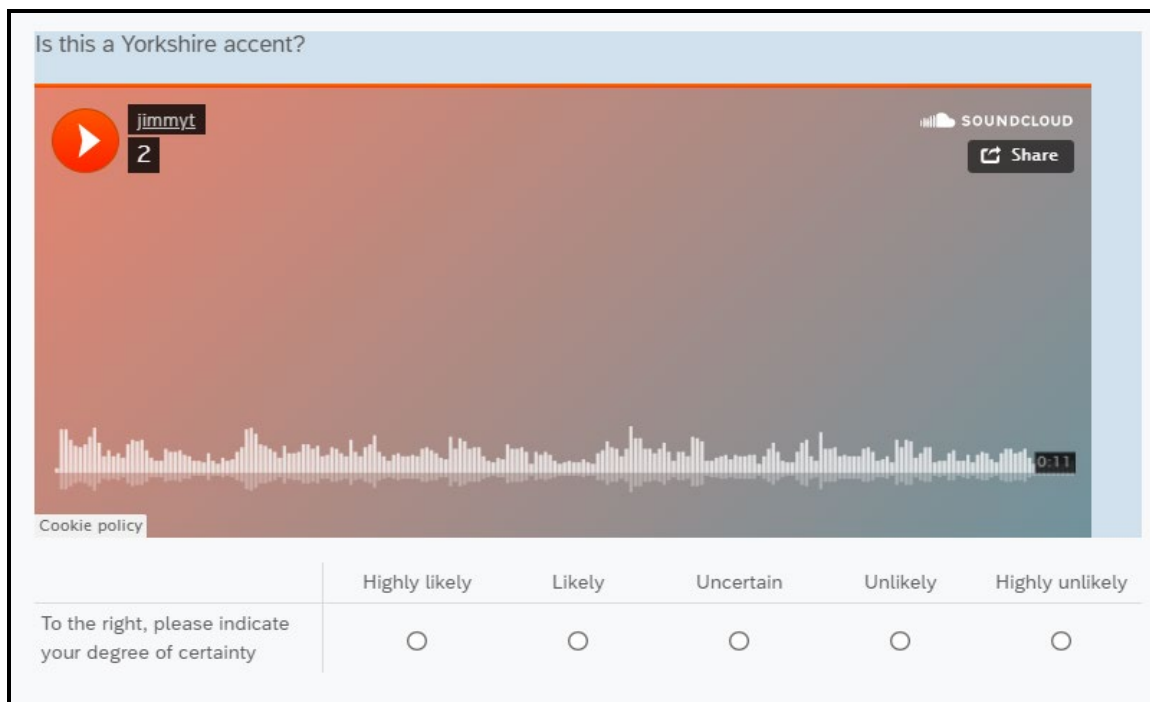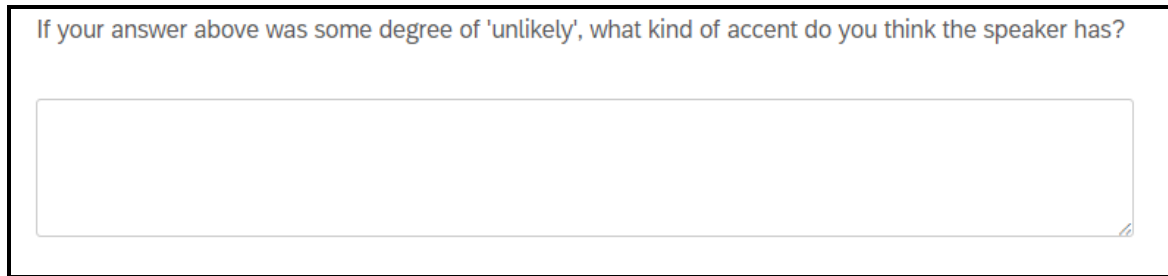
Figure 8.2: Free-text box for entry of additional comments

If your answer above was some degree of 'unlikely', what kind of accent do you think the speaker has?

Unlike in Eng-1, a five-point Likert scale was used in Eng-2 because it was thought possible that the lack of nuance involved in a simple three-way choice may have led listeners in Eng-1 to respond 'don't know' unless they were absolutely convinced of their answer. In addition, the employment of a five-point Likert scale permits an assessment of the relative confidence, and not just the accuracy, embedded in 'decision' (i.e. correct and incorrect) responses. An analysis of this type would have the potential to reveal, for example, correlations between degree of confidence in 'decision' responses and overall accuracy (cf. Wilson 2009). However, a statistical analysis accounting for responses at each point on the Likert scale was not attempted in Eng-2, since it concerns language varieties (British English) which are not directly relevant to LAAP.

Responses on the five-point scale were instead combined into two pairs of two categories for the purposes of statistical analysis. The first pair, 'correct' vs. 'incorrect', was used in one series of model comparisons to assess the relative *accuracy* of group response patterns. The second pair, 'decision' (i.e. correct or incorrect) vs. outright uncertain responses, was employed in a further series of model comparisons to assess relative *confidence*. This binary measure of confidence is subjected to inferential statistical analysis in Eng-2 (and later in Ar-3). As noted in Chapter 5, a similar examination of confidence was not attempted in Wilson (2009).

Predictions about the effect on results of the variable of confidence, and its possible interaction with accuracy, were not incorporated into the thesis' guiding hypotheses (the reasons for this are explained in Chapter 5). However, in Eng-2 confidence is explored in addition to accuracy as a measure of intergroup response variability in order to evaluate the assertion that NSNLs in LAAP are more confident in their judgements than are linguists (cf. Fraser 2009, 2011, 2019).

8.5. Listeners

8.5.1 Recruitment

Listeners were recruited in the same manner as in Eng-1 (refer to Chapter 7, subsection 7.5.1). There were 328 participants in total. Complete responses were submitted by 198 English native-speaking listeners. This represented an improvement on the Eng-1 completion rate (60% versus 41%) and was interpreted as justification for the revisions to stimuli number and duration specified in Chapter 7, section 7.8.

Table 1 shows the number of listeners whose responses were included, by education and (sub)national background. Of non-linguist listeners there were 159 in total: 55 Yorks B&R, 39 Yorks Res, 45 Brit non-Yorks and 20 Non-Brit. Of linguists there were 39: 12 Yorks B&R, 9 Yorks Res, 17 Brit non-Yorks and 1 Non-Brit. The small number of Non-Brit linguist respondents meant that a group of listeners of this type could not be composed. The responses of the single Non-Brit linguist were therefore omitted entirely, meaning that the responses of 197 listeners were included in the final analysis.

Table 8.1: Number of listeners, by education and (sub)national background

Education /Background	Yorks B&R	Yorks Res	Brit-non-Yorks	Non-Brit	Totals
Non-linguists	55	39	45	20	159
Linguists	12	9	17	n/a	38
Totals	67	48	62	20	197
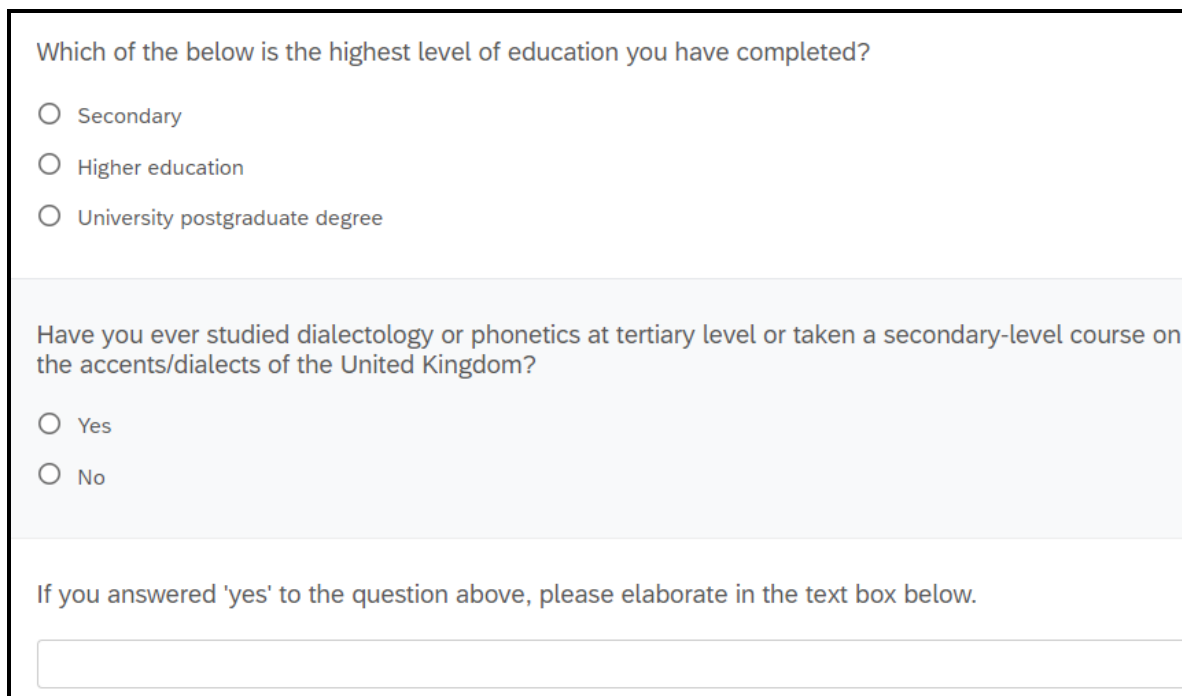
8.5.2. Method used to allocate listeners to groups

The classification of listeners followed the same criteria as in Eng-1 (see Chapter 7, subsection 7.5.2), with one difference. In Eng-1, placing listeners into one of four groups involved asking whether they had received formal academic training in linguistics. Qualtrics' skip logic function prevented progression further into the survey by listeners who answered 'yes' to this question.

In Eng-2, listeners were instead asked, 'Have you ever studied dialectology or phonetics at tertiary level or taken a secondary-level course on the accents/dialects of the United Kingdom?' If they answered 'yes' to this question, they were asked to elaborate. Those who answered 'yes' were not prevented from continuing with the survey. Inclusion of their responses allowed comparative analysis of the accuracy/certainty of linguistically-trained and -untrained listeners, which led in turn to the contribution made by this experiment to the highly vexed NSNL vs. linguist debate (see below, section 8.8, for discussion)

Listeners were also asked to specify the highest level of education they had completed. The intention was to allow for independent control of this variable, though in the present analysis this was not done. Figure 8.3 shows the display of survey questions, on Qualtrics, in connection with educational background.

Figure 8.3: Display on Qualtrics of survey questions about educational background

Which of the below is the highest level of education you have completed?

○ Secondary

○ Higher education

○ University postgraduate degree

Have you ever studied dialectology or phonetics at tertiary level or taken a secondary-level course on the accents/dialects of the United Kingdom?

○ Yes

○ No

If you answered 'yes' to the question above, please elaborate in the text box below.

8.6. Stimuli

In Eng-2, stimuli were presented initially in fixed order and then in pseudo-random order. A statistical evaluation of order effects was conducted and showed no significant result.

8.6.1. Stimuli: selection

Ten stimuli were obtained from the same source and edited as detailed in subsection 7.6.1 of Chapter 7. Stimuli again featured Yorkshire and non-Yorkshire speakers—the latter a revised selection from (historical) Lancashire, as well as one foil from Birmingham. Stimuli included five from males and five from females.

Four Yorkshire stimuli (Greetland, Stainland, Leeds and Halifax) and three foils (Manchester 3, Wigan 1 and Salford—all within Greater Manchester) were retained from Eng-1. It was thought that the Yorkshire/non-Yorkshire ratio was thus unbalanced and the range of stimuli too geographically restricted, so three further non-Yorkshire stimuli from further afield in England were selected as replacements. These featured speakers from Blackpool, Liverpool and Birmingham. The first two towns, like the locations of the Lancashire stimuli re-used from Eng-1, are within the dialectological north-west of England. Birmingham is on the border of north and south, and its accent contains both northern and southern features (Wells 1982).

Like the accents in the foil stimuli used in Eng-1, the accent of Blackpool numbers among those of Lancashire. The city of Liverpool, too, formerly belonged to the county of Lancashire. However, the Liverpool accent is in many senses quite distinct from those of the rest of the historical county, as well as of the remainder of northern England (Wells 1982; Watson & Clark 2013). The Liverpool stimulus was thus included in Eng-2 as (what was predicted to be) an obvious outlier; the Birmingham stimulus was included for the same reason. The Blackpool stimulus was used partly because the featured speaker had a history of residence in Liverpool, and features of her speech were impressionistically identified as influenced by the distinctive Scouse accent. These particular stimuli were thus chosen because they seemed to contain

features fairly identifiable as non-Yorkshire. The relevant features are specified below, in subsection 8.6.5.

8.6.2. Stimuli: speakers' geographical origins

As in Eng-1, stimuli were downloaded from [IDEA: International Dialects of English Archive](). All of the IDEA stimuli used here were recorded in the early 2000s. In Table 8.2, stimuli are identified as 'Yorkshire' or 'non-Yorkshire' (not as 'Lancashire', as in Eng-1), in accordance with the fact that they featured a wider range of accents than did the stimuli in Eng-1. An asterisk next to the location of a stimulus indicates that it was also used in Eng-1.

Table 8.2: Geographical origins of stimuli, with biographical information from *IDEA*

Stimulus origin	Description in *IDEA*
Greetland (Yorkshire)*	England 55 female, 91, 1912, Greetland and Calderdale (West Yorkshire)
Halifax (Yorkshire)*	England 57 male, 52, 1951, white, Leeds and Halifax (West Yorkshire)
Leeds (Yorkshire)*	England 81 male, teens, 1989, white, Dewsbury and Leeds (West Yorkshire)
Stainland (Yorkshire)*	England 56 male, 48, 1955, white, Stainland (West Yorkshire)
Wigan (non-Yorkshire)*	England 15 male, 34, 1966, white, Wigan (Lancashire)
Manchester (non-Yorkshire)*	England 71 male, 39, 1968, white, Manchester
Salford (non-Yorkshire)*	England 54 female, late 40s, 1950s, white, Salford (Lancashire)
Blackpool (non-Yorkshire)	England 14 female, 28, 1972, white, Blackpool and Liverpool
Birmingham (non-Yorkshire)	England 102 female, 63, 1954, white, Birmingham
Liverpool (non-Yorkshire)	England 44 female, 31, 1970, white, Kirkdale (Liverpool) and Manchester

Figures 8.4 and 8.5 show the places of origin within/outside Yorkshire of the speakers featured in the stimuli.

Figure 8.4: Map of speaker origins; Midlands/Yorkshire/Lancashire partially shown (Yorkshire stimuli in blue, non-Yorkshire in purple; approximate line of historical Yorkshire border in orange)

Figure 8.5: Map of speaker origins; historical counties of Yorkshire/Lancashire partially shown (Yorkshire stimuli in blue, non-Yorkshire in purple; approximate line of historical Yorkshire border in orange)
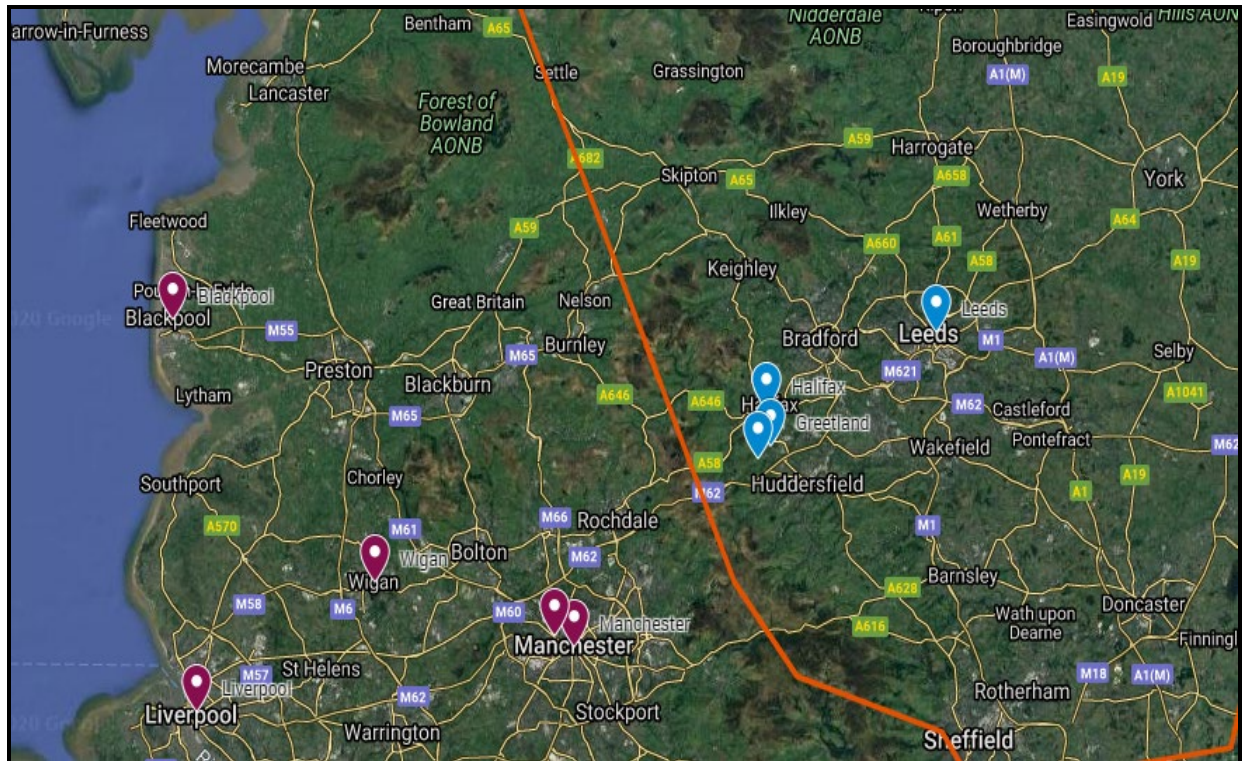
8.6.3. Stimuli: duration

As specified above, stimuli were edited, using Audacity, to an average duration of 10 seconds (range: 9 to 11 seconds). The reduced duration (cf. c. 30 seconds in Eng-1) was motivated partly by the attempt to increase the number of completed responses. It was also predicted that the reduced duration might serve to accentuate differences in response patterns between listener groups.

8.6.4. Stimuli: text

In Eng-1, stimuli featured readings of both *The Rainbow Passage* and *Comma Gets a Cure*. For Ar-2, in the interests of uniformity of text, all stimuli featured readings of *Comma Gets a Cure*.

The section of text included was as below.

Well, here's a story for you: Sarah Perry was a veterinary nurse who had been working daily at an old zoo in a deserted district of the territory, [so she was very happy to start a new job at a superb private practice in North Square near the Duke Street Tower].

The section in square brackets was included only in the Wigan stimulus, because it alone featured a realisation of the SQUARE-NURSE merger (as [əː], in *square*). As discussed in Chapter 7, subsection 7.6.5.2, this vowel merger is often identified as typical of NW England. It was thought that its occurrence might distinguish this, the sole Lancashire stimulus of a predominantly 'traditional Northern' character, as not of Yorkshire. This was important because three of the four Yorkshire stimuli were also mainly of a traditional Northern character and shared many features (but not the SQUARE-NURSE merger as [əː]) with the Wigan stimulus.

8.6.5. Stimuli: summary of features included

Four stimuli were of Yorkshire English. These featured speakers from Greetland, Leeds, Stainland and Halifax. The discussion in subsection 7.6.5.1 of Chapter 7 covers in full the range of features included.

There were six non-Yorkshire stimuli. The features present in the three stimuli retained from Eng-1 (Manchester 3, Wigan 1 and Salford) are discussed in subsection 7.6.5.2 of Chapter 7. Some additional exposition is necessary, though, of features of the three new non-Yorkshire stimuli.

The features of the Blackpool stimulus were (it seemed to me) broadly GNE, with the exception of the speaker's pronunciation of *nurse*, which is approximately [ɛː]. Its presence in the speaker's repertoire may result from her declared period of residence in Liverpool. This pronunciation of *nurse* is common in Liverpool English (Wells 1982, p. 361). It was predicted that this feature, in combination with the other 'General Northern but not necessarily

Yorkshire' features observed in the stimulus, might be sufficient to identify it as non-Yorkshire.

The Birmingham stimulus featured (in *Perry*) a characteristically West Midlands diphthongal realisation of happY, [ɜɪ], which is generally not associated with any accent of Yorkshire (Wells 1982). This was the only stereotypically non-Yorkshire feature in the stimulus but was considered enough to mark it as such. There were also variably tense monophthongal tokens of happY, in *daily* and *territory*.

The Liverpool stimulus contained abundant stereotypical tokens: the stressed vowels in *nurse*, *Sarah* and *working* as [eː], affricated realisations of /t/ (in *veterinary*, *territory*, *deserted* and *district*), initial /d/ (in *deserted*, *daily* and *district*) and /k/ (in *working*). The stimulus was thus thought to be fairly obviously non-Yorkshire.

Of the three new stimuli, Birmingham and Liverpool proved reliable in separating listener group performance, with reliability defined as occasioning above-chance accuracy in one listener group or more. Of the seven stimuli retained from Eng-1, Greetland, Stainland, Halifax and Salford proved reliable in this regard. These results are detailed in subsection 8.7.2.4.

8.7. Results

8.7.1. Organisation

Subsection 8.7.2 presents descriptive statistics, by percentage totals, indicating how the response patterns of the seven groups of listeners varied when confronted with the question, 'Is this a Yorkshire accent?' In Figures 8.6 and 8.7, response patterns by all seven groups to, respectively, Yorkshire and Non-Yorkshire stimuli are compared (first non-linguists, then non-linguists). The seven groups' accuracy/confidence on all stimuli combined is compared in Figures 8.8 and 8.9. Group responses to individual stimuli are compared in Figures 8.10 and 8.11, with groups divided along national rather than regional and/or educational lines. Subsection 8.7.2.5is a brief discussion of the overall picture suggested by descriptive statistics prior to inferential statistical analysis.

Where Yorkshire stimuli are concerned, a correct response means 'correctly accepting' as from Yorkshire the stimulus in question—that is, the response must be either 'highly likely' or 'likely'. The term 'false rejection' is used to describe an incorrect response (either 'highly unlikely' or 'unlikely') to Yorkshire stimuli.

In connection with non-Yorkshire stimuli, a correct response means 'correctly rejecting' as from Yorkshire the stimulus in question, meaning that the response must be either 'highly unlikely' or 'unlikely'. The term 'false acceptance' is used to describe an incorrect response (either 'highly likely' or 'likely') to non-Yorkshire stimuli.

'Uncertain' responses are considered as a response type distinct from 'decision' responses. The reason for the treatment of such responses here is the thesis' attempt to investigate confidence in a manner that has not previously been attempted in LAAP related work (see Chapter 5 for discussion of alternative conceptions of confidence, e.g. in Wilson 2009). Table 8.3 specifies the designation of 'decision' responses by stimulus type.

Table 8.3: Designation of 'decision' responses by stimulus type

	'Yes' response (i.e. this is a Yorkshire accent)	'No' response (i.e. this is not a Yorkshire accent)
Yorkshire	Correct acceptance	False rejection
Lancashire	False acceptance	Correct rejection

Subsection 8.7.3 presents the results of an inferential statistical analysis of accuracy, in the form of a series of model comparisons, with correct and incorrect responses as the dependent variables. This is an attempt at a rigorous investigation of research objective 1 (i.e. examining the respective accuracy of group response patterns to the question 'Is this a Yorkshire accent?') and, by extension, an evaluation of the extent to which hypotheses (i) and (ii) of the thesis are supported.

An additional aim of the inferential analysis, as mentioned in subsection 8.4.1, is an initial investigation of intergroup differences in confidence. Accordingly, a series of model

comparisons is also conducted in subsection 8.7.3, using the binary dependent variables of uncertain and 'decision' responses.

Inferential statistical results are discussed in the relevant subsections of 8.7.3 with reference to their implications for the validity of the thesis' hypotheses, as well as for the familiarity effect. In section 8.8. I summarise these results and their possible import, firstly for LAAP as currently practiced, secondly for the envisaged supplementary test of applicant perceptions, and thirdly for the design of the Arabic experimental series.

8.7.2. Descriptive statistics

8.7.2.1 Non-linguists' responses to Yorkshire and non-Yorkshire stimuli

Figure 8.6 shows non-linguists' response pattern, by percentage of correct acceptances, false rejections and uncertain responses, to Yorkshire and non-Yorkshire stimuli respectively.

Figure 8.6: Non-linguists' responses to Yorkshire and non-Yorkshire stimuli
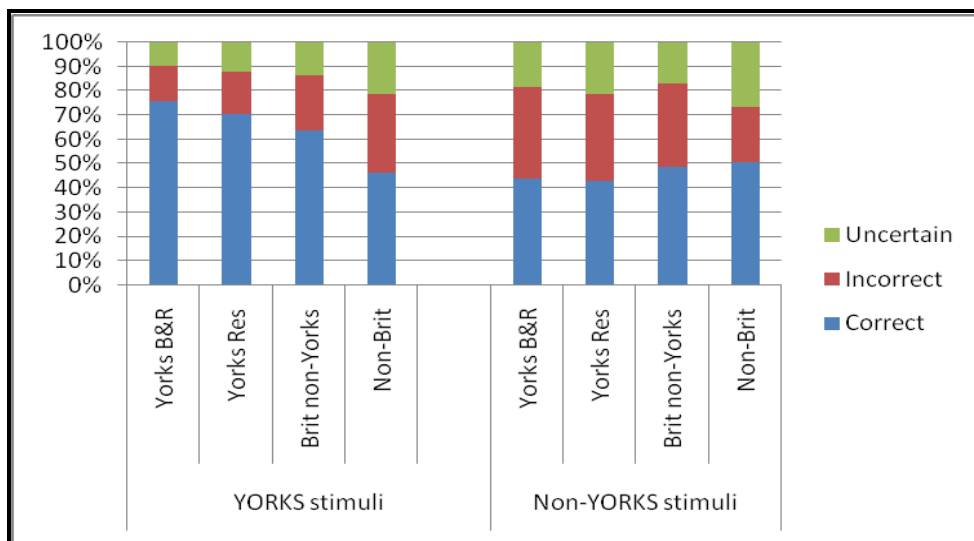
On Yorkshire stimuli, Yorkshire-associated listeners were the most accurate and the least uncertain of the four non-linguist groups (e.g. Yorks B&R: 76% correct acceptances, 15% false rejections and 10% uncertain responses). On stimuli of this type Yorks Res were slightly lower in accuracy and slightly more uncertain than Yorks B&R but more accurate and more certain than either of the non Yorkshire-associated groups (cf. Non-Brit 46% correct acceptances, 33% false

rejections and 21% uncertain responses).

On non-Yorkshire stimuli, however, the pattern is broadly the inverse, at least in relative accuracy. Only one group scored above chance on correct rejections (Non-Brit, at 51%). Interestingly, Non-Brit also entered considerably fewer incorrect acceptances than any other group (23% vs. e.g. c. 38% Yorks B&R). Yorkshire-affiliated listeners were much less certain than they were on Yorkshire stimuli: c. 20% uncertain vs. c. 10%. The other two groups were also somewhat less confident on non-Yorkshire stimuli than they were on Yorkshire stimuli (cf. Non-Brit c. 27% vs. c. 21% uncertain).

8.7.2.2. Linguists' responses to Yorkshire and non-Yorkshire stimuli

Figure 8.7 shows linguists' response pattern, by percentage of correct acceptances, false rejections and uncertain responses, to Yorkshire and non-Yorkshire stimuli respectively.

Figure 8.7: Linguists' responses to Yorkshire and non-Yorkshire stimuli
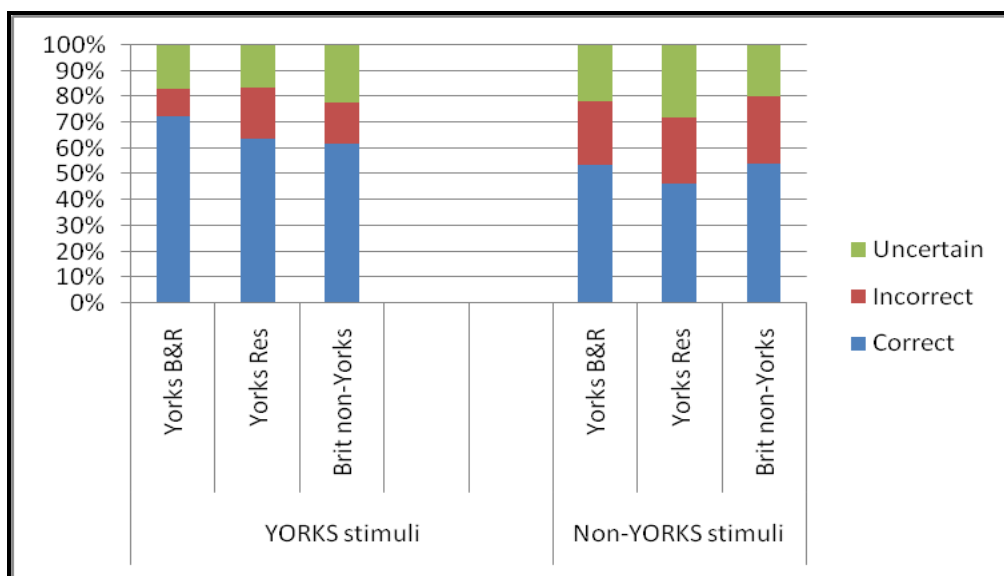
Observations about the groups' relative accuracy in Figure 8.6 apply broadly to Figure 8.7, in that non-linguists and linguists alike performed more accurately on Yorkshire than on non-Yorkshire stimuli, with the two Yorkshire-affiliated linguist groups entering the largest proportion of correct acceptances (Yorks B&R 71%, Yorks Res 64%). There are nevertheless some notable differences between Figures 8.6 and 8.7.

Firstly, all three linguist groups entered slightly fewer correct acceptances than did the non-linguists (e.g. Yorks B&R linguists, 71%; Yorks B&R non-linguists 76%). Yorks B&R and Brit non-Yorks linguists also entered appreciably fewer false rejections than did their non-linguist counterparts (e.g. Brit non-Yorks linguists 16%, Brit non-Yorks non-linguists 23%). Yorks Res is an exception on this measure, in that linguists entered marginally more false rejections (19%) than did non-linguists (16%).

Secondly, on non-Yorkshire stimuli, all three linguist groups entered a higher percentage of correct acceptances than did non-linguists, with Yorks B&R and Brit non-Yorks linguists each achieving above chance accuracy (54%) on this measure (cf. Yorks B&R non-linguists 44%, Brit non-Yorks non-linguists 49%). The three linguist groups also entered fewer incorrect acceptances than did non-linguists (e.g. Yorks Res linguists 26%, Yorks Res non-linguists 35%).

Thirdly, linguists entered more uncertain responses than did non-linguists, to both kinds of stimuli. On non-Yorkshire stimuli, for example, Yorks Res linguists entered 28% uncertain responses, compared to Yorks Res non-linguists 21%.

Fourthly, on Yorkshire stimuli, Non-Brit presents in Figure 8.6 as an obvious outlier among the non-linguist groups. It was the only one of the four groups to enter a percentage of correct acceptances at below-chance level (46%). Its percentages of both false rejections and uncertain responses (21%) were correspondingly anomalous. On percentage of false rejections, for example, the distance between Non-Brit (33%) and Brit non-Yorks (23%) was four times that between Yorks B&R (15%) and Yorks Res (17%), and almost double that between Brit non-Yorks and Yorks Res.

8.7.2.3. Non-linguists' responses to all stimuli combined

Figure 8.8 displays non-linguists' response pattern, by percentage of correct, incorrect and uncertain responses, to all stimuli combined.

Figure 8.8: non-linguists' responses to all stimuli combined
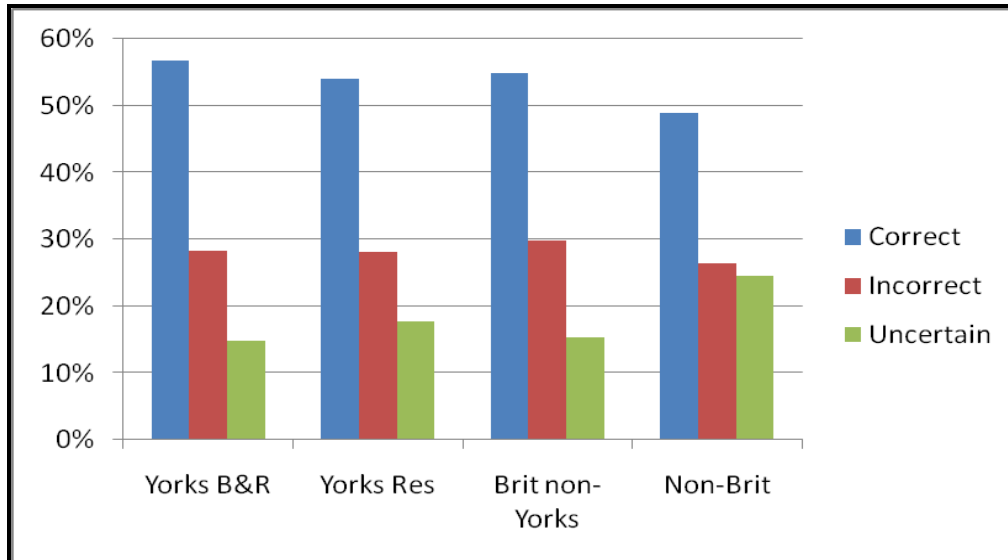
All three British groups entered approximately the same percentage of correct responses, at c. 55%, with Yorks B&R slightly higher in the range (57%) and Yorks Res slightly lower (54%). The Non-Brit group achieved slightly below chance-level accuracy (49%).

All groups entered c. 25-30% incorrect responses, with Brit non-Yorks highest in the range (30%) and Non-Brit (26%) the lowest. Differences among the three British groups in percentage of uncertain responses were minimal (all c. 15%); only the Non-Brit group entered a notably higher percentage of uncertain responses (25%). As in the comparisons displayed in Figure 8.6, the Non-Brit group again appears as an outlier here.

8.7.2.3. Linguists' responses to all stimuli combined

Figure 8.9 displays linguists' response pattern, by percentage of correct, incorrect and uncertain responses, to all stimuli combined.

Figure 8.9: linguists' responses to all stimuli combined
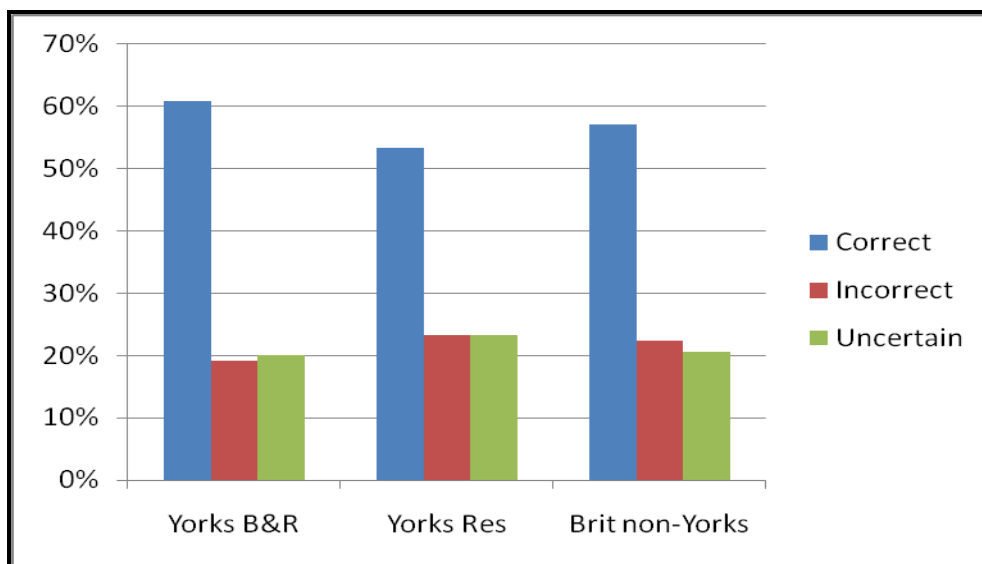
All three groups achieved above chance-level accuracy, with Yorks B&R (61%) highest in the range. Yorks Res, however, scored only slightly above chance level (53%).

On incorrect responses all three groups were roughly equal (c. 20%), with Yorks B&R at the lower end (19%) and Yorks Res at the upper end (23%) of the range. Yorks B&R and Brit non-Yorks entered a nearly identical percentage of uncertain responses (c. 20%), while Yorks Res entered slightly more than this (23%).

Several observations can be made on the basis of Figures 8.8 and 8.9. First, the Non-Brit (non-linguist) group is an outlier in having entered correct responses at below chance level, and uncertain responses at 25% (cf. the three British non-linguist groups at above-chance correct and c. 15% uncertain). Second, all linguist groups entered more correct and fewer incorrect responses than did all non-linguist groups (Yorks Res excepted, where the percentage of correct and incorrect responses entered by non-linguists and linguists was approximately equal). Third, among non-linguists of the three British groups, the percentage of incorrect responses was almost double that of uncertain responses, while among linguists the two response types

occurred in nearly equal proportion.

8.7.2.4. Group responses to individual stimuli

Figures 8.10 and 8.11 together compare group responses to individual stimuli. This comparison is not directly relevant to the research objectives of the thesis but is included to demonstrate the extent to which each of the stimuli proved reliable in discriminating among group response patterns. Reliability is here defined as occasioning above-chance accuracy in one listener group or more.

For brevity's sake, groups are stratified by national (i.e. not educational or regional) background. This is because few notable differences in results were observed to obtain on the variables of regional or (linguistic) educational background. An exception was the Manchester stimulus, on which the two Yorkshire-affiliated groups, irrespective of linguistic education, entered correct rejections at slightly below-chance level (i.e. nearly the same as the Non-Brit group), while Brit-non-Yorks achieved slightly above chance-level accuracy on the same measure. This marginal difference meant that it was not taken as a reliable discriminator.

Figure 8.10 shows all British groups' responses to individual stimuli; Figure 8.11 shows the Non-Brit group's responses to individual stimuli.

Figure 8.10: Combined British groups' responses to individual stimuli
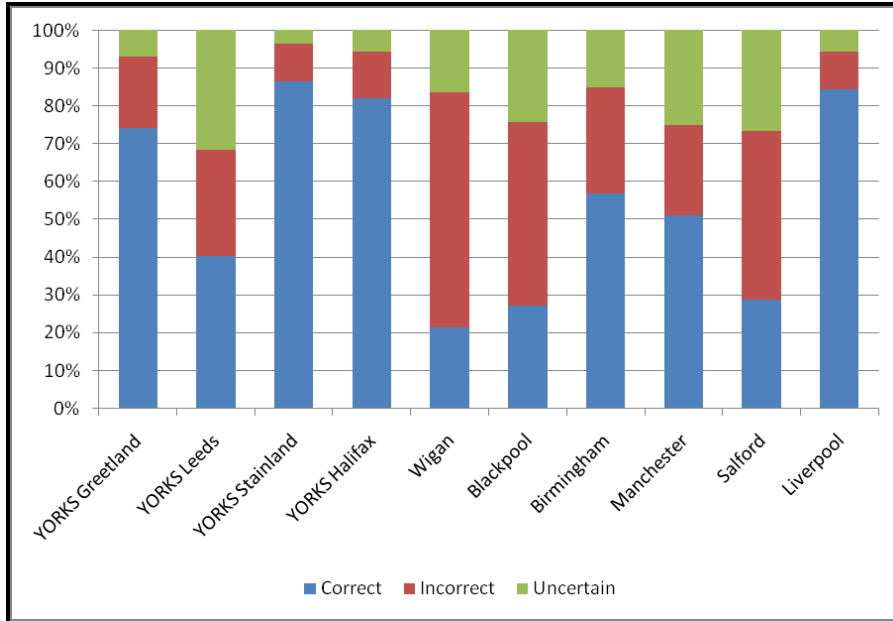
Figure 8.11: Non-British group's responses to individual stimuli
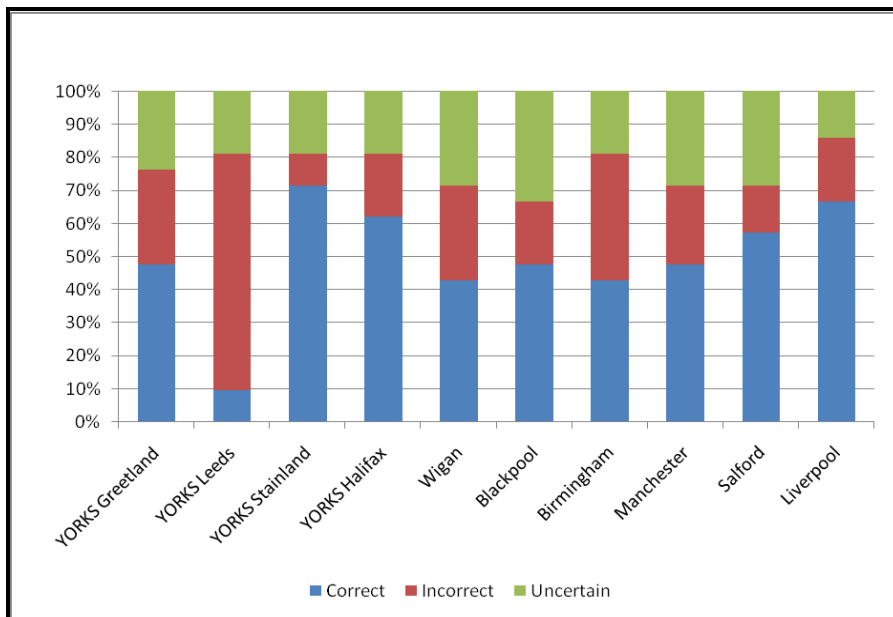
Of the Yorkshire stimuli, Greetland provoked the greatest divergence in correct acceptances by the two groups (British 75% vs. non-British below chance); Halifax (82% vs. 62%) and Stainland (85% vs. 71%) were respectively second and third most reliable. All three (especially Greetland) include a fairly large number of traditional Yorkshire features.

The fourth Yorkshire stimulus, Leeds, with its (in my judgement) much greater GNE influence, occasioned below-chance performance by both groups. All four Yorkshire stimuli did, however, provoke considerable differences in percentage of uncertain responses (e.g. Stainland < 5% vs. 20%, Leeds 30% vs. 20%).

Of the non-Yorkshire stimuli, Birmingham, Liverpool and Salford turned out to be reliable discriminators. Salford, though to my ear overwhelmingly an exponent of GNE, was most reliable: 58% correct rejections among non-British listeners but only 19% in the British group. This result perhaps suggests greater awareness among British listeners of the prevalence of GNE in Yorkshire.

Birmingham, the most reliable discriminator, elicited 56% correct rejections among British listeners, with non-British listeners at 41% on this measure. This difference in accuracy perhaps indicates British listeners' greater awareness of canonical divergences in the phonology of West Midlands and Yorkshire speech (e.g. monophthongal vs. diphthongal happY).

Liverpool, replete with local phonology, proved relatively easy for both groups to correctly reject, though the difference was nonetheless notable (British 85% vs. non-British 68%). It also resulted in a substantial group-level difference in (un)certainty: British 5% vs. non-British 15% uncertain.

Wigan and Blackpool were of little use as discriminators, though likely for different reasons. The Wigan stimulus, as an exponent of a traditionally Northern dialect, may have been relatively easy to mistakenly identify with Yorkshire. Blackpool, apart from a single token suggesting the influence of Scouse phonology, has a predominantly GNE character and might thus have been

difficult to reject as Yorkshire.

8.7.2.5. Discussion

Preliminary to inferential statistical analysis, three main observations of interest emerge from the preceding review of descriptive statistics. The first is that the non-linguist and linguist Yorks B&R listener groups were each more accurate, both on Yorkshire stimuli and overall, than any of the other five groups, linguist or non-linguist. Yorks B&R non-linguists slightly outperformed linguists of the same background on Yorkshire stimuli. This finding contrasts notably with the results of Eng-1, in which marginally the most accurate group on Yorkshire stimuli was Yorks Res, though on Yorkshire stimuli Yorks B&R and the two non-Yorkshire-affiliated groups still performed broadly in the direction predicted by the other-accent effect. It is possible that the superior performance of the Yorks Res group in Eng-1 is attributable to the much smaller overall sample size (cf. six Yorks Res listeners of 28 in Eng-1 vs. 39 of 198 in Eng-2).

The second point of interest is that the linguist and non-linguist Yorks Res groups were more accurate than the non Yorkshire affiliated groups. This result lends preliminary support to the validity of the familiarity effect.

Third, it is notable that linguists, in addition to their greater accuracy, generally evidence a greater propensity than non-linguists towards entering uncertain responses—that is, they are less confident in their responses. It is also noteworthy, however, that Non-Brit non-linguists appear to be the least confident group of all.

Two further remarks are worth making about these descriptive results. The first is that the pattern observed in Eng-1, in which all groups entered a noticeably higher percentage of false acceptances than false rejections, is also applicable here—except in the case of non-Brit, in which the (marginally) opposite tendency is discernable. As in Eng-1, this may have been caused by the relatively non-particular character of the GNE-like stimuli used in Ar-2, in spite of attempts made in stimuli selection to minimise their number. As will be shown in subsequent chapters, this pattern is partially violated in the Arabic series of experiments, where only Syrian

(i.e. 'local') listeners consistently record a larger percentage of false acceptances than false rejections, the latter being very uncommon among them.

Secondly, in Eng-2, the Non-Brit group presents as a general outlier. In addition to being unique in recording more false rejections than false acceptances, it evinces a distinct pattern on overall accuracy and confidence—lower on each than any of the six British groups—and in its relatively low-accuracy and low-confidence responses to some individual Yorkshire stimuli. As I suggest below, in subsections 8.7.3.3 and 8.7.3.5, the response patterns of the non-linguist Non-Brit group—unmatched as it was with a *linguist* Non-Brit group—may have skewed inferential statistical analysis of both accuracy and confidence.

8.7.3. Inferential statistics

8.7.3.1. Procedure

For Eng-2, inferential statistical analysis was conducted in *R* (R Core Team 2022), by means of a series of binomial logistic mixed effects models (a form of glmer) built with the lme4 package. Fixed effects were speaker geographical background (two levels: Yorkshire or non-Yorkshire), listener geographical background (four levels: Yorks B&R, Yorks Res, Brit non-Yorks, Non-Brit) and listener education (two levels: linguistic education or no linguistic education). Participant was included as a random effect. Paired interactions between all three fixed effects were also incorporated into the full model.

To test the significance of each variable in the full model, a partial model was composed for each variable at issue. These partial models were the same as the full model except that in each instance the variable in question was excluded. Model comparisons were conducted by means of ANOVAs.

Listener responses to the question 'Is this a Yorkshire accent?' constituted the dependent variable. Two series of dependent variables (DVs) were analysed: one assessing accuracy of response, and the other assessing confidence. Accuracy was modelled as '0' or '1' according to

an incorrect or correct response (regardless of degree of confidence as expressed by Likert-scale ratings). Confidence was also modelled as '0' or '1', according to whether the response was an 'uncertain' or 'decision' response, irrespective of accuracy or degree of confidence.

Outright 'uncertain' responses (i.e. point three on the relevant Likert scale) were excluded from the analysis of accuracy. This decision was made because of the difficulty in deciding whether listeners' selection of this response type constitutes a simple admission of 'don't know' (i.e. lack of competence/familiarity) or caution in identifying stimuli as Yorkshire/non-Yorkshire due to familiarity/competence; this epistemological problem is discussed in Chapter 5. The exclusion of 'uncertain' responses means that the y-axes in Figures 8.12 to 8.14 display a higher probability of accuracy than would have resulted from their inclusion.

Each series of DVs was analysed in a separate series of model comparisons. Accuracy is dealt with in subsection 8.7.3.2 and 8.7.3.3, followed by confidence in subsections 8.7.3.4 and 8.7.3.5.

8.7.3.2. Listener accuracy: results

Table 8.4 presents the coefficients and significance values for the model of best fit, as determined by glmer, for accuracy of listener response. It also specifies significance values, where present, for the fixed effects. The intercept encodes three baseline fixed effects: 'education linguist', 'speaker Non-Yorks' and 'background Brit Non-Yorks'.

Table 8.4: Coefficients and significance values for glmer analysis of accuracy

```
Fixed effects:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                           0.6567     0.1589   4.133 3.58e-05 ***
Educationnon-linguist                -0.3064     0.1438  -2.132 0.033040 *
SpeakerYorks                          0.6816     0.1985   3.434 0.000595 ***
BackgroundNon-British                 0.3890     0.2520   1.544 0.122706
BackgroundYorkshire B+R              -0.1561     0.1623  -0.962 0.336085
BackgroundYorkshire res              -0.1777     0.1791  -0.992 0.321071
SpeakerYorks:BackgroundNon-British   -1.0454     0.3880  -2.695 0.007047 **
SpeakerYorks:BackgroundYorkshire B+R  0.7708     0.2897   2.661 0.007794 **
SpeakerYorks:BackgroundYorkshire res  0.4963     0.3089   1.607 0.108150
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results demonstrate significant effects (by means of predicted probabilities) on accuracy for four independent variables, or pairs of independent variables in interaction. The first is education (i.e. whether the listener declared secondary-level or higher training in British phonetics). The second is speaker origin (i.e. whether the stimulus was of Yorkshire or not). The third and fourth specify interactions between speaker origin and listener background (i.e. the listener's association with Yorkshire through birth and/or upbringing, solely residence as an adult).

Figures 8.12 to 8.14 illustrate in turn the predicted probability of correct response (i.e. accuracy) associated with education, then speaker (i.e. whether the stimulus featured a Yorkshire or non-Yorkshire speaker) and finally the interaction between the background of the speaker in the stimulus and that of the listener.

Figure 8.12: Predicted probability of correct response associated with, respectively, linguistic education and no linguistic education
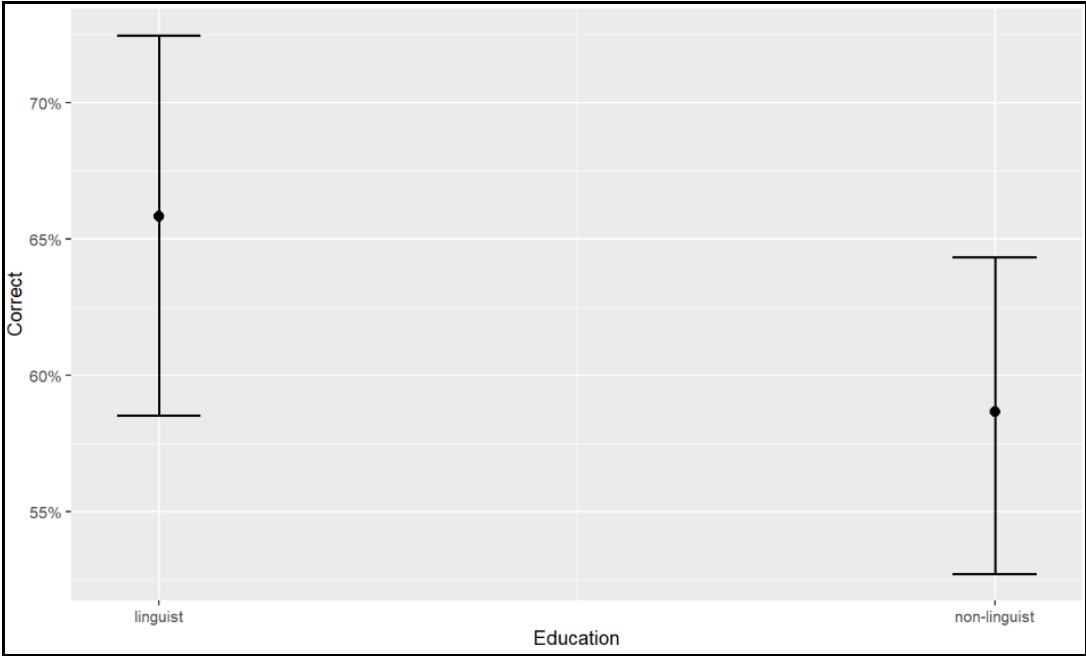
Figure 8.13: Predicted probability of correct response associated with, respectively, non-Yorkshire and Yorkshire stimuli

Figure 8.14: Predicted probability of correct response associated with interaction between listener and stimulus type
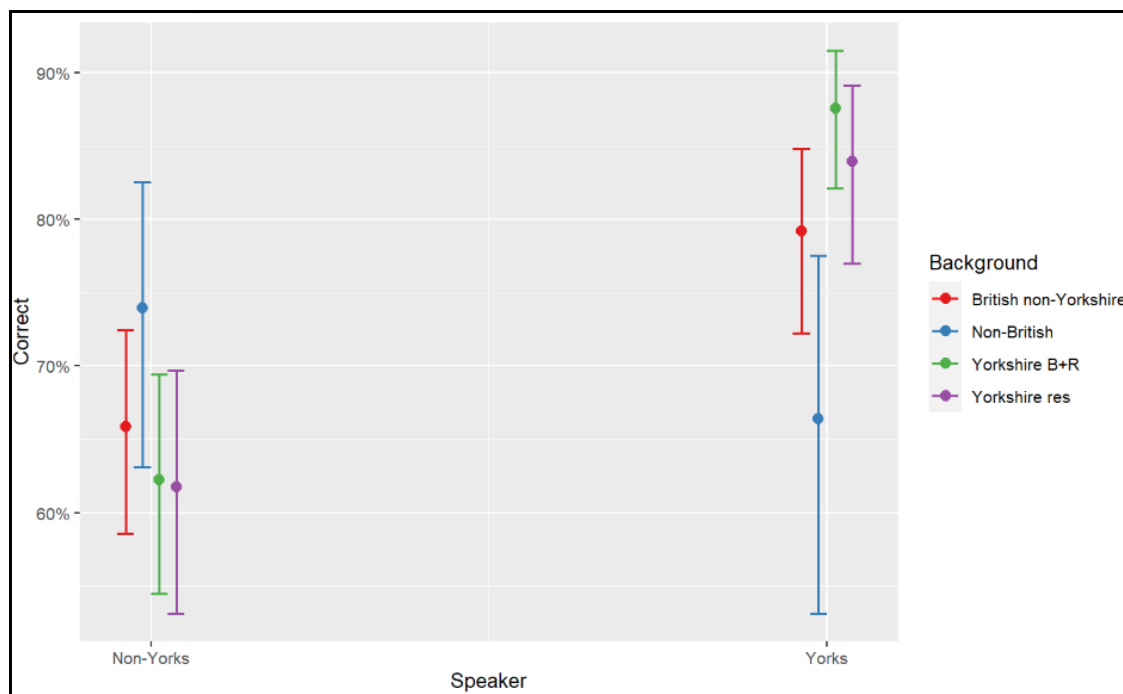
8.7.3.3. Discussion of results: listener accuracy

Results indicate support for hypothesis (i), that local listeners will be most accurate in identifying local speakers. Yorks B&R were significantly more accurate on Yorkshire stimuli (p = .00779) than other groups. This finding simultaneously supports the other-accent effect: speakers find it easier to recognise their own accent (cf. Chapter 5). Moreover, listeners belonging to the Non-Brit group (i.e. the least local listeners) were significantly less accurate than all three British groups on Yorkshire stimuli (p = .007047). This finding seems to further confirm the other-accent effect.

In addition, Figure 8.14 shows a close negative association between variability in group response patterns to Yorkshire stimuli and accuracy. The narrowest range was found among the most accurate group (Yorks B&R) and the highest among the least accurate (Non-Brit). On non-Yorkshire stimuli, however, there appears to be a (rougher) positive correspondence between

accuracy and variability in response patterns. The Non-Brit group trends highest on both accuracy and variability, while the three British groups trend lower on each. This may be interpreted as a reflection that the Non-Brit group's response pattern was overwhelmingly random in character—an impression reinforced by Figures 8.6 and 8.8, which shows accuracy at around chance level.

No significant advantage in accuracy on Yorkshire stimuli was associated with belonging to the Yorks Res group. This group was specifically constructed in order to explore the validity of the familiarity effect, that residency in a particular area makes the local accent easier to identify (cf. Chapters 5 and 7). This operation of this effect appears not to be supported. Nonetheless, as implied by descriptive statistics, all three British listener groups evidence a notably higher— albeit not statistically significant—probability than Non-British listeners of correct responses to Yorkshire stimuli. This tendency seems to be graded by degree of familiarity (cf. Figure 8.14, where Yorks Res is marginally less accurate than Yorks B&R but more so than Brit Non-Yorks).

Hypothesis (ii), that local listeners will be more accurate in identifying local speakers irrespective of linguistic education, appears not to be supported. Non-linguists (including those belonging to the Yorks B&R and Yorks Res groups) were in fact significantly less accurate than linguists (p = .033) in identifying Yorkshire stimuli. One possible confound should be mentioned in this connection. This is that listener groups were unbalanced: there was no non-Brit linguist group, while groups of all three British backgrounds were divided into a linguist and a non-linguist group. As shown by descriptive results (cf. Figure 8.8), the Non-Brit group was generally less accurate than any of the other non-linguist groups. This may have skewed inferential statistical results to show greater accuracy on the part of linguists.

Results show that listeners overall were significantly more accurate on Yorkshire than on non-Yorkshire stimuli (p = <.001). However, the only significant result returned on the interaction of listener and speaker background—apart from the positive result involving Yorkshire B&R—was negative (non-Brit, p = .004). Thus it seems reasonable to infer that the majority of the effect of greater accuracy on Yorkshire speakers came from the responses of the three British groups—

Yorks B&R in particular (cf. Figure 8.14).

A final finding of interest is that there was no significant effect in connection with non-Yorkshire stimuli. As we will see, this pattern is observed also in the inferential statistical analysis of the results of Ar-3.

8.7.3.4. Listener confidence: results

As specified in 8.7.3.1, confidence is operationalised here in binary fashion. The dependent variables are 'uncertain' responses (coded as '0') and 'decision' responses (coded as '1'), regardless of degree of confidence (highly likely, unlikely etc.). The fixed effects are the same as they were in the analysis of accuracy: Yorkshire vs. non-Yorkshire stimuli, listener geographical background (Yorkshire vs. non-Yorkshire) and listener linguistic education.

*Degree*s of listener confidence (and its relationship to accuracy), incorporating a full consideration of Likert-scale ratings, is examined only in connection with the results of Ar-3 (Chapter 11), in the shape of a formal analysis of correlation. The decision not to do so in Eng-2 was motivated by the fact that a graded consideration of the two variables is especially relevant to differentiating group response patterns in LAAP-like tasks, which commonly involve Arabic but never British English.

Table 8.5 shows the coefficients and significance values for the model of best fit, as determined by glmer, for confidence of listener response. The intercept shown in Table 8.5 encodes two baseline fixed effects: 'education: linguist', and 'speaker: non-Yorks'. Unlike in the analysis of accuracy, no interaction was found between any of the fixed effects.

Table 8.5: Coefficients and significance values for glmer analysis of confidence

```
Fixed effects:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)            1.3189     0.1684   7.834 4.74e-15 ***
Educationnon-linguist  0.2004     0.1779   1.126 0.260094
SpeakerYorks           0.4481     0.1272   3.522 0.000429 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results demonstrate a significant effect (by means of predicted probabilities) on confidence for only one independent variable: speaker origin (i.e. whether the stimulus was of Yorkshire or not). No significant effect is shown for listener education. In illumination of these results, Figures 8.15 and 8.16 show in turn the predicted probability of a 'decision' response being given (i.e. confidence) associated with listener education and speaker (i.e. stimulus) origin.

Figure 8.15: Predicted probability of 'decision' response associated with, respectively, linguistic education and no linguistic education
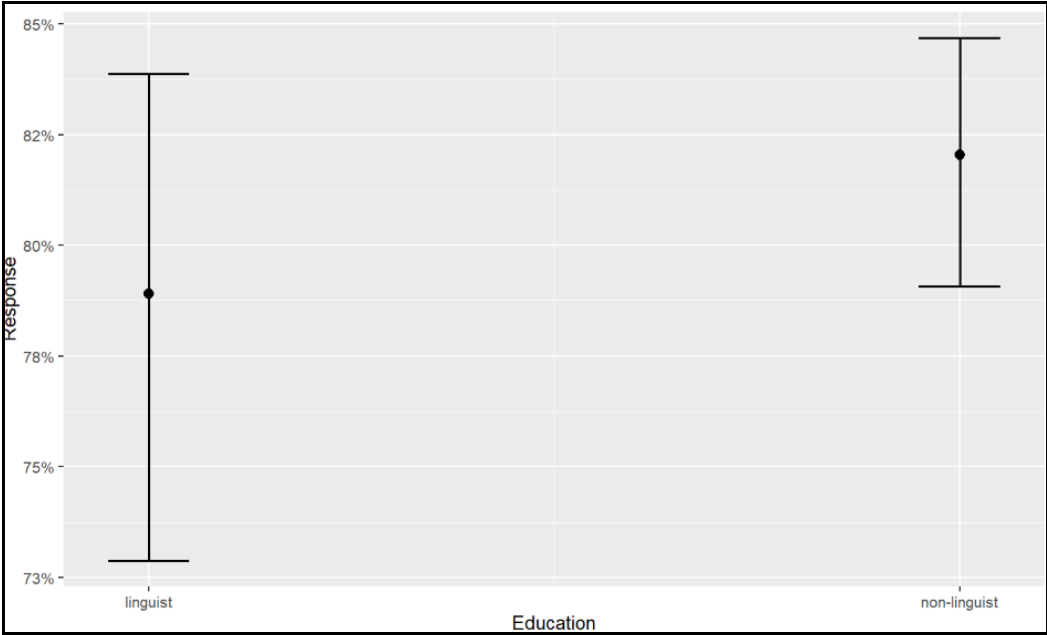
Figure 8.16: Predicted probability of 'decision' response associated with, respectively, non-Yorkshire and Yorkshire stimuli

8.7.3.5. Discussion of results: listener confidence

As discussed in Chapter 5, owing to the overall equivocal findings of previous experimental work on the subject, no prediction as to the relative confidence of listener groups was hypothesised in this thesis. However, in the LAAP context it has been argued repeatedly—adducing, almost exclusively, evidence from outside the field—that linguistic training has a tempering effect on listener confidence (e.g. Fraser 2009, 2011, 2019; Patrick 2010). The above results are therefore of interest as an empirical examination of the matter from within a LAAP-like experiment design.

Results indicate that in Eng-2 linguistic education had no significant effect on listener confidence. This finding contradicts the assertion that linguistic training engenders greater caution in the expression of conclusions (cf. Chapter 5). However, as Figure 8.15 shows, listeners with linguistic training evince greater variability in confidence than do non-linguists.

Furthermore, as in connection with accuracy, it must be pointed out that the unbalanced nature of the listener sample—whereby there was a Non-Brit non-linguist but not a Non-Brit linguist group—may have distorted results here. As indicated by descriptive statistics (cf. Figure 8.8), the non-linguist Non-Brit group was the least confident in its overall response pattern. It is at least possible that, had a non-Brit linguist group been composed for Eng-2, the probability of finding an association between linguistic training and confidence would have been increased.

Results demonstrate that all listeners were significantly more confident on Yorkshire than on non-Yorkshire stimuli ($p = <.001$). A similarly significant effect on accuracy, involving the same independent variables, was located. This suggests the operation of some sort of relationship between confidence and accuracy, on Yorkshire stimuli only. I leave this suggestion aside for investigation in future work.

8.8. Listener accuracy and confidence: concluding discussion

In summary, inferential statistical analysis of the results of Eng-2 demonstrates the following.

(1) Local listeners (i.e. the Yorks B&R group) were significantly more accurate than non-locals in identifying local stimuli (i.e. featuring speakers from Yorkshire), supporting both hypothesis (i) of the thesis and the other-accent effect.

(2) The other-accent effect is further supported by the fact that the least local group (Non-Brit) was significantly less accurate than all British groups on Yorkshire stimuli.

(3) The other locally-affiliated group (Yorks Res) showed no significant advantage in identifying Yorkshire stimuli, apparently in contradiction of the familiarity effect; all British listeners did, however, show a higher probability than the Non-Brit group of entering correct responses to Yorkshire stimuli.

(4) Linguistic education significantly affected overall accuracy on Yorkshire stimuli, though this may be due to the unbalanced nature of the Non-Brit listener sample, which included non-linguists only.

(5) Contrary to tenacious claims in the LAAP literature, linguistic education revealed no significant effect on listener confidence—although this result, too, might have been influenced by the non-presence of a Non-Brit linguist group.

(6) All listeners were significantly more confident on Yorkshire than on non-Yorkshire stimuli, suggesting a relationship of undefined kind between confidence and accuracy on responses to the former stimulus type.

For current LAAP practice, these findings carry several possible implications. First, support adduced here for the other-accent effect—and the lack of it for the familiarity effect—suggests that the selection of analysts with native knowledge of the particular (claimed) variety spoken by the applicant is essential to accurately identifying the applicant's true origin. Second, the revealed significant relationship between linguistic education and accuracy (but only among native speakers, local or otherwise, of the *same language*) indicates that the ideal LAAP analyst is likely to be a native speaker with linguistic training. Third, linguistic training of native

speakers is unlikely to lead to fewer decisions being made in LAAP, especially in cases involving truly local applicants (cf. significantly greater confidence among all listener groups on Yorkshire but not non-Yorkshire stimuli).

Results also lead to the tentative conclusion that true local native speaker-listeners would be more accurate than non-locals in a future supplementary test of perception (outlined in Chapter 1 and described in detail in Chapter 12). The significant effect of linguistic education on accuracy of responses to Yorkshire stimuli does not nullify this inference, since in Eng-2 local listeners *generally* (i.e. regardless of education) were significantly more accurate than non-locals.

Eng-2 was the final experiment in the English series; the remaining three experiments concerned Arabic. Many elements of experiment design trialled in the English series were re-used in the Arabic series, yet text and stimulus selection had, of course, to be completely reconsidered. In addition, the lack of statistically significant support revealed here for the familiarity effect prompted the re-composition of listener groups along national rather than sub-national lines. I detail both the constants in and the changes to experiment design in the following chapter.

Chapter 9: Arabic experiment 1

9.1. Introduction

The final series of experiments, including the pilot described in this chapter (henceforth Ar-1, followed by Ar-2 and Ar-3), focused on Arabic. Syrian (Damascene) Arabic was selected as the target variety, for two reasons.

First, Arabic is at present among the languages most frequently assessed in LAAP. More than one bureau reports that Arabic speakers are the subject of a greater number of LAAP referrals than speakers of any other language (Stan pers. comm. for Verified; Hubbuch pers. comm. for LINGUA). Further, Syrian origin was at the time of writing claimed more frequently than any other nationality among asylum seekers subjected to LAAP, at least at Verified (Stan pers. comm.). Eurostat (2022) statistics on asylum applications by claimed (non-EU) nationality appear to confirm these reports, though they do not specify the number of referrals to LAAP (Figure 9.1).

Figure 9.1: Eurostat statistics on asylum applications by claimed (non-EU) nationality

Top 30 citizenships of first-time asylum applicants (non-EU citizens), EU, 2020 and 2021 (thousands)

Source: Eurostat (online data code: migr_asyappctza)
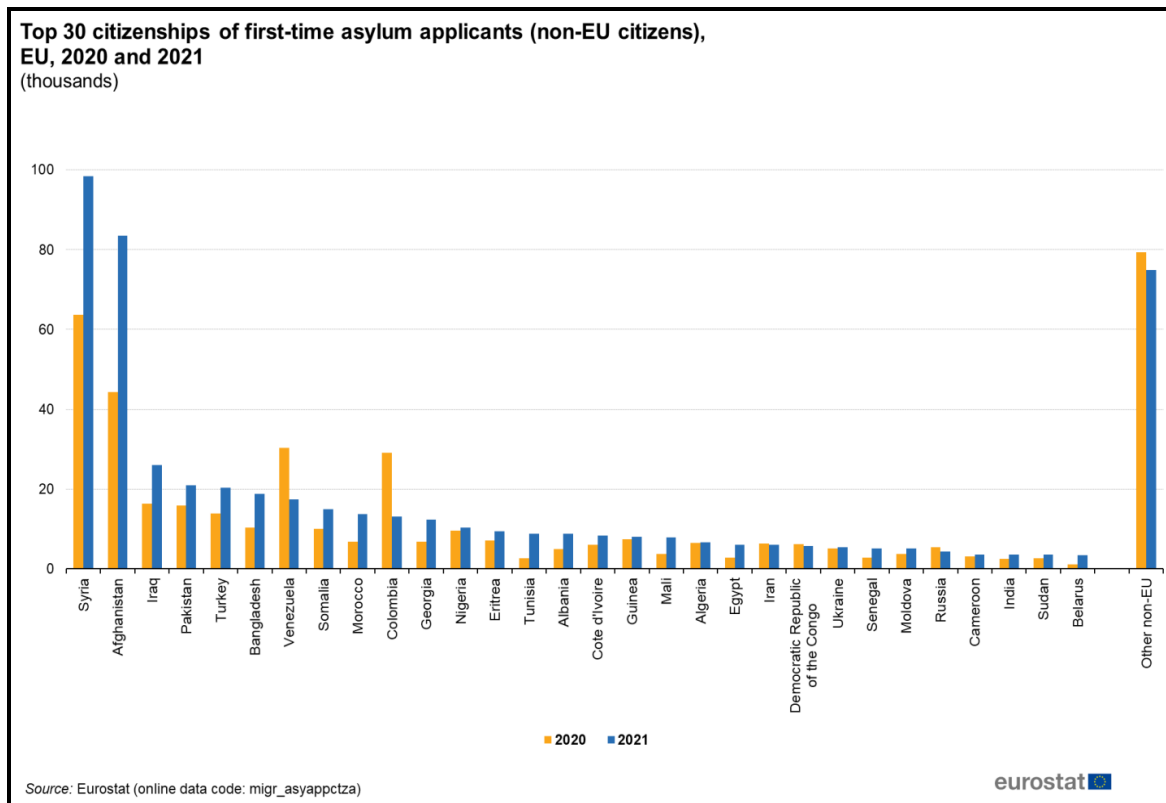
The second reason for the selection of Arabic had to do with the availability of a corpus well populated with vernacular varieties: the [Intonational variation in Arabic Corpus](#) (IVaR) (Hellmuth & Almbark 2017). This is not the case with most other languages of interest to LAAP (e.g. Dari, Kurdish, Somali). The availability of an extant corpus was especially valuable during COVID-related restrictions, when fieldwork was effectively ruled out. Damascene Arabic is the only Syrian variety represented in IVaR. This fact was not considered a limitation, since Damascene is reported to be the Syrian variety most widely known to Arabic speakers (Palva 2006, p. 604; Versteegh 2001, p. 140). Its wide recognition was expected to narrow but not eliminate the predicted gap in performance by Syrian vs. non-Syrian listeners.

As in Eng-1 and Eng-2, listeners in all experiments in the Arabic series were presented with short voice samples and asked whether and to what degree of certainty they were able to associate the samples with a particular location. In the Arabic series the question asked of listeners was, 'Is this a Syrian accent?' Chapter 7, section 7.1 explains the general reasoning

behind the phrasing of this question.

9.2. Ar-1: hypotheses

The thesis began with three hypotheses, of which only the first and third are relevant to Ar-1. They are as follows:

(i) locals (Syrians) will be more accurate than non-locals (non-Syrians) in identifying stimuli featuring fellow local speakers; (ii) the predicted pattern of superior local accuracy will not be affected by linguistic education; (iii) feature citations will vary in number or kind (or both) according to listeners' 'localness'.

9.3. Ar-1: research objectives

Ar-1 has two research objectives. As in all experiments in this series, research objective 1 involves presenting audio stimuli featuring Syrian and non-Syrian Arabic speakers to two groups of listeners—also Syrian and non-Syrian—and analysing their respective response patterns to the question, 'Is this a Syrian accent?' Research objective 1 thus represents the operationalisation of the first of the three hypotheses above and bears directly on the ultimate aim of the thesis: developing a novel perception test to be performed by asylum applicants. If significant variance in response patterns can be detected between, respectively, local and non-local speaker-listeners of Arabic, the first steps can be taken towards the development of such a test.

However, owing to the small number of listeners recruited for Ar-1 (N = 10), an inferential statistical analysis of this variance—and thus a rigorous evaluation of the extent of support for the hypotheses—is unsuitable. In addition, owing to the educational background of the listeners in Ar-1 (all were linguists) no observations are made in this chapter about the validity of hypothesis (ii).

Research objective 1 is instead approached largely as a means of narrowing down the 52 stimuli used here to a more workable total (< 25), retaining those which occasion differing response patterns (i.e. in accuracy and/or degree of certainty) among listeners. Accordingly, descriptive differences in group response patterns are evaluated in subsections 9.8.1.2 to 9.8.1.5, and the resulting decisions as to inclusion/omission of stimuli for the following experiment (Ar-2) are described in subsection 9.8.1.6.

Research objective 2 is the operationalisation of hypothesis (iii). Its investigation in Ar-1 is effected by eliciting listeners' comments, in free-text boxes, on the cues that guided their responses to the question 'Is this a Syrian accent?' In Ar-1, comparisons are made on the dual basis of: (1) the degree to which cues recognised by listeners coincide with those acknowledged in the dialectological sources and; (2) how the cues recognised vary between the Syrian and non-Syrian groups.

Research objective 2 is relevant to the thesis in two senses. The first is theoretical; several related propositions about the nature of ordinary native speakers' perceptions of language, as opposed to those of linguists, are advanced in chapter 4. The second sense is practical: Is the format of LAAP reports used by agencies fit for purpose? Do such reports accurately reflect the perceptions of NSNLs, or are their perceptions merely retrofitted into a framework which is unsuitable? Research objective 2 is an attempt to answer such questions. But its ultimate aim is to develop a second type of supplementary test for asylum applicants. This test is visualised as broadly similar to the listen-repeat kind described by Moosmüller (2011), prefigured in chapter 1 and detailed in Chapter 12—i.e. a test of production to augment the perception test associated with research objective 1. In section 9.9 I survey the features of Syrian Arabic that were most frequently commented upon, by Syrians and non-Syrians respectively. It is envisaged that such a survey—though its illumination of differences between the two groups in feature perception—might inform the composition of text(s) to be used in a production test of the type described.

9.4. Methodology

9.4.1. Task design

The basic procedure of the Arabic experimental series was identical to that of the series on English: to present listeners with voice recordings and have them judge to what degree of likelihood each speaker is from a particular location. The operative question here was, 'Is this a Syrian accent?'—fundamentally the same question as was asked in the English series. The form of this question was vetted by various academic authorities on Arabic, prior to administration of the survey, and judged satisfactory. Accordingly, task design did not vary radically between the Arabic and English experimental series.

In both series, a degree of ecological validity was secured by allowing listeners to access stimuli multiple times, as is possible in LAAP. Further task design features of the English series had proved functional (use of Qualtrics), relatively easy for listeners to interpret (the phrasing of the main question as 'Is this a [local] accent?') and were judged likely to attract a comparatively large number of completed responses while making the task sufficiently challenging for listeners (stimulus duration, at c. 10 seconds). Free-text comment boxes in the English experiments permitted some insights into response patterns; this fact motivated the retention of comment boxes in the Arabic series, with some modification of wording.

As in Eng-2, all experiments in the Arabic series employed a five-point Likert scale—from 'highly likely' to 'highly unlikely'—followed by a free-text comment box headed by the question, 'What clues led you to your judgement of the speaker's accent?'

All text in Ar-1 was in English; Arabic translations were used in Ar-2 and Ar-3. Figures 9.2 and 9.3 display these questions in the same manner as they appeared in the survey.

Figure 9.2: Display of main question with Likert scale

Is this a Syrian accent?
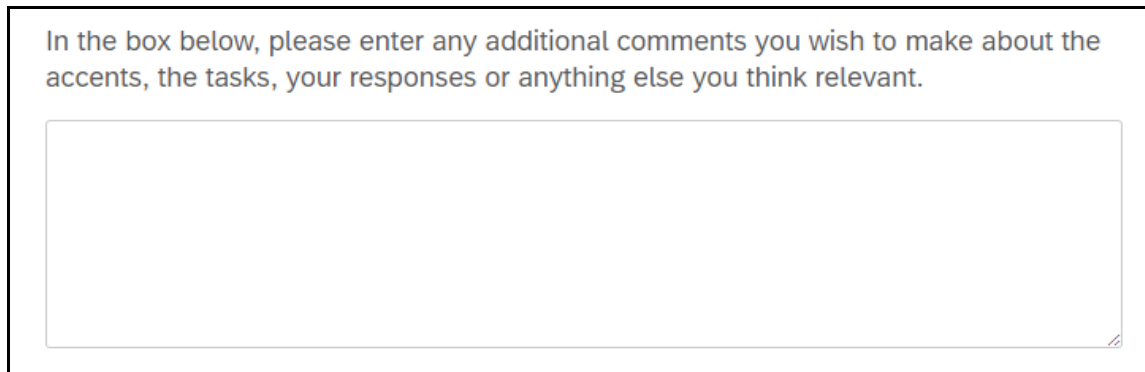
	Highly likely	Likely	Uncertain	Unlikely	Highly unlikely
To the right, please indicate your degree of certainty	○	○	○	○	○

Figure 9.3: Free-text box 1

What clues led you to your judgement of the speaker's accent?

Additional comments were elicited in a free-text box at the end of the survey. Figure 9.4 displays it as in Qualtrics.

Figure 9.4: Free-text box 2

In the box below, please enter any additional comments you wish to make about the accents, the tasks, your responses or anything else you think relevant.

The tasks in the Arabic series were thus conceptually similar to those in the English experiments, where the question was, 'Is this a Yorkshire accent?' followed by a free-text box headed by the question 'If your answer above was some degree of 'unlikely', what kind of accent do you think the speaker has?' The change in the substance of the latter question was motivated by the aim of obtaining a broader view of the grounds on which respondents made their decisions (i.e. on accents both rejected and accepted as Syrian).

9.5. The dialectology of Arabic and possible difficulties with aspects of task design

It may be objected that the phrasing of the question, 'Is this a Syrian accent?' could have caused confusion among respondents. This is because Syria is home to a range of dialects belonging to various Arabic subfamilies. Most widely spoken are the Levantine dialects, a continuum of related varieties prevalent in western Syria, Palestine, western Jordan, Alexandretta/Iskenderun/Hatay (in coastal southeastern Turkey) Cilicia (southern central Turkey) and Lebanon (Behnstedt 2012; Behnstedt &Woidich 2013). Levantine dialects, like those of most other modern Arabic dialect groups, are of the sedentary (as opposed to Bedouin) type.

Sedentary varieties are conventionally divided into two further categories: urban and rural. Levantine urban dialects have generally replaced interdentals with alveolar or dental equivalents and unvoiced uvular stops/voiced velar stops with their unvoiced glottal equivalent, although this has not occurred in Jordan. In many rural varieties the former in each pair has been retained, with the retention of either /g/ or /q/ varying among them (Behnstedt&Woidich 2013).

Central southern Syria, meanwhile, is home to Bedouin Najdi Arabic (or North Arabian), variants of which are also spoken in Iraq, Saudi Arabia and some areas of Jordan (Behnstedt 2012; Versteegh 2001). Shawi Bedouin dialects, marked *inter alia* by the occurrence of the voiced velar stop rather than the unvoiced uvular stop of other Bedouin-type dialects, are spoken among (semi-)nomadic shepherds and others in various parts of Syria, especially but not only in the desert/steppe regions east of the Euphrates, bordering western central Iraq (Procházka 2018). Bedouin dialects of both types are present also in ecologically similar parts of the Arab world (Behnstedt 2012; Versteegh 2001).

The picture is further complicated by the presence of conservative, Bedouin-like features in various sedentary dialects. In inner northeastern Syria, Iraqi Mesopotamian dialects, closely related to those of neighbouring northwestern Iraq, are found (Behnstedt 2012; Versteegh 2001). Meanwhile, in outer northeastern Syria, northwestern Iraq and southeast inland Turkey, Anatolian, or North Mesopotamian, varieties—another subdivision of the Mesopotamian dialects—are spoken (Jastrow 2012a). The Anatolian dialects are referred to by linguists as *qeltu* and the Iraqi dialects as *gilit* varieties. These last two designations refer the presence of either the unvoiced uvular or the voiced velar stop—both Bedouin features—as well as the presence or absence of the marker of the first person singular perfective, /tu/ (Jastrow 2012a).

Other sedentary varieties, too, commonly evidence a mixture of innovative and Bedouin-type characteristics (Behnstedt & Woidich 2013). This mixture occurs in the Syria-Jordan cross-border upland region known as the Horan, on both sides of which a distinctive Horani dialect is spoken (Behnstedt 2012). Horani Arabic displays a number of Bedouin-type features (interdental fricatives, for example). Some sedentary varieties in Syria—Horani and rural

Palestinian Arabic among them—share some or all of these conservative features with both Najdi Arabic and Gulf Arabic, the latter group comprising dialects spoken by many citizens of Kuwait and other states in the east of the Arabian peninsula (Holes 2018).

In summary, then, the terms 'Syrian', 'Levantine', 'Syro-Lebanese' and 'Greater Syrian' do not refer to dialects spoken only within Syria's borders; neither are they the only forms of Arabic spoken by Syrian citizens.

Figure 9.5 shows the geographical distribution of Arabic varieties spoken in Syria; those designated 'A' to 'Z' are Levantine. Figure 9.6 shows the situation of the Levantine group relative to non-Levantine varieties. Note that, reflecting the varying schema imposed by dialectologists, the two maps do not agree completely in the location/classification of non-Levantine varieties.

Figure 9.5: Geographical distribution of Arabic varieties spoken in Syria (Behnstedt 2012)

Map 1. Syria: Languages and dialects.
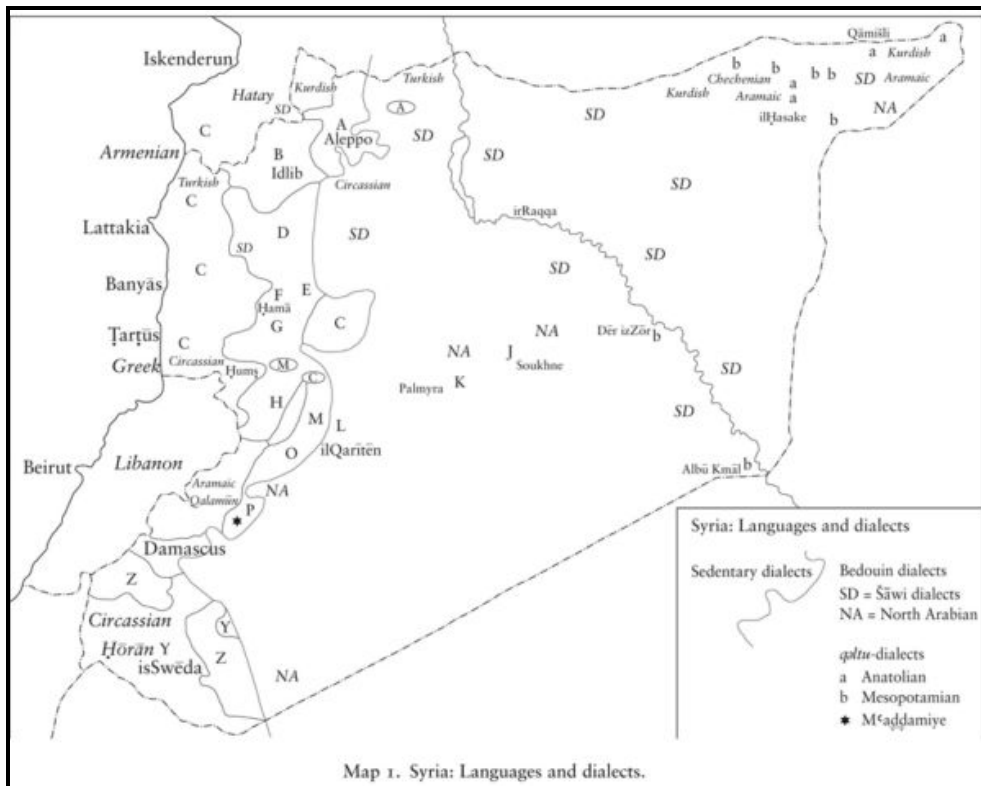
Figure 9.6: Geographical situation of Levantine Arabic varieties, in light green (Wikimedia Commons contributors 2022)
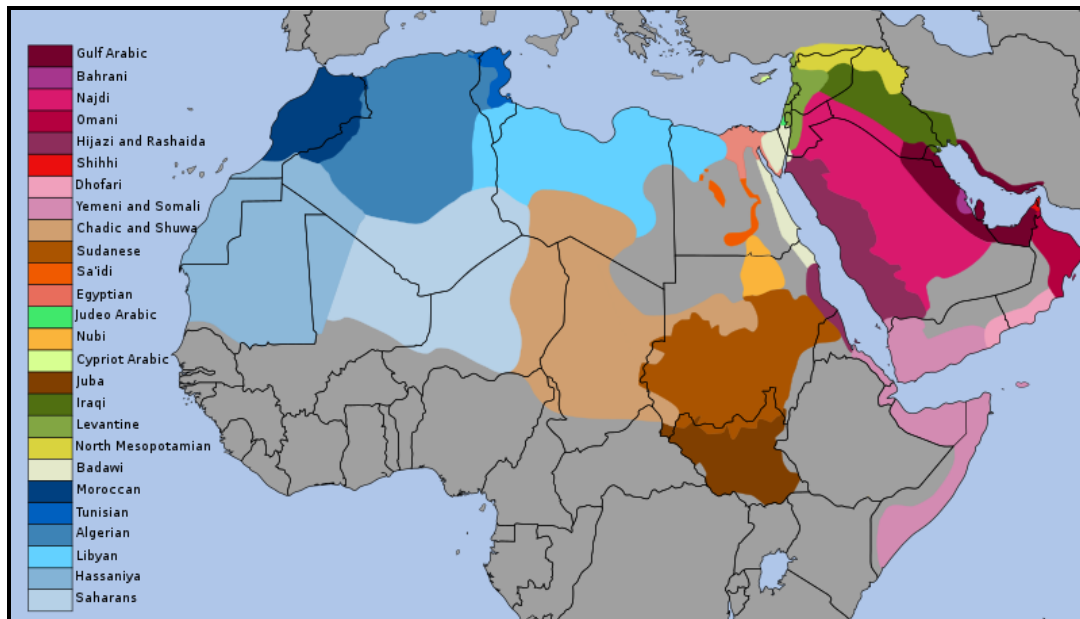
Even from this brief and necessarily simplified description, the potential for confusion as to the meaning of the term 'Syrian accent' in the survey question should be clear. Is the question asking whether the stimulus at issue is a distinctively Syrian example of the Levantine dialects, or is it asking whether it belongs among varieties spoken in Syria as a political entity, which encompasses a much wider range of Arabic varieties?

On the one hand, comments from Syrian and non-Syrian listeners, linguists in Ar-1 and NSNLs in subsequent experiments alike, suggest that 'Syrian accent' was understood—in line with Versteegh's (2001) statement on the status of Damascene Arabic in Arabophone lay dialectology—to refer principally to the Levantine varieties of western Syria proper, including that of Damascus. On the other hand, as reviewed in subsection 9.8.1.7, listener comments—particularly by Syrians—indicate a strong awareness of the existence of cross-border varieties and the related fact that the designation 'Syrian' can be construed to include dialects from outside the (western Syrian) Levantine group (this is noted in Behnstedt 2012). As discussed in subsection 9.8.1.5, this awareness may be reflected in the frequency in results of false acceptances: the misidentification of e.g. Jordanian, Kuwaiti and Iraqi stimuli as Syrian.

Nevertheless, listeners' generally strong performance across all three Arabic experiments in correctly identifying Syrian stimuli suggests that the question was interpreted as conceived— that is, something like, 'Is this an accent of a speaker of (Syrian) Levantine Arabic?'

9.6. Listeners

9.6.1. Recruitment

Listeners for Ar-1 were recruited through contacts at the universities of York and Newcastle. Ten native-speaker listeners from various Arabic-speaking countries responded. Two listeners declared Syrian citizenship; there were in addition two Jordanians, one Egyptian, four Saudis and one Libyan. This cross-section of listeners was not ideal, since it included only two Syrians. There were also listeners from countries (Saudi Arabia and Libya) which according to LAAP agencies infrequently produce asylum applicants (Cambier-Langeveld, pers. comm.; Hubbuch, pers. comm.; Stan, pers. comm.).

However, since there is no known evidence that Arabic speakers of any non-Syrian nationality find it easier or harder to distinguish among Syrian and other Arabic accents, it was thought unnecessary to expend the time and effort required to seek listeners from a more representative cross-section of Arabic-speaking countries. Thus the invitation to participate in the Arabic series was not restricted to listeners of particular national backgrounds.

Several preliminary questions were asked to permit the classification of listeners into groups. Following completion of a section on consent, listeners were first asked whether they were native speakers of Arabic, then about their country of origin (i.e. of what country they were born a citizen). Following this, in a free text box, they were asked to specify any periods, with approximate dates, of residence outside their country of birth/citizenship. Finally, they were asked whether they had studied linguistics at postgraduate level.

Figures 9.7, 9.8 and 9.9 display in turn the questions asked to ascertain listeners' native speaker

status, national origin/citizenship/residential history and level of training in linguistics. All listeners surveyed responded in the affirmative to the latter question.

Figure 9.7: Question asked to ascertain listeners' linguistic background

Are you a native speaker of Arabic?

○ Yes

○ No

Figure 9.8: Question asked to ascertain listeners' national origin/citizenship and periods/ approximate dates of residence outside country of birth

Of which country were you born a citizen? Have you ever lived outside your original country of citizenship? If so, please provide locations and approximate dates below.

Figure 9.9: Question asked to ascertain listeners' educational background in linguistics

Have you studied linguistics at postgraduate level?

○ Yes

○ No

9.6.2. Classification of listeners

9.6.2.1. Linguists vs. NSNLs

For Ar-1, only responses from Arabic native-speaker linguists (i.e. not from NSNLs) were sought. This decision was made because it was predicted that native-speaker linguists would have a better grasp of the metalanguage necessary to express their insights about the cues informing their decisions. This would in turn permit a better-informed initial exploration of research objective 2 (section 9.9). Furthermore, since the intention of subsequent experiments was to recruit NSNLs, using only linguists in this experiment would serve to reserve a larger potential pool of NSNL respondents for future recruitment.

The definition of 'linguist' in the Arabic experimental series differed from that in the English series. In Eng-2, linguists were classified as listeners who declared training in British dialectology to secondary level or above; in Ar-1, they were defined as listeners with postgraduate qualifications in linguistics (as Figure 9.9, above, indicates). This change was made largely because of the wording of the third of the *Guidelines*, which states,

> "...expertise can be evidenced by holding of higher degrees in linguistics, peer reviewed publications, and membership of professional associations."

(LNOG 2004, p. 262).

This passage implies, though it does not explicitly state, that undergraduate qualifications in linguistics were not considered sufficient by the authors of the *Guidelines*.

Wilson (2009, 2016) found that undergraduates were not significantly less accurate than others in identifying Ghanaian English speakers. She nonetheless argues (2016, p. 32) that

> "Undergraduate linguistics courses do not necessarily provide students with enough practical phonetic work for them to be able to distinguish between small differences in speech sounds in the same way that a postgraduate course would"

In order to investigate the accuracy advantage yielded by postgraduate training postulated by the *Guidelines* and in Wilson 2016, only listeners holding qualifications beyond undergraduate level were recruited for Ar-1.

9.6.2.2. National vs. regional origin

In the English experiments, listeners were divided into groups at the regional (i.e. historical county) as well as national level, with stimuli selected according to a similar degree of resolution. In the Arabic series, listeners were stratified and stimuli selected solely according to national origin. This difference in listener classification and stimuli selection was motivated by two related considerations, both derived from the results of Eng-2.

First, a significant difference in accuracy was found between; (1) Yorks B&R (i.e. 'locals' in the narrowest sense) and other listeners, but only on Yorkshire stimuli; and (2) non-Brit and other listeners, also only on Yorkshire stimuli. Second, no significant advantage in accuracy was associated with the 'familiarity effect' (i.e. among the Yorks Res group). Together these results suggested that the fourfold geographical classification of listener groups essayed in the English series would not produce significant intergroup differences in performance in experiments on Arabic. It was thus considered that a simple two-way distinction between local and non-local listeners would be sufficient. For the Arabic series the local/non-local distinction was made at the level of nationality. This decision had to do with two factors not connected to the results of Eng-2, as follows.

One was the difficulty of obtaining a broadly representative range of within-country Syrian stimuli. The IVaR database, the best source of stimuli available for use during COVID-related fieldwork restrictions, at the time of research contained stimuli only of Damascene speakers (i.e. not of other Syrian dialects). This meant that a comparison of stimulus recognition at the within-country level could not be made, further reinforcing conclusions reached on the basis of Eng-2.

Another was my interpretation of British government policy on LAAP. This states that, for a

claim to be considered credible without adequate documentation, an asylum applicant's "dialect or accent" should be consistent with "their claimed country or region of origin" ([Home Office 2018](#), p. 8). It is not known whether similar criteria obtain among the immigration agencies of other countries which use LAAP. Indeed, even statements on this score obtained from Home Office representatives (Kirk, pers. comm.) are vague. But it appears that, for the British government at least, establishing or refuting an asylum applicant's likely linguistic origin in Syria—and not necessarily a particular location within the country—may be sufficient. This again appeared to justify the sorting of listeners by Syrian/non-Syrian origin.

For these three reasons, then, the allocation of listeners to regional/within-country categories, and the selection of stimuli representing a range of within-country dialects, was abandoned in the Arabic experiments.

9.7. Stimuli

9.7.1. Selection

Ar-1 featured 52 stimuli. Of these, 42 were of native Arabic speakers, six from each of Egypt, Iraq, Jordan, Kuwait, Morocco, Oman and Tunisia; 10 were of native Arabic speakers originally from Syria, specifically Damascus. All speakers in the IVaR database were between 18 and 30 years of age at the time of recording, in the early to mid 2010s, on location in the Middle East. All Syrian speakers in the stimuli were originally from Damascus but resident in Jordan at the time of data collection. Speakers of non-Syrian varieties were resident in their respective countries of origin.

The initial 52 stimuli were selected to include a cross-section of speakers from all eight of the countries featured in the IVaR database. The greater number of Syrian stimuli was motivated by its selection as the 'target' variety. An even number of stimuli featuring male and female speakers was selected. Unlike in Eng-2, all stimuli were presented in pseudo-random order. This was done in the interests of eliminating any possible order effects, even though none had been revealed by inferential statistical analysis of the results of Eng-2.

In line with Eng-2, which attracted a higher completion rate than Eng-1, stimuli were edited for Ar-1 to c. 10 seconds in duration (average 10.5 seconds). All stimuli featured the same selection of read text. The orthography of the text, an extract of an Arab folktale about the character Juha, represented as closely as possible the dialect of each of the eight countries, with phonetics/phonology lexis and morphosyntax native to it. This, combined with control exercised during data collection over style-shifting, meant that interference via code switching with Modern Standard Arabic, herein MSA, was not possible (on code switching in Arabic see e.g. Bassiouney 2018; Bentahila et al. 2013). Following the distinction between the two terms drawn by Foulkes & Hughes (in press), 'Phonology' is used here as distinct from 'phonetics' because a number of Arabic dialects differ such that some sounds (e.g. interdentals) are present in some varieties but entirely absent from others. Such distinctions are denoted by the term 'phonology', while articulatory differences in the realisation of the same allophone are 'phonetic'.

Whereas the texts used in the English experimental series concentrated on a fairly subtle range of phonetic variation (i.e. accent in the narrow sense), the *Juha* text was selected for the Arabic series because of the presence in it of a wide range of phonetic/phonological, morphosyntactic and lexical cues. It was expected that this broadening in the dimensions of variability would serve assist listeners in distinguishing the featured seven non-Syrian dialects from Syrian (Damascene) Arabic, the equivalent task having perhaps been excessively difficult in the English series.

Appendix C to this thesis present the scripts used in the readings performed by speakers from each of the eight countries. In the left column is a transliteration in broad IPA, with the original Arabic and an English translation in the remaining two columns. It should be noted that, though in no instance does a speaker deviate significantly from the realisations specified, the stimuli feature some degree of phonetic and even lexical variability among speakers of the same nationality. IPA transliterations are based on those provided by Almbark (pers. comm.), one of the two linguists who worked on the IVaR corpus.

9.7.2. The dialectology of Arabic and its bearing on stimuli selection

The dialectological literature reveals that the linguistic situation in the Arabic-speaking world is highly fragmented (e.g. Al-Wer et al. 2015; Behnstedt 2012; Behnstedt&Woidich 2013; Bentahila et al. 2013; Holes 2018; Khattab & Foulkes in press; Owens 2001, Versteegh 2001). Arabic dialects are conceptualised as diverging and intersecting according to some or all of multiple parameters: geography, urban vs. rural, diglossia, codeswitching, lifeways (e.g. Bedouin), tribal identity and ethno-religious background (e.g. some Shi'ites, the Druze, the Alawites, some Christians and, at least prior to their wholesale migration to Israel, the Iraqi and Maghrebi Jews). In Syria, the situation is further complicated by the occurrence of varieties across national borders (Syria-Lebanon, Syria-Jordan, Syria-Iraq, Syria-Turkey).

There are nevertheless a number of well-known shibboleths that identify speakers with particular countries or places within them, e.g. the Cairene realisation of Classical Arabic /ʤ/ as /g/. Some varieties—e.g. Beiruti, Damascene and (especially) Egyptian Arabic—have further the status of de facto national or international pop-cultural standard varieties, and they are all the more widely recognised for it. For this last reason it was anticipated that at least some of the Syrian stimuli selected, all of which feature speakers originally from Damascus reading in a closely controlled version of their native variety, would permit comparatively easy recognition by all respondents—with the prediction, in line with hypothesis (i) of the thesis, that Syrian listeners would be more accurate in identifying Syrian speakers than would non-Syrians.

9.7.3. Predicted diagnostic features

Behnstedt (2012) argues that features unequivocally distinctive of Syrian Arabic are difficult to define. He numbers among them 'intonation combined with certain pausal features', *imāla* (i.e. vowel-raising under phonotactic constraints) and personal pronouns. The experiments on Arabic conducted for this thesis, based as they were on a short sample of read text, presented only a few of these features.

Even within this limited range, considerations of space preclude an exhaustive description of all features of all dialects which are present in the stimuli. Instead, I concentrate on features

occurring in the stimuli which, according to a survey of the dialectological literature, are typical of the seven non-Syrian dialects of Arabic and which may have served to differentiate Syrian Arabic from them (and vice versa).

Tables 9.1 and 9.2 summarise and contrast in turn the segmental phonetic/phonological and morphosyntactic features covered in a survey of the available sources. Features which distinguish a *single* (non-Syrian) dialect, not just from Damascene Arabic but from all other varieties—e.g. the distinctively Moroccan genitive marker /dial/ (Aguade 2018, p. 54)—are not included in Tables 9.1 and 9.2.

For each feature in these two tables, reflexes of the MSA realisation are compared. Relevant suprasegmental properties are described, together with results connected with research objective 2, in section 9.9. Lexis is rarely described in cross-dialectal comparative terms. No predictions were therefore made as to its salience, despite obvious lexical differences in the eight featured dialects. It is not considered here, excepting where it arises in connection with other classes of feature. However, some listeners cited lexical items in their own right; these are covered in subsection 9.9.5.

Table 9.1: Summary of phonetic/phonological features and their location in the Juha text

Dialect/feature & text reference	MSA /q/ (lines 1, 4)	MSA /ʤ/ (lines 2, 3)	MSA /ð/ (in the word for *they*) (line 3)	Final imāla (lines 2, 5)
Egyptian	/ʔ/, /ʔ/	/g/	/du:l/	-
Iraqi	/g/, /q/	/ʤ/	/ðo:la/	-
Jordanian	/g/, /g/	/ʤ/	/haðo:l/	/-e/
Kuwaiti	/g/, /ʔ/	/ʤ/	/haðo:la/	-
Lebanese	/ʔ/, /ʔ/	/ʒ/	/hajdo:l/	/-e/
Moroccan	/g/	/ʒ/	-	-
Omani	/g/	/ʤ/	/haðo:la/	-
Syrian (Damascus)	/ʔ/, /q/	/ʒ/	/hado:l/	/-i/
Tunisian	/q/, /q/	/ʒ/	/haðu:/	-

Table 9.2: Summary of morphosyntactic features and their location in the Juha text

Dialect/feature & text reference	3MSg marker (line 1)	preposition *in* (line 2)	Adj. Neg. (in *not good/(very) bad*) (line 3)	3PL marker (lines 4, 5, 6)	future/imperfective(?) (line 6)
Egyptian	-/u/	/l/-	/wiʃi:n (2awi)/*	-/u/	/haj/-
Iraqi	-/a/	/bil/-	/(kul:iʃ)* mu: ze:ni:n/	-/un/	/raħ/
Jordanian	-/u/	/fil/- or /bil/-	/miʃ kwajsi:n/	-/u/	/raħ/
Kuwaiti	-/a/	/fil/-	/mu: ze:ni:n (killiʃ)/*	-/un/	/raħ/
Lebanese	/-u/	/bil/-	/ma mni:ħ/	-/u/	/raħ/
Moroccan	-/u/	/l/-	/welad lħram/	-/u/	/kaj/-
Omani	-/a/	/l/-	/wa:jid (xa:jsi:n)/*	-/un/	/baj/-
Syrian (Damascus)	-/u/ or -/o/	/bil/-	/mu: mnaħ/	-/u/	/raħ/
Tunisian	-/u/	/fil/-	/ma: ra:hum (ħle:lif)/*	-/u/	/taw:a/

Notes to Table 9.2:

- /kaj/- may be identifiable with the description by Eid (2012) of a 'preverbal particle', /ka/ or /ta/, employed in "Moroccan koine" to express the imperfective aspect.

- Words in parentheses and marked with an asterisk are intensifiers, with the broad meaning of 'very'. Thus Egyptian Arabic, for example, does not employ adjective negation; instead it intensifies the adjective *wiʃi:n* ('brutal'/'monstrous') by means of *ʔawi* ('very').

- One Jordanian female Jordanian speaker uses /fil/-, while the others use /bil/-. This suggests variability ('structured heterogeneity' in Labovian terms; cf. Chapter 4) in the form of this morpheme in Jordanian Arabic.

9.8. Research objective 1: results

9.8.1. Organisation

This section details results related to research objective 1: How and to what extent did the response patterns of Syrians and non-Syrians differ when confronted with the question, 'Is this a Syrian accent?' 'Decision' responses are shown at closer resolution than they were in the two

English language experiments: 'highly likely', 'likely', 'highly unlikely' and 'unlikely' are not simply categorised as correct or incorrect responses but are assessed independently. Consistent with the fact that Arabic is relevant to LAAP in a way that British English is not, the aim here is to discern further group differences in response patterns than were perceptible from the broader approach taken in Eng-1 and Eng-2. Owing to the small number of listeners and the fact that this experiment was a pilot, results are represented only in the form of descriptive statistics.

Subsection 9.8.2 deals with the two groups' response patterns to Syrian stimuli and non-Syrian stimuli respectively. Subsections 9.8.3 and 9.8.4 detail group response patterns by *individual* stimuli—first Syrian, then non-Syrian. Moroccan, Omani and Tunisian stimuli are omitted from consideration because they were not subject to false acceptances by any listener, Syrian or non-Syrian.

In subsection 9.8.5, I discuss overall trends observable in these results. Decisions as to the retention or omission of stimuli for Ar-2 are detailed in subsection 9.8.6. In section 9.8.7, I examine the raw data, in combination with listener comments, for what it reveals about Syrians' and non-Syrians' respective knowledge of the dialectology of Syria.

9.8.2. Syrian vs. non-Syrian stimuli

Figure 9.10 compares group response patterns to, respectively, all Syrian and non-Syrian stimuli.

Figure 9.10: combined group response patterns to Syrian and non-Syrian stimuli
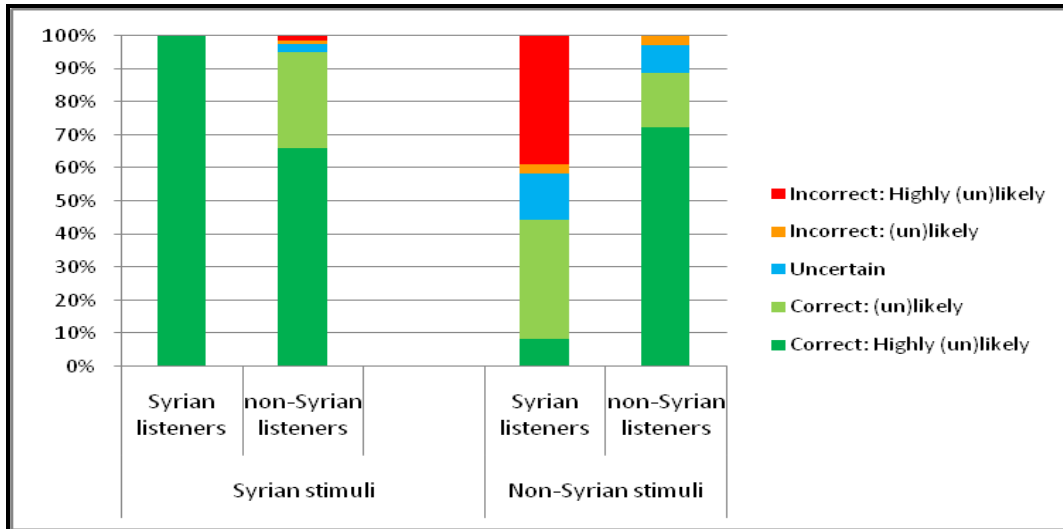
On Syrian stimuli, Syrian listeners entered 100% correct acceptances at the highest degree of certainty ('highly likely'). Non-Syrians were also very accurate, at c. 95%, though they were less certain in their correct acceptances (c. 30% 'likely') and entered a small percentage of incorrect and uncertain responses (< 5% combined).

On non-Syrian stimuli, Syrians evidence below chance-level accuracy (c. 45%) as well as low confidence in their correct rejections (< 10% 'highly unlikely' vs. c. 35% 'likely'). They also entered notably more uncertain responses than non-Syrians (c. 15% vs. < 10%).

Non-Syrians were considerably more accurate than Syrians on non-Syrian stimuli (c. 90% correct rejections). Non-Syrians were nevertheless slightly more accurate on Syrian than on non-Syrian stimuli. Aside from correct rejections, the majority of non-Syrians' responses to non-Syrian stimuli were of the 'uncertain' type, with a negligible percentage of false acceptances.

9.8.3. Syrian stimuli

Figure 9.11 illustrates response patterns, to Syrian stimuli only, by non-Syrian and Syrian listeners respectively.

Figure 9.11: Group response patterns to Syrian stimuli
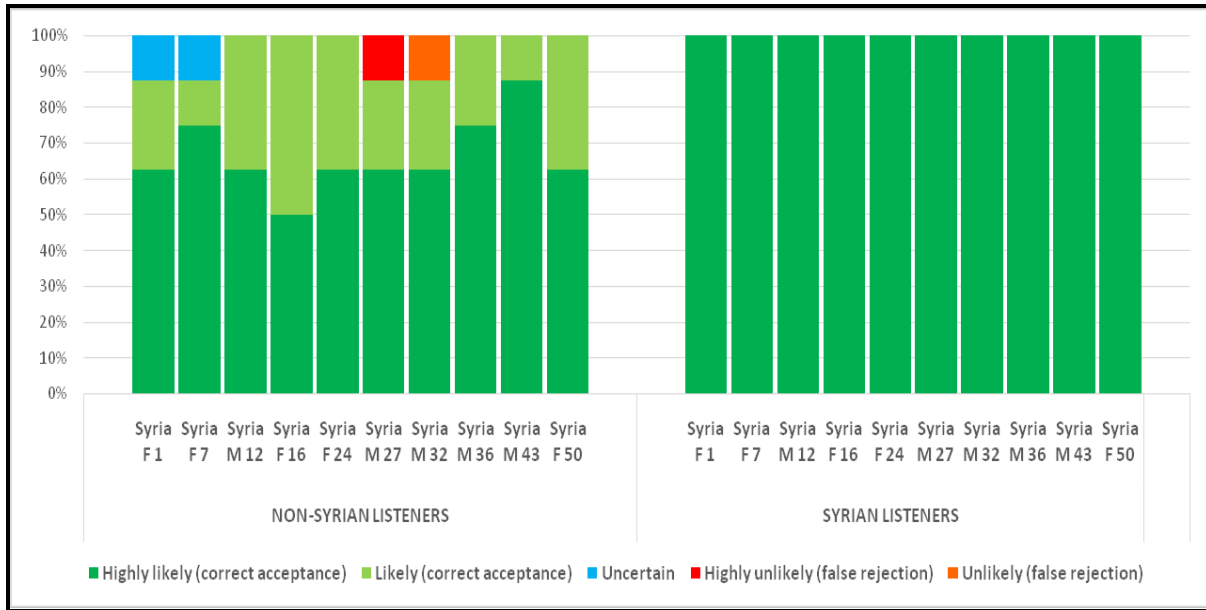
As already noted, both of the Syrian listeners correctly accepted all stimuli to the maximum possible degree of certainty (i.e. 'highly likely'). Some non-Syrians falsely rejected two of the 10 stimuli and were uncertain of the provenance of two others. The remaining six stimuli were correctly accepted by all non-Syrians, but the degree of certainty was less uniform than in the case of Syrian listeners: no individual stimulus attracted solely 'highly likely' responses.

9.8.4. Non-Syrian stimuli

Figure 9.12 shows the response patterns of non-Syrian and Syrian listeners, respectively, to non-Syrian stimuli.

Figure 9.12: Group response patterns to non-Syrian stimuli
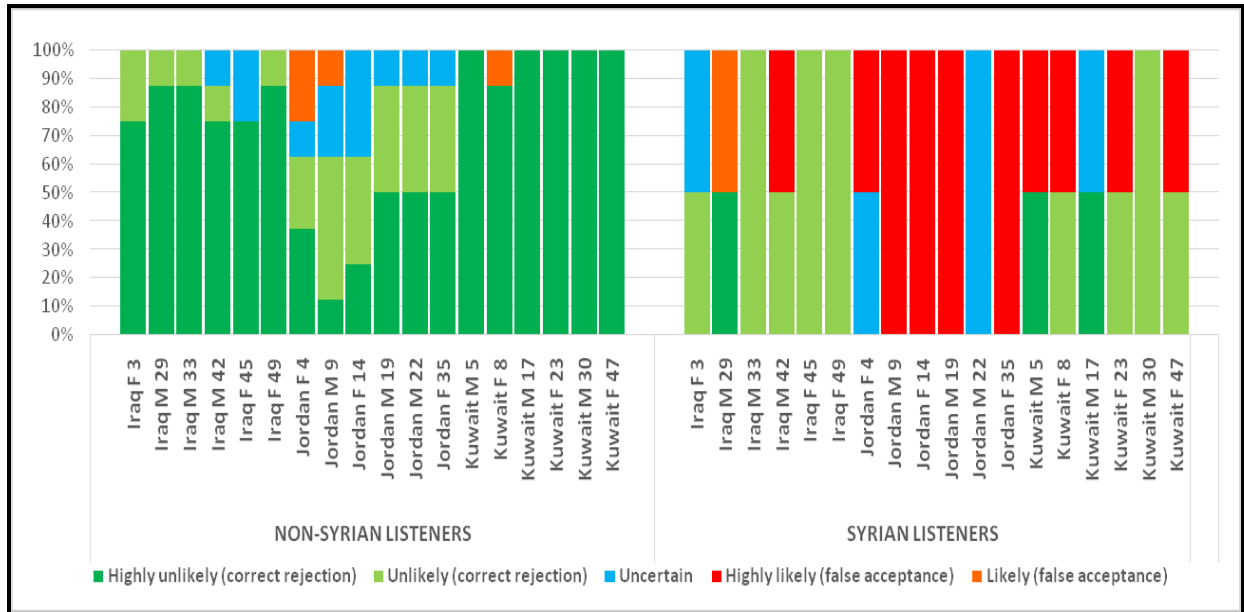
Both of the Syrian listeners falsely accepted four of the six Jordanian stimuli, to the highest available degree of certainty (i.e. 'highly likely'). A fifth Jordanian stimulus attracted a mixture of 'highly likely' and 'uncertain' responses, and the sixth solely 'uncertain' responses.

Non-Syrians, too, had greater difficulty in correctly rejecting Jordanian stimuli than they did stimuli of any other national origin. None of the six stimuli prompted correct rejections by all listeners, with 'uncertain' responses occurring in all instances. Two of the Jordanian stimuli occasioned in addition some false acceptances, though none to the degree of 'highly likely'.

One of the two Syrian listeners falsely accepted two of the six Iraqi stimuli. A third Iraqi stimulus occasioned an 'uncertain' response from one of the two listeners. The rest were correctly rejected, though with less than complete certainty (i.e. 'unlikely') in each case. None of the non-Syrian listeners falsely accepted any of the Iraqi stimuli, and they were generally fairly certain in their correct rejections (at least 75% 'highly unlikely'), though two Iraqi stimuli prompted 'uncertain' responses.

Four of the six Kuwaiti stimuli were falsely accepted by one of the Syrian listeners. Of the remaining two Kuwaiti stimuli, one was subject to one 'uncertain' response and one correct rejection and the other correctly rejected by both listeners. Non-Syrians correctly rejected five of the six Kuwaiti stimuli to the highest possible degree of certainty. The sixth gave rise to a false acceptance by a single listener. Syrian and non-Syrian listeners all correctly rejected, to the highest degree of certainty, each of the Egyptian, Moroccan, Omani and Tunisian stimuli.

9.8.5. Discussion

There are four principal findings here. First, both listener groups had particular difficulty in correctly rejecting Jordanian stimuli; comparisons of group performance on stimuli of this type are of interest to later experiments in this series.

Second, Syrian listeners falsely accepted or were uncertain about the origin of non-Syrian stimuli much more frequently than non-Syrian listeners. Syrians also evidenced lesser confidence than non-Syrians on stimuli of this type, as measured by 'uncertain' responses.

Third, Non-Syrians were highly accurate on both kinds of stimulus, though they were more so on the Syrian type. They also entered fewer 'uncertain' responses to non-Syrian stimuli than did Syrians, and were more confident in their 'decision' responses.

The fourth and most interesting finding is that in no instance did Syrians falsely reject Syrian stimuli or express uncertainty to any degree about their origin. This latter finding supports hypothesis (i) of the thesis, that local listeners will better recognise fellow local speakers.

The support for hypothesis (i) adduced from the inferential statistical results of Eng-2, wherein the Yorkshire B&R group was significantly more accurate than others in correctly accepting Yorkshire stimuli, is thus reinforced by the raw data in Ar-1. The obvious caveats about the low number of respondents and the almost equally accurate performance of non-Syrians on Syrian stimuli weaken but do not negate this conclusion.

Inter-listener response variability was fairly low within each group. On non-Syrian stimuli, all non-Syrian listeners except two (an Egyptian and a Jordanian) entered at least one 'uncertain'

response, while three listeners (not including the aforementioned Jordanian or Egyptian) entered at least one false acceptance. One of the two Syrian listeners entered four 'uncertain' responses and 10 false acceptances to non-Syrian stimuli, while the other produced one 'uncertain' response and five false acceptances.

Interestingly, on Syrian stimuli, the abovementioned Egyptian and Jordanian listeners were the only non-Syrians to enter false rejections (one each) or 'uncertain' responses (two by the Jordanian, none by the Egyptian). See Chapter 10, subsection 10.5.5, where this pattern, among individuals, of high accuracy on non-Syrian stimuli combined with relatively low accuracy on Syrian stimuli is repeated. Both Syrian listeners entered only correct acceptances of Syrian stimuli.

It is of course notable that the Arabic speaker-listeners in this experiment (and in this series generally) were far more accurate than the English speaker-listeners in the two preceding experiments. This may have had mainly to do with the selection of stimuli here along national rather than, as in Eng-1 and Eng-2, within-country lines. Differences between Yorkshire and other northern English dialects are much less pronounced than those between, e.g. Syrian and Jordanian or (still more) Syrian and Moroccan or Tunisian varieties of Arabic, and the tasks in the English experiments may therefore have been more difficult.

A further observation of interest is that the response pattern of Syrian listeners— unsurprisingly, given their 100% accuracy rate on Syrian stimuli—does not fully accord with the trend encountered in the English series and in Wilson (2009), where all listener groups recorded a higher percentage of false acceptances than false rejections (except, in Eng-2, Non-Brit). Non-Syrians here, however, show an approximately equal inclination towards both types of error, with a much greater percentage of uncertain responses (indicative of generally lower confidence) to non-Syrian than to Syrian stimuli.

The patterns discussed here, however, generally comport with those of the English series. Local listeners are more accurate and more certain on local stimuli than are non-locals (thus providing evidence, further to that gleaned from Eng-2, of the validity of the other-accent effect), whereas on non-local stimuli the advantage in accuracy is inverted. As chapters 10 and

11 demonstrate, succeeding experiments on Arabic show a similar though not identical pattern.

9.8.6. Stimuli selection for Ar-2

In Ar-1, the other aim associated with research objective 1 was determining retention/omission of stimuli for Ar-2. To assist in this task, I was guided primarily by the responses of Syrian listeners. This decision was taken in order to keep the task of selection relatively simple, since taking into account variables in response patterns by both groups would have made the task very complex.

Two observations are relevant here. First, *all* listeners correctly rejected *all* Egyptian, Moroccan, Omani and Tunisian speakers to the highest degree of certainty. Second, in the face of Iraqi, Kuwaiti and (especially) Jordanian stimuli, Syrians were less certain and less accurate than were non-Syrian listeners, as well as less certain and less accurate than they were when confronted with Syrian stimuli. Some stimuli of these nationalities proved especially problematic for Syrians, and these were the non-Syrian stimuli retained for Ar-2. They were:

- the six Jordanian stimuli, to which in each case Syrian listeners had either entered a false acceptance or responded 'uncertain';

- Iraqi stimuli 29 & 42, the only two of this nationality which Syrian listeners (and only Syrian listeners) had accepted incorrectly;

- all Kuwaiti stimuli (except 30), to each of which one Syrian listener at least had incorrectly accepted or responded 'uncertain';

- six Syrian stimuli, selected at random, since responses to all Syrian stimuli by Syrian listeners were homogeneous in both their degree of certainty and their accuracy.

Owing to their comprehensive correct rejection by all listeners, it was decided to omit all Egyptian, Moroccan, Omani and Tunisian stimuli. The selection described above, with the addition of two stimuli from Lebanese speakers, reduced the total number of stimuli from 52 in Ar-1 to 21 in Ar-2. The Lebanese stimuli were recorded by Lebanese Arabic speakers on request, using the Syrian script of the *Juha* story, and included in Ar-2 precisely because of the close

relationship between Syrian and Lebanese Arabic. Transcripts of the Lebanese stimuli are also included in Appendix C.

It was expected that the use of these 21 stimuli would serve to induce still more clearly differentiated response patterns in Syrian and non-Syrian listeners than had occurred in Ar-1. Further detail on decisions made about stimuli selection for Ar-2 is provided in Chapter 10.

9.8.7. Listeners' demonstrated awareness of Syrian dialectology

It is notable that the Jordanian, Iraqi and Kuwaiti stimuli represent dialects from three distinct Arabic subfamilies. The Jordanians speak a Levantine variety, the Iraqis Mesopotamian and the Kuwaitis Gulf Arabic, all of which share features with or belong to dialects found on Syrian territory (e.g.Horani Arabic in southern Syria/northern Jordan).

The relative inability of Syrian listeners to reject these stimuli correctly—with a comparable degree of confidence/accuracy either to non-Syrians or to their own judgements of Syrian stimuli—suggests that the difficulty may stem from Syrians' knowledge of the variegated character of the country's dialectology. In other words, one interpretation of the response patterns reviewed in section 9.8 is that Syrian listeners—and to a lesser extent, non-Syrians— misidentify non-Syrian stimuli as Syrian not because of what they *do not* know but because of what they *do*. As shown below, this interpretation is validated by listener comments.

Comments in Ar-1 indicate strong awareness of cross-border varieties. Both non-Syrians and (especially) Syrians entered comments revealing knowledge of both the close relatedness of the Levantine dialects and the existence of non-Levantine varieties in Syria—particularly the (Iraqi) Mesopotamian and Horani dialects. Since these comments often co-occur with false acceptances of or 'uncertain' responses to Jordanian, Iraqi or Kuwaiti stimuli, this awareness would seem to have sometimes led them to error. Once again: the 'problem' here is what listeners *do* know rather than what they *do not* know. There were also comments in this vein, often similarly associated with false acceptances, by non-Syrian listeners. The comments reproduced verbatim below have two sources: first, in connection with judgements of

individual stimuli (in the free-text boxes associated with each stimulus); and, second, the free-text box at the conclusion of the survey.

Comments by Syrian listeners included the following, reproduced verbatim.

- *"This is more Iraqi to me but regions nearby speak similarly"* ('uncertain' response to an Iraqi stimulus)

- *"It looks more Jordanian or Palestinian to me but there is a chance that can be also spoken by nearby areas in Syria"* ('uncertain' response to a Jordanian stimulus)

- *"Accent can be related to Eastern Syrian"* (false acceptance of a Kuwaiti stimulus)

- *"Eastern Syria again"* (false acceptance of a Kuwaiti stimulus)

- *"Can be Eastern Syrian"* (false acceptance of a Jordanian stimulus)

- *"Probably Eastern Syrian"* (false acceptance of a Jordanian stimulus)

- *"It sounds more Iraqi to me but nearby regions speak similarly"* ('uncertain' response to a Kuwaiti stimulus)

- *"This can be Jordanian or Palestinian but also can be related to some areas in Syria"* (false acceptance of a Jordanian stimulus)

- *"It is more Jordanian but again regions nearby might share similar accent"* (false acceptance of a Jordanian stimulus)

- *"It can be Eastern Syrian"* (false acceptance of an Iraqi stimulus)

- *"It is close to Iraqi but also Eastern Syrian speak similarly"* (false acceptance of an Iraqi stimulus)

- *"It is more inclined to Jordan and Palestine but also some Syrian might speak similar"* (false acceptance of a Jordanian stimulus)

- *"Again Eastern Syria"* (false acceptance of an Iraqi stimulus)

- *" In eastern Syria such as Deir ez-Zor; Al Hasakah, Raqah, etc. people speak in similar accent"* (false acceptance of an Iraqi stimulus)

- *"Probably Horani"* (false acceptance of a Jordanian stimulus)

- *"This is quite likely Horani"* (false acceptance of a Jordanian stimulus)

- *"Horani"* (false acceptance of a Jordanian stimulus)

- *"Horani"* (false acceptance of a Jordanian stimulus)

- *"Likely Hirani but could easily be Jordanian as well"* ('uncertain' response to a Jordanian stimulus)

- *"This is tricky, but I would go with Horani Arabic (dialect from southern Syria on the border with Jordan). It is quite easy to confuse with a Jordanian or Palestinian dialect since the share a lot of the same features, but I am inclined to judge it as Syrian because of the conditional raising in the word 'city'."* (false acceptance of a Jordanian stimulus)

Comments of this type by non-Syrians were less abundant, and all were entered by a single (Libyan) listener.

- *"I wouldn't be surprised if it's a variety spoken in the Syrian/Iraqi border"* (correct rejection of an Iraqi stimulus)

- *"I'm not sure if it's Syrian or Jordanian. It's Bedouin. The first part (second 1) had the masculine possessive pronoun which sounded Levantine, the word 'hathoul' (second 4) is also Levantine but that changed in second 8 when she said 'garyeh'. It sounds Bedouin and I have limited knowledge of the Bedouin variety of Syrian"* ('uncertain' response to a Jordanian stimulus)

- *"It sounds Syrian based on lexical choice but the phonology could be a Bedouin or rural*

variety; the voiced affricate and the velar plosive" (false acceptance of a Jordanian stimulus)

- *"Sounds like a Bedouin/rural (away from the city) variety of Syrian"* (false acceptance of a Jordanian stimulus)

The willingness of the native-speaker linguists in Ar-1 to accept such stimuli as Syrian, or to enter 'uncertain' responses to them, suggests that linguists are, as Fraser (2009, 2011, 2019) asserts, cautious in their conclusions. However, the validity of Fraser's argument that *only* linguists exercise suitable caution in their judgements is revisited in Chapter 12.

9.9. Research objective 2: Results

9.9.1. Organisation

Research objective 2 emerges from hypothesis (iii) of the thesis: speaker-listeners will vary by region of origin in the number and kinds of cues they recognise. The means chosen to evaluate research objective 2 was a simple comparison of the types of cues cited by Syrian and non-Syrian listeners, respectively. The first type of cue was modelled as those that are identifiable with features described in the dialectological literature; the second was modelled as those that are not. In adopting this approach the ultimate intention was to gain insights that might inform the development of a supplementary production test for LAAP.

In the following review of listener comments, phonetics/phonology is covered initially, followed in turn by morphosyntax, suprasegmentals and lexis. In subsections 9.9.2 to 9.9.5, individual features of all four classes are described, as closely as is practicable, with reference to the dialectological sources (these features, and their location in the *Juha* text, are specified in subsection 9.7.3, Tables 9.1 and 9.2). An informal prediction is then made of the diagnostic value of each feature. Following this, citations of each feature are examined for the extent to which each seems to have guided listeners' judgements.

Subsection 9.9.6 covers citations of features *not* mentioned in the consulted dialectological

literature. Subsection 9.9.7 is a summary of the feature citations of Syrian and non-Syrian listeners respectively. Subsection 9.9.8 is an interpretative discussion of observed divergences between features cited by Syrian/non-Syrian listeners; also covered in 9.9.8 is the strength of support for hypothesis (iii) that can be inferred from these results.

In describing and classifying listener comments, a distinction is made between diagnostic and salient features. Diagnostic features are those cited in connection with correct responses only. Salient features are those cited in connection with either correct or incorrect/uncertain responses. A feature can be classified as both salient and diagnostic, provided it is cited in association with a correct response but as diagnostic only of mentioned in connection with a correct response. If, however, it is cited in association with any other type of response, it is classified as salient only.

This framework permits an informal weighing of the degree to which listener impressions agree with those of the reviewed literature on Arabic dialectology (salience) and the extent to which these impressions are brought to bear in correctly identifying accents (diagnostic value).

9.9.2. Phonetics/Phonology

MSA /q/

According to Lentin (2012), in Damascene Arabic [ʔ] is the most common reflex of MSA /q/ (see line 1, /ʔalu:lu/), though [q] also occurs (see line 5, /qarji/). Compare this to the reflex [g] in the Iraqi, Jordanian, Kuwaiti, Omani and Moroccan transliterations, and [q] in the Tunisian.

In Iraqi Arabic [q] occurs as a variant (here, in line 7, /qarji/). The variant [ʔ] is likewise present in Kuwaiti Arabic in its reflex of the same word (see line 7, /ʔarja:f/).

As well as in the Syrian version, [ʔ] occurs in the Egyptian (Wilmsen 2012). However, it co-occurs in line 1 with a differing word-order (i.e. Egyptian [ʔasħa:bu ʔalu:lu] vs. Syrian [ʔalu:lu sħa:bu]). Naïm (2012) states that, in Lebanese Arabic,[ʔ] is the commonest reflex of MSA /q/ but that [q] occurs in loanwords from MSA. In line 1, [ʔ] is indeed used by both Lebanese speakers. But, in line 7, [q] (in [qarje]) is used by one of the speakers, and [ʔ] (i.e. [ʔarje]) by the

other.

The reasons for this inter-speaker variation are difficult to discern; suffice it to say that the distribution of variants in Lebanese Arabic may be more complex than some of the sources admit. Nonetheless, given that Lebanese stimuli were not used in this experiment, the expectation was the various reflexes of MSA /q/ would have considerable diagnostic value for Damascene Arabic.

Two of the 10 total listeners commented on the occurrence of the various reflexes of MSA /q/ in line 1.

In correctly accepting all six Syrian stimuli, an Egyptian listener repeatedly cited "realisation of /q/ as a glottal stop", presumably in line 1. This seems to lend additional support to the prediction that Damascene [ʔ] in line 1 may function as strongly diagnostic of Syrian Arabic.

A Saudi listener, in expressing uncertainty about the provenance of a Jordanian speaker, commented: "Syrians are known to use أ instead of ق . This speaker used g" (note that in the Arabic alphabet أ represents /ʔ/, and ق represents /q/).

Two listeners, with three comments between them, remarked on the reflexes of /q/ in line 7. A Saudi listener (not the same person as mentioned above) correctly rejected a Jordanian stimulus with reference to the occurrence of [g]: "the speaker might be Jordanian (algaryah)".

A Libyan listener, meanwhile, identified a Jordanian stimulus as possibly Bedouin on the basis of "the velar plosive", though in the event incorrectly accepted the stimulus in question. The same respondent was uncertain whether the speaker in a second Jordanian stimulus was Syrian or Jordanian but cited [g]in "garyeh" as a Bedouin feature.

Thus the variability in Damascene Arabic of reflexes of MSA /q/, as exemplified by [ʔ] in line 7, appears to have been of less help to listeners than that encountered in line 1 (as [q]).

The occurrence of various reflexes of MSA /q/ was the feature most widely commented on by respondents in Ar-1. The feature thus evidenced higher salience than all others; its salience, however, was not matched by its diagnostic value among listeners in general (i.e. apart from the solitary Egyptian, all of whose citations of MSA /q/ as [ʔ] were associated with correct acceptances). Notably, all respondents who commented on the variants of MSA /q/ were non-

Syrian.

MSA/ʤ/

The MSA voiced palato-alveolar fricative, /ʤ/, becomes [ʒ] in Damascene Arabic (see lines 1 and 7). Damascene Arabic shares this realisation with Moroccan and Tunisian. In Iraqi, Jordanian, Kuwaiti and Omani it is, broadly, [ʤ]. In Cairene Arabic it is [g].

On its own this feature would appear to have considerable diagnostic value in distinguishing a Cairene speaker from a non-Cairene. It was not, however, predicted to have great diagnostic value as an identifier in its own right of Syrian (Damascene) Arabic, since its realisation as [ʒ] is shared by the Moroccan and Tunisian dialects.

Despite reports in the literature, and the evidence of the stimuli, that [dʒ] is rare or non-existent in Damascene Arabic, an Egyptian listener thrice cited the occurrence of [ʤ] in the stimuli as diagnostic of Syrian Arabic: e.g. 'realisation of /dʒ/ as [dʒ]'. Interestingly, the respondent correctly accepted all the corresponding stimuli as Syrian.

A Libyan listener, meanwhile, identified a Jordanian stimulus as possibly "Bedouin" on the basis of "the voiced affricate", though opting in the event to incorrectly reject the stimulus in question. In this case, it seems likely that the listener thought that the stimulus featured a speaker from a border region (cf. subsection 9.8.1.7).

Accuracy was, as with MSA /q/, mixed in cases where listeners commented on this feature, suggesting for the time being that it is more salient than diagnostic in character. Again, neither Syrian respondent commented on it.

MSA/ð/

In Damascene Arabic stimuli the reflex of the MSA voiced interdental fricative, /ð/, is [d] (see line 3, [ha:dol]). Lebanese, Moroccan and Egyptian share this reflex with Damascene Arabic. In Jordanian, Kuwaiti, Omani and Tunisian, /ð/ occurs. Among the eight featured dialects, even in those which share the [d] reflex, there is considerable variability in the form of the relevant lexeme (compare e.g. Egyptian [du:l], Tunisian [haðu:], Iraqi [ðo:la], Jordanian/Kuwaiti/Omani

[haðo:la], Moroccan [hade:]).

Indeed, according to the sources, [ha:dol] apparently bears a good deal more salience and diagnostic value than merely as an exponent of the /ð/ - /d/ opposition. Brustad (2000, p. 114) reports that in Egyptian Arabic /ha-/ is omitted from the pronoun (which in this text means *they*); this is in fact the case in the featured Egyptian stimuli.

Also of interest is a discussion of the sociolinguistic indexicality of (ha)dol by Owens (2001). Here it is stated that in one study Egyptians showed no reluctance to maintain their use of the dialectal form /du:l/ in a formal context in which Syrians were observed to switch from their colloquial form to the MSA equivalent. Owens interprets this to mean that "Egyptians may be less reticent about maintaining their local norms in formal situations than are certain other nationalities" (2001, p. 433). Apparently, then, the regional variation in tokens of *they* is known to (some) Arabic speakers, with Egyptian /du:l/ being a marker of Egyptian cultural self-confidence, and Syrian shifting to the MSA form a sign of the relative lack of it among Syrians.

The inference is, then, that speakers of at least Syrian and Egyptian Arabic are overtly conscious of dialectical variability in this lexeme. In this connection it is worth recapitulating Behnstedt's (2012) mention of the distinctiveness of Syrian personal pronouns, of which /ha:dol/ would presumably be considered an example.

With such considerations in mind, and in the absence from this experiment of Lebanese stimuli, /ha:dol/ and its variants, as a whole lexeme rather than simply as an exponent of /d/, was predicted to function as highly salient and diagnostic.

A Saudi listener cited the word for *they* (/ha:dol/ in Damascene Arabic) in two instances. In each case this was interpreted as a reference to the /ð/ - /d/ opposition, owing to the varying informal transcriptions he/she employed ("hathool" vs. "hadool"). This listener apparently remarked upon its realisation as [ð] in correctly rejecting one (Jordanian) stimulus and its realisation as [d] in correctly accepting a second.

A Libyan respondent mentioned its realisation in a Jordanian stimulus "in the word hathoul"

(again interpreted as a reference to /ð/), but classified the stimulus in question as of uncertain origin. It seems likely that the listener was unsure which side of the Syrian-Jordanian border the speaker was from.

MSA /ð/ thus proved to have greater salience than it did diagnostic value, though the latter was not negligible. It has been interpreted here as arising from the /ð/ - /d/ opposition, because in all cases its citation occurred in discussing the provenance of Levantine (i.e. Syrian or Jordanian) stimuli, in which the opposition between these two sounds alone—as opposed to, e.g., reduced forms such as Egyptian [du:l]—is salient. As was the case with all previously reviewed features, this one received no acknowledgement from Syrian listeners.

Final imāla

Final imāla, typical of the Levantine dialects, is a process whereby word-final /a/ (or /a:/) is raised in the environment of /i/ (Versteegh 2001, p. 153). The resulting realisation varies, in the Levantine dialects, between [e] and [i] (Owens 2006). It is sometimes characterised as a sort of vowel harmony (Owens 2006).

Hellmuth (2019, p. 65) laments that "... the facts of many aspects of the phonology of many varieties of Arabic are still unknown," and numbers vowel fronting/raising (of which imāla is an exemplar) among the underexplored aspects of Arabic phonology. Final imāla is thus a particular manifestation of a general phenomenon, especially marked in Syrian Arabic, which is as yet poorly described, notwithstanding Behnstedt's (2012) very general citation of "intonation combined with certain pausal features" and imāla.

In these stimuli, final imāla occurs only in Syrian and Jordanian Arabic, i.e. in examples of the Levantine dialects, as a process operating on the final vowel, here marking a feminine noun, in /madina/, *city* (see line 4), and /qarja/, *countryside* (see line 7). In Damascene Arabic it is realised, respectively, as /madi:ni/ and /qarji/; in Jordanian it is /madi:ne/ and /garje/. It was predicted that final imāla in -/i/ would carry substantial diagnostic value for Syrian Arabic, in contradistinction to the only other Levantine dialect present (i.e. Jordanian) as well as to the other dialects, in which it does not occur.

Notably, final imāla was the only feature in the literature cited by a Syrian respondent. This occurred in the context of a false acceptance of a Jordanian stimulus, on the explicitly stated grounds that it was likely from the Syrian-Jordanian border region known as the Horan. No details were given, beyond the description "conditional raising in the word city", as to how the particular phonetic properties of the imāla featured in this stimulus led the respondent to this conclusion.

A Saudi respondent was uncertain as to the origin of a Jordanian stimulus but mentioned the occurrence of final imāla, citing the "long vowel in madineih". Apparently the respondent heard imāla here as suggestive of some form of Levantine Arabic but could not conclusively identify the stimulus in question as Syrian. The predicted strongly diagnostic value of final imāla was therefore not validated; neither did it turn out to be very salient for listeners.

9.9.3. Morphosyntax

Adjective negation

Adjectives are negated in Damascene Arabic by /mu:/ (see line 3), a form which is shared with Iraqi and Kuwaiti. All other featured varieties use other negation particles (e.g. Jordanian /miʃ/, Tunisian /ma:/).

In this text, Damascene Arabic combines the negation particle with the adjective /mna:ħ/ (i.e. *good*). Lebanese uses the near-isomorphic form /mni:ħ/ in conjunction with /ma/ (yielding the phrase /ma mni:ħ/; compare to Damascene /mu: mna:ħ/). Other dialects employ more or less differing lexical and syntactic tools to express the same concept (e.g. Moroccan /welad lħram/).

The various phrases equivalent to *very bad*/*not good* therefore contain phonetic, lexical and syntactic cues for dialect identification. As Lebanese Arabic did not feature among the stimuli in Ar-1, this phrase was expected to be recognised as strongly diagnostic of Damascene Arabic.

Two listeners, an Egyptian and a Libyan, cited the negation phrase as typical of Syrian Arabic— the Egyptian respondent in connection with each of the six Syrian stimuli—and in every case they correctly accepted the relevant stimulus. It was therefore among the most salient and the most diagnostic of all the features in the stimuli. However, neither of the two Syrian listeners

mentioned it.

The preposition *in*

The Damascene Arabic form of the preposition semantically equivalent to *in* is /bi(l)/- (see line 2, /bilmadi:ni/). The alternation of forms, /∅/ vs. -/l/ is conditioned by the phonotactics of Arabic; further variability in form is syntactically conditioned (Shetewi, pers. comm.).

According to Lentin (2012) in most Levantine dialects /fi/- co-occurs with pronouns and /bi/ with nouns. Procházka (2012) states that only /bi/- occurs in Iraqi Arabic, while in Marghrebi varieties /fi/ is used for the locative (and the existential) and /bi/- for the instrumental (see also Aguade 2018).

In these stimuli, /bi(l)/- is indeed shared among Syrian, Iraqi and Lebanese, while all other dialects use either /fil/- or /l/-. The former, thus encountered beyond the Levantine group, was not predicted to act as crucial diagnostic and in fact was not mentioned by any listener.

Future marker

In Damascene Arabic /raħ/ (see line 6) is employed as a future marker (alternating, according to Lentin 2012, with /laħ(a)/). In these stimuli the form /raħ/ is shared with Iraqi, Lebanese, Jordanian and Kuwaiti.

Various alternatives occur in other dialects (c.f. Moroccan /kaj/, Tunisian /tawa/, Omani /baj/-, Egyptian /haj/-). Aguade (2018) explains that /ka/ is used in Moroccan Arabic as a durative marker. Eid (2012) states that in Moroccan Arabic /ka/ denotes the imperfective aspect, associated with habitual or progressive actions.

In these stimuli, the near variant /kaj/ seems to be a variant of /ka/, which in turn appears to function in syntactically and semantically in the same manner as /raħ/. In any case /raħ/, which, owing to its fairly wide distribution even beyond the Levantine group seemed unlikely to be perceived as especially diagnostic of Syrian Arabic as such, was not mentioned by any of the respondents.

3MSg marker

In Damascene Arabic the (pro)nominal suffix -/u/ (with -/o/ as a variant) marks the 3MSg in

both the objective and the genitive case (line 3). The use of -/u/ applies in all other featured dialects except Iraqi (Abu-Haidar, 2012), Kuwaiti and Omani (Holes 2012a, 2012b), in which it is -/a/.

Again, though, no listener cited this feature, which, because of its shared distribution among the Levantine, Egyptian and Maghrebi dialects, was not predicted to function as highly diagnostic of Syrian Arabic.

3PLSubj marker

The third-person plural subject marker in Damascene Arabic is the suffix -/u/ (Lentin 2012), a form which it shares with all dialects featured except the Iraqi, Kuwaiti and Omani, in all of which it is -/un/ (Holes 2012a, 2012b; Jastrow 2012b). Its wide distribution suggested that it would not function as a crucial diagnostic, and indeed it was not cited by any listener.

9.9.4. Suprasegmentals
Stress

As Hellmuth (2019) points out, information on the stress patterns of the various Arabic dialects is relatively scarce—an observation which she extends to comparative (i.e. cross-dialectal) studies of Arabic suprasegmentals in general (Hellmuth 2013, p. 66). As she also remarks,

> "...minimal stress pairs that crucially distinguish one dialect from another are relatively infrequent, occurring only in words of certain prosodic shapes"

(2013, p. 62).

Lentin (2012) states that in Damascene Arabic words with three syllables or more and containing a long initial syllable (i.e. (C)VC or (C)V:) take stress on the initial syllable. Woidich (2012) specifies that, in Cairene Arabic, words with the same syllabic structure receive stress on the post-initial syllable. This stress pattern would appear to contrast neatly two of the most widely spoken Syrian and Egyptian Arabic varieties.

Comments indicate that an Egyptian respondent may have relied on this feature in accurately identifying all of the six Syrian stimuli: "stressing the first (antepenultimate) syllable in

[lahdʒitak]". A Saudi listener, too, offered "lahjtaak (the tone)" as a cue which had guided (correct) identification of a Syrian stimulus. Whether the latter amounts to a citation of initial-syllable stress in Damascene Arabic or some quality of intonation (discussed below) is difficult to decide; certainly comments such as this demonstrate the difficulty of interpreting non-technical remarks with reference to the literature on Arabic dialectology. Again, neither of the Syrian listeners commented upon this feature.

Hellmuth (2019, p. 174) reports that stress in Gulf Arabic (including Kuwaiti) is assigned to a "superheavy" syllable, where it occurs, and otherwise to the penultimate syllable. This feature, however, was not mentioned by any listener.

Vocalic mass/speech rhythm

On the matter of speech rhythm, Hellmuth (2019, p. 175) summarises findings that V% measurements—the quantity of vocalic material present in equivalent utterances from dialect to dialect—combined with comparative measurements (ΔC) of the duration of non-vocalic intervals broadly increases from west (i.e. Morocco, Tunisia) to east (i.e. Syria, Lebanon) in the Arabic-speaking world.

Hellmuth (2019) also notes that, while they can be used to measure speech rhythm, V% metrics vary as a function of speech rate. She also reports the results of an experiment which showed that native speakers of Arabic were able accurately to identify speakers of western or eastern Arabic varieties solely by the rhythm in which samples were spoken (i.e. on the basis of vocalic mass alone, with semantic material not present).

This general pattern of east to west variation in speech rhythm was not reflected in listener comments on any of the stimuli.

Syllable Structure

Versteegh (2001) explains that the Syro-Lebanese and Mesopotamian dialects differ from Egyptian Arabic in their treatment of epenthesis, particularly -CCC- clusters. In Damascene and Mesopotamian Arabic an epenthetic vowel intervenes between the first and second consonant;

in Egyptian Arabic it occurs before the third. While this feature may have been expected to help distinguish Syrian stimuli from Egyptian, no listener cited it.

Versteegh (2001) also specifies a special type of syllable structure for the Maghrebi dialects (here Moroccan and Tunisian), in opposition to the pattern prevailing elsewhere: CvCC>CCvC.Again, this feature received no acknowledgement any listener.

Intonation

Rising intonation in utterance-final contexts is reported to be a distinctive feature of Syrian (particularly Damascene) Arabic; this "drawl" is common in both interrogative and declarative structures (Hellmuth 2020, p. 594). Whether such an intonation pattern would occur in read passages of the type featured in the stimuli is questionable. In any case, no listener explicitly cited intonation in their comments.

Nevertheless, recalling once again Behnstedt's (2012) mention of (unspecified) intonation patterns as typical of Syrian Arabic—as well as the possible interrelationship between the parameters of intonation, vowel length/height and speech rhythm—the question of intonation remained as a consideration for later experiments in this series.

9.9.5. Lexis

Some listeners cited lexis as diagnostic of Syrian Arabic. The Egyptian listener, for example, repeatedly cited the construction /di:r ba:lak/ as distinctively Syrian in correctly identifying three Syrian stimuli, though since it occurred also in the Iraqi, Jordanian and Kuwaiti stimuli it seems likely that these references to its distinctiveness were founded on something other than its purely lexical properties.

A Syrian listener alluded to "lexical choice" as partially informing the correct rejection of an Iraqi stimulus. The same listener used an identical phrase in explaining the false acceptance of a Jordanian stimulus. As reviewed above, Syrian /mna:ħ/ (*good*) was cited by two (non-Syrian) respondents, and in multiple instances by one of them, though always in association with the

negation particle /mu:/.

As also detailed above, the word for *they* (/ha:dol/ in Damascene Arabic) was cited by a Libyan listener as "Levantine", independently of the negation phrase and in connection with a Jordanian stimulus, where it was transliterated informally as "hathoul". It was also cited twice by a Saudi listener ("hathool" vs. "hadool"), both times in relation to stimuli of Levantine origin. For reasons previously discussed, references to the word for *they* were construed as commentaries on phonological rather than lexical variability.

9.9.6. Citations of features not mentioned in the literature

The Egyptian listener cited "voice quality" in comments correctly identifying all the Syrian stimuli. The Libyan listener mentioned "vowel quality" in connection with an Iraqi stimulus, correctly rejected partly on this basis. To what precisely the qualities mentioned in each (or any) instance refer is impossible to say for certain with reference to the available literature, although they may have to do with intonation, vowel raising/lengthening or a combination of all.

9.9.7. Specification of features cited, by listener group

Tables 9.3 and 9.4 specify, for Syrians (N = 2) and non-Syrians (N = 10) respectively, the number of citations of features occurring in the relevant dialectological literature. These are divided according to whether they are associated with any response type ('salient') and/or correct responses ('diagnostic').

Table 9.3: Features cited by Syrian listeners, by salience and diagnostic value

Feature	/q/	/dʒ/	/ð/	Final imāla	Adjective negation	Preposition *in*	Future marker	3MSg	3PLSubj	Suprasegmentals (all types)	Totals
Salient	0	0	0	1	0	0	0	0	0	0	1
Diagnostic	0	0	0	0	0	0	0	0	0	0	0
Totals	0	0	0	1	0	0	0	0	0	0	1

Table 9.4: Features cited by non-Syrian listeners, by salience and diagnostic value

Feature	/q/	/dʒ/	/ð/	Final imāla	Adjective negation	Preposition *in*	Future marker	3MSg	3PLSubj	Suprasegmentals (all types)	Totals
Salient	10	4	3	1	7	0	0	0	0	6	31
Diagnostic	7	3	2	0	7	0	0	0	0	6	25
Totals	17	7	5	1	14	0	0	0	0	12	56

9.9.8. Discussion

Table 9.3 demonstrates that features referenced in the sources had no diagnostic value for Syrians—for whom, moreover, only final imāla exhibited (minimal) salience. Table 9.4 shows that Non-Syrians located considerable salience and diagnostic value in the same range of features. As reviewed in subsection 9.9.5, there were also some comments by non-Syrian listeners on lexis, and a single comment in this connection by a Syrian listener.

Reflexes of MSA /q/ evidence the greatest degree of salience *and* diagnostic value for non-Syrians. Both MSA /dʒ/ and /ð/ exhibit somewhat lesser salience and diagnostic value, although the salience and diagnostic value of /dʒ/ was demonstrated only in the responses of the single

Egyptian listener. For non-Syrians taken as a whole, adjective negation appears both highly salient and diagnostic, though in six of the seven instances this was the case only for the sole Egyptian listener. Suprasegmentals were also highly salient and diagnostic—but again only for the Egyptian listener. Apart from two references by non-Syrians to voice/vowel quality, features *not* mentioned in the literature were completely absent from comments. This may be taken as suggesting that the reviewed dialectological surveys match fairly well the perceptions of (non-Syrian) native linguists.

There is therefore some evidence for the cues on which non-Syrians depended in making their decisions, with the various reflexes of /q/ and /ð/ being most prominent and widespread. However, the almost total absence of comments by Syrians means that the cues upon which they relied are largely unknown.

Hypothesis (iii) of the thesis is that speaker-listeners will vary by place of origin in the cues they recognise (in quantity, type or both). With appropriate caution owing to the relatively small number of respondents, the above results appear to support this hypothesis. Further exploration of the validity of hypothesis (iii) is undertaken in the two subsequent experiments on Arabic. I proceed now to discuss the design and results of the second experiment in the Arabic series, which predominantly involved NSNL listeners, Syrian and non-Syrian.

Chapter 10: Arabic experiment 2

10.1. Introduction

The background to the series of Arabic-language experiments to which this experiment (Ar-2) belongs is discussed in Chapter 9, section 9.1. Once more, listeners were presented with short voice samples and asked whether and to what degree of certainty (on a five-point Likert scale) they were able to associate the samples with Syria. The question asked was again, 'Is this a Syrian accent?' Ar-2 differs from its immediate predecessor in two respects: first, in the number and provenience of the stimuli included; second, in that the majority of listeners here are NSNLs, whereas all listeners in Ar-1 were native-speaker linguists.

10.2. Ar-2: hypotheses

As in Ar-1, only the first and third of the thesis' three hypothesis are examined in this chapter. They are:

(i) Syrian listeners will be more accurate than non-Syrians in identifying Syrian stimuli;

(iii) Syrians and non-Syrians will differ in the features they cite as having informed their decisions as to the national identity of the speakers in the stimuli.

The extent of support for hypothesis (ii)—which involves examining differences between linguists and NSNLs—is not susceptible to examination in this chapter because listeners were (chiefly) NSNLs. For an explanation of the equivalent limitations in Ar-1, consult section 9.2 of Chapter 9.

10.3. Ar-2: research objectives

Ar-2 has two research objectives, both of which are identical to those of Ar-1 (Chapter 9, section 9.3). The limitation imposed by sample size on a full investigation of research objective 1 applies here as it did in Ar-1.

Research objective 1 is approached here partly as a means of determining retention/omission of stimuli for the succeeding experiment (Ar-3), and partly with reference to the validity of hypothesis (i). The criteria used in decisions made on the former are the same as those set out in Chapter 9, section 9.3.

In its investigation of research objective 2, Ar-2 differs from Ar-1 in the sense that the majority of listeners here are NSNLs, whereas all listeners in Ar-1 were native-speaker linguists. The same analytical procedure nonetheless applies here as it did in Ar-1: cataloguing and quantifying listener citations of the cues that informed their identification of stimuli as Syrian or non-Syrian.

10.4. Methodology

10.4.1. Task design

Task design differed from that of Ar-1 (Chapter 9, subsection 9.4.1) in the number and national origins of stimuli used (see below, subsection 10.4.3). Otherwise it was identical to that of Ar-1, except that all material in the Qualtrics user interface was translated into Arabic. Examples of the two primary questions are displayed in Figures 10.1 and 10.2. Figure 10.3 shows the free-text box and accompanying question inviting general comments on the tasks/stimuli. The free-text box appeared at the end of the survey.

Figure 10.1: Main task interface; script at top-right reads, Is this a Syrian accent? Script at bottom reads, 'To the right, please indicate your degree of certainty', with five alternatives (Highly likely, Likely, Uncertain, Unlikely, Highly unlikely) specified above each check box.
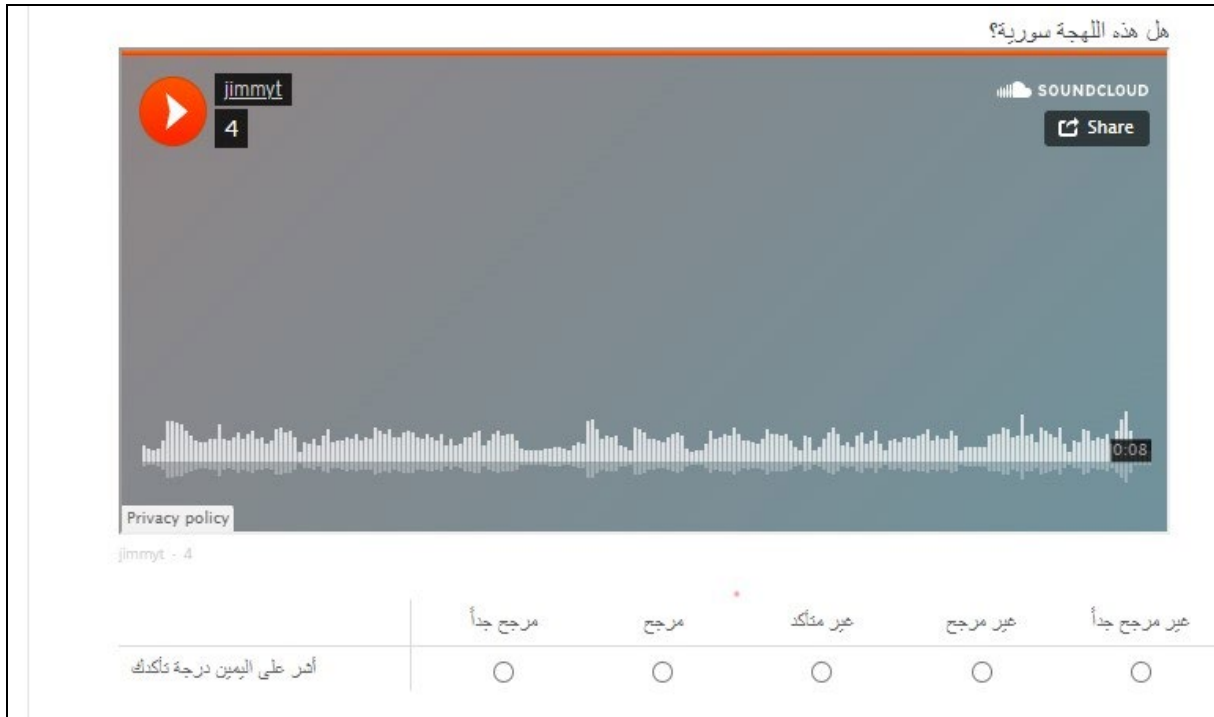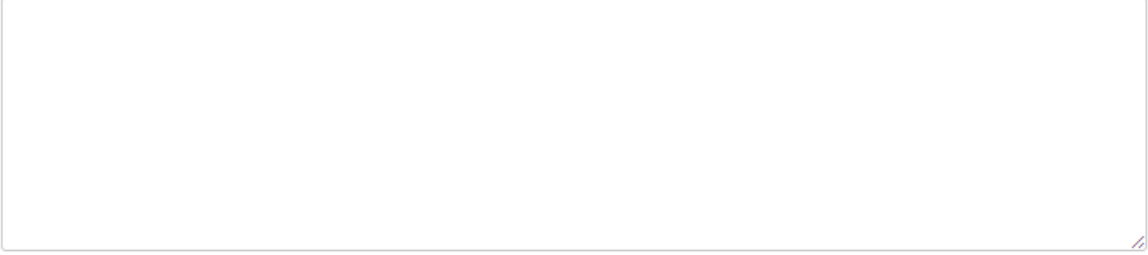
Figure 10.2: Free-text box displayed below main task; script at top-right reads, 'What clues lead you to your judgement of the speaker's accent?'

Figure 10.3: Free-text box for general comments on the task/stimuli; text reads, 'In the box below, please enter any additional comments you wish to make about the accents, the tasks, your responses or anything else you think relevant.'

في الصندوق في الأسفل، الرجاء إدخال أية تعليقات إضافية ترغب في إعطائها حول اللهجات، الإستبيان، إجاباتك، أو أي شي آخر تظن أنه ذو صلة.

10.4.2. Listeners

As in Ar-1, listeners were recruited through contacts at the universities of York and Newcastle. Questions asked to ascertain listeners' linguistic background (i.e. whether a native speaker of Arabic), national origin/citizenship and any periods of residence outside their country of birth, as well as whether they had received postgraduate linguistic training, were substantively the same as in Ar-1.

Here, though, all questions were translated into Arabic. This was done to allow participation by NSNL listeners who do not speak/read English. Figures 10.4, 10.5 and 10.6 display these questions as they appeared in Qualtrics.

Figure 10.4: Question asked to ascertain listeners' language background; script at top left reads, 'Is Arabic your mother tongue?'

هل العربية لغتك الأم؟

○ نعم

○ لا

Figure 10.5: Question asked to ascertain listeners' national origin/citizenship and period/approximate dates of residence outside country of birth/citizenship; script reads, 'In which country were you born a citizen? Have you lived in any country outside your country of citizenship? If you answered yes, please give approximate locations and dates below.'

في أي دولة ولدت كمواطن؟ هل عشت في أي دولة خارج البلد الذي تحمل جنسيته؟ إذا أجبت بنعم، الرجاء إعطاء المواقع و التواريخ بشكل تقريبي في الأسفل.
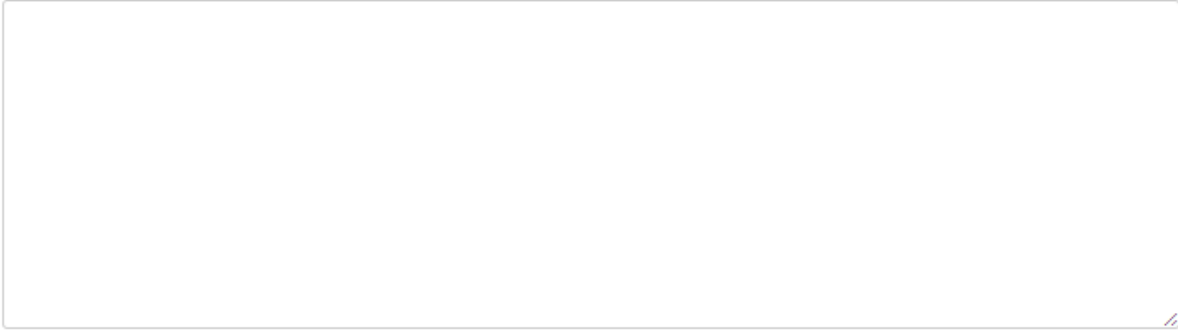
Figure 10.6: Question asked to ascertain listeners' educational background in linguistics; script at top reads, 'Have you studied linguistics at a postgraduate level?'; script beside check-boxes reads, 'Yes' and 'No'.

هل درست اللغويات بمستوى دراسات عليا؟

○ نعم

○ لا

Unlike in Ar-1, in which responses by Arabic native speaker linguists were sought, here the aim was to recruit only Arabic NSNLs. This decision was made so that participants would better reflect the educational profile of most asylum claimants. In this way Ar-2 would act as a pilot study for its successor (Ar-3), as Ar-1 had for Ar-2.

Twelve listeners from various Arabic-speaking countries responded. Three listeners were of Syrian background. Among the nine non-Syrians there were six Saudis, one Jordanian, one Iraqi and one Moroccan.

Despite the intention to recruit only NSNLs, one each of the Syrians and non-Syrians declared postgraduate linguistic training. Their responses were not excluded, because it was as yet

unknown whether Arabic-speaking linguists and non-linguists would differ in their response patterns—a question that is explored in Ar-3 (Chapter 11).

Table 10.1: National origin/educational background and number of listeners

Origin/ Education	Syria	Jordan	Morocco	Saudi Arabia	Iraq	Totals
PGling	1	0	1	0	0	2
NSNL	2	1	0	6	1	10
Totals	3	1	1	6	1	12

10.4.3. Stimuli

As in all previous experiments stimuli were uploaded to Soundcloud, to which they were linked in Qualtrics. Where stimuli are concerned the same range of considerations applies—i.e. in connection with selection, Arabic dialectology and predicted diagnostic features—as documented in Section 9.7 of chapter 9. The same extract of the *Juha* story was employed as in Ar-1; consult Appendix C for transcripts. As in Ar-1, stimuli were presented in pseudo-random order.

Whereas Ar-1 included 52 stimuli from eight countries of origin, in Ar-2 there were 21 stimuli from five countries (Table 10.2). Of the latter, two stimuli were of Lebanese speakers, and the remainder were from Iraq (two), Jordan (six), Kuwait (five) and Syria (six).

The Lebanese stimuli, each featuring an originally Beiruti female speaker reading the same extract of the *Juha* story as in the IVaR extracts, were not featured in Ar-1. The similarity between Damascene and urban varieties of Lebanese Arabic motivated their inclusion here, with the prediction that they would actuate further differences in the respective response patterns of the two groups, Syrian and non-Syrian. With the close relatedness of the two varieties in mind, the written script recited by the Lebanese speakers was identical to that used by the Syrian speakers in the IVaR extracts; see the transcripts in Appendix C.

Table 10.2: National origin and number of stimuli

Origin	Syria	Lebanon	Jordan	Kuwait	Iraq	Total
Number	6	2	6	5	2	21

10.5. Research objective 1: results

10.5.1. Organisation

This section details results emerging from research objective 1: How and to what extent did the response patterns of Syrians and non-Syrians differ in answer to the question, 'Is this a Syrian accent?' This objective relates directly to determining the extent of support for hypothesis (i).

For the same reasons as described in Chapter 9, subsection 9.8.1, 'decision' responses are considered in their own right according to Likert-scale ratings ('highly likely', 'highly unlikely' and so on); they are not consolidated into 'correct' and 'incorrect' categories. Again, too, 'non-decision' responses (i.e. 'uncertain' ratings on the Likert scale) are considered independently. Also as in Ar-1, due to the small number of listeners and the fact that this experiment is a pilot for its successor, results are evaluated solely in the form of descriptive statistics.

Subsection 10.5.2 deals with the two groups' combined response patterns to Syrian and non-Syrian stimuli, respectively. Subsections 10.5.3 to 10.5.4 detail group response patterns by *individual* stimuli—first Syrian, then non-Syrian. Subsection 10.5.5 is a discussion of results with reference to the strength of support for hypothesis (i). Decisions as to the retention or omission of stimuli for Ar-2 are detailed in subsection 10.5.6. In subsection 10.5.7, I examine the raw data, combined with listener comments, to examine Syrians' and non-Syrians' respective knowledge of Syrian dialectology.

10.5.2. Syrian vs. non-Syrian stimuli

Figure 10.7 contrasts group response patterns to all Syrian and non-Syrian stimuli, respectively.

Figure 10.7: combined group response patterns to Syrian and non-Syrian stimuli
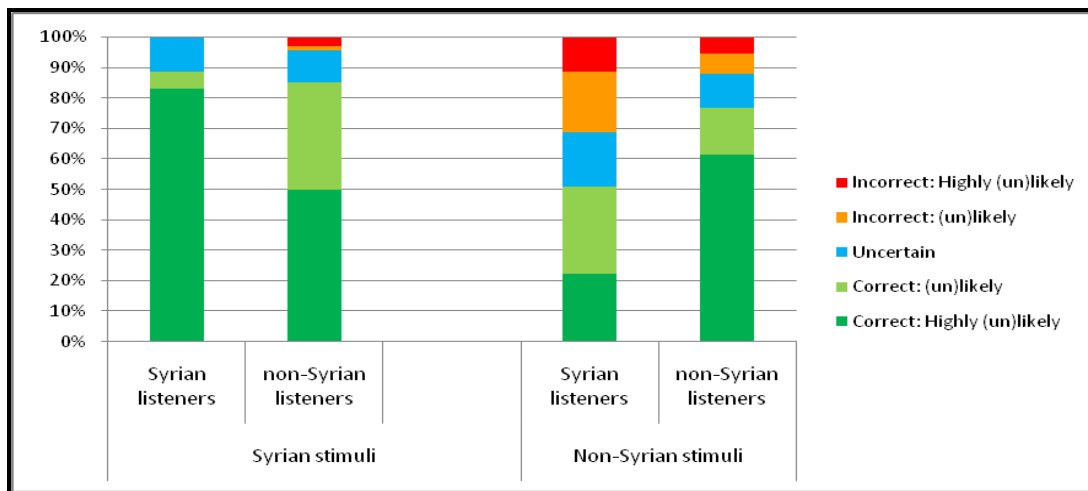
On Syrian stimuli, Syrian listeners entered c. 90% correct acceptances: 85% at the highest degree of certainty ('Highly likely'). The balance of their responses was of the 'uncertain' type; there were no false rejections.

With c. 85% correct acceptances, non-Syrians were almost as accurate as Syrians on Syrian stimuli, though they were less confident in their correct acceptances (c. 50% 'Highly likely'). They entered c.10% 'uncertain' responses, with the remainder consisting of false rejections at one or the other degree of confidence. In fact, non-Syrians were markedly more accurate on Syrian than on non-Syrian stimuli.

On non-Syrian stimuli, non-Syrians were considerably more accurate than Syrians, with c. 75% correct rejections. Syrians achieved only chance-level accuracy combined with low confidence in their correct rejections (c. 20% 'Highly unlikely' vs. c. 30% 'Likely'). They were also distinctly less certain than non-Syrians (c. 20% vs. c. 10% 'uncertain' responses).

10.5.3. Syrian stimuli

Figure 10.8 shows non-Syrian (N = 9) and Syrian listeners' (N = 3) response patterns to the six Syrian stimuli.

Figure 10.8: non-Syrian and Syrian listeners' response patterns to Syrian stimuli
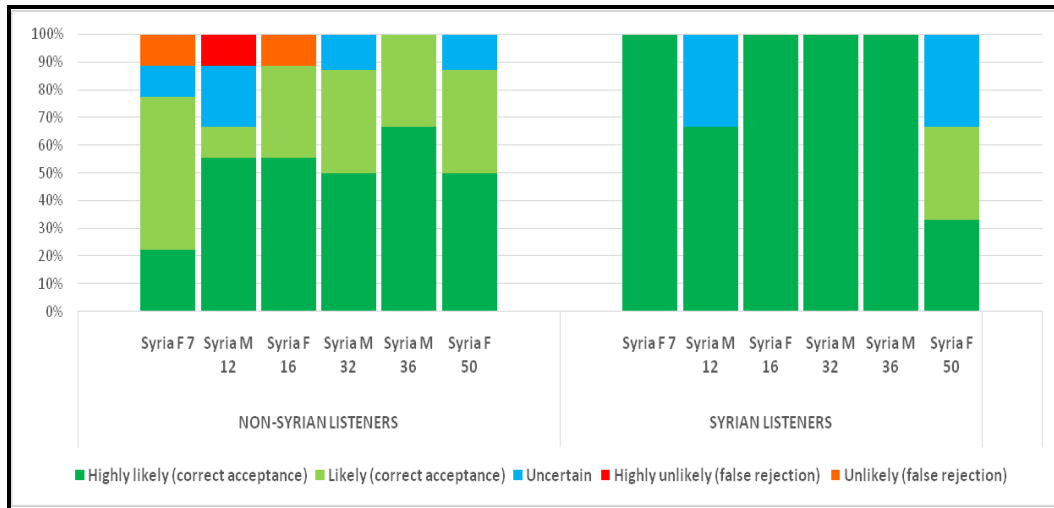
Among non-Syrians there were false rejections, at various degrees of certainty, of three stimuli; four attracted outright uncertain responses by at least one listener. Only one stimulus was correctly accepted by all non-Syrians. Even when accurate, non-Syrians were collectively less certain than Syrians in their responses; no stimulus was correctly accepted as 'highly likely' by all non-Syrians.

Four stimuli occasioned correct acceptances by all Syrians, to the maximum possible degree of certainty (i.e. 'highly likely'). To one of the stimuli there was a correct acceptance at lesser certainty ('likely'). Two stimuli occasioned an 'uncertain' response from one of the three Syrians. There were no false acceptances by Syrians. This pattern contrasts with that seen in Ar-1, where all responses by Syrians were correct to the highest degree of certainty.

10.5.4. Non-Syrian stimuli

Figure 10.9 demonstrates non-Syrian and Syrian listeners' response patterns to the 15 non-Syrian stimuli. Note that the naming of stimuli (M 29, F 47 etc.) signifies that male/female speakers are featured in the stimuli; the numbers are used for reference purposes only and do not refer to any attributes of the speakers.

Figure 10.9: non-Syrian and Syrian listeners' response patterns to non-Syrian stimuli
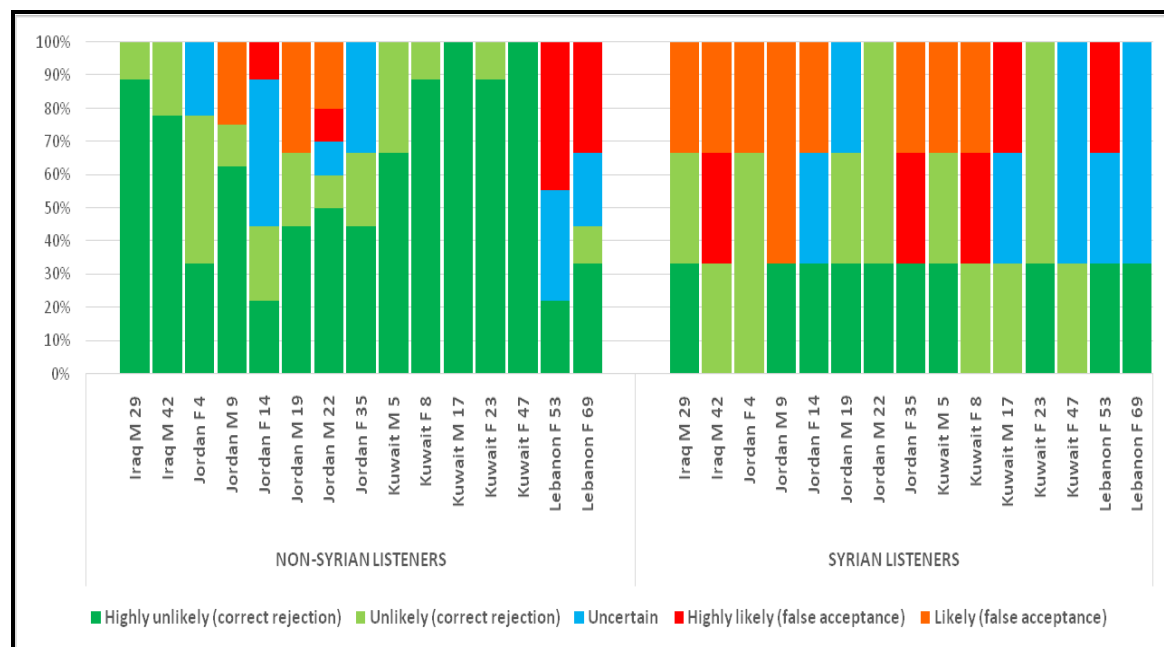
Non-Syrian listeners were overall more accurate than Syrians, though only two stimuli, Kuwait M 17 and Kuwait F 47, resulted in a homogeneous response pattern (100% 'Highly unlikely') among all non-Syrians.

Unanimously, non-Syrian listeners correctly rejected both of the Iraqi and all five Kuwaiti stimuli (i.e. all non-Levantine stimuli). Four of the six Jordanian stimuli (i.e. all except F 4 and F 35), however, occasioned some false acceptances by non-Syrians; F 4 and F 35 each provoked some 'uncertain' responses.

A similar pattern on Jordanian stimuli obtained among Syrian listeners. Four of the six(i.e. all except M 19 and M 22) occasioned a false acceptance by at least one Syrian listener, while M22 provoked only correct rejections and M19 a single 'uncertain' response.

Only two stimuli—one Jordanian and one Kuwaiti—were correctly rejected by all three Syrian listeners, while one Lebanese, one Jordanian and one Kuwaiti stimulus each attracted a combination of correct rejections and 'uncertain' responses. The two groups' respective response patterns to Jordan M 22 and Lebanon F 53 are interesting in the sense that they were the only non-Syrian stimuli on which Syrians entered more correct rejections than non-Syrians (100% vs. 60% on the former, 30% vs. 20% on the latter). All stimuli except Jordan M 22 and Lebanon F 69 were falsely accepted by at least one Syrian listener, though none were falsely accepted by all.

10.5.5. Discussion

The four chief findings here are broadly consistent with those gleaned from Ar-1. Firstly, on non-Syrian stimuli—especially those of the Iraqi and Kuwaiti (non-Levantine) type—non-Syrians were considerably more accurate and more certain than Syrians. However, this divergence is violated by a comparison of group response patterns to Levantine-type stimuli (i.e. Jordanian and Lebanese), on which non-Syrians evidence a response pattern more similar to that of Syrians, with roughly comparable overall uncertainty and inaccuracy. Differences in groups' perception of Levantine and non-Levantine varieties emerge once more in Ar-3.

Secondly, Syrians were again more accurate and more confident on Syrian than they were on non-Syrian stimuli. Thirdly, on Syrian stimuli Syrians were again more accurate than non-Syrians (providing further evidence of the validity of the other-accent effect), though here the difference was much less pronounced than it had been in Ar-1. Finally, there were once more no false rejections by Syrians of Syrian stimuli, though here (unlike in Ar-1) there were some uncertain responses. This last finding supports guiding hypothesis (i) of the thesis, that Syrians will show greater accuracy in identifying Syrian voice samples with Syria. As in Ar-1, however, a note of caution must be acknowledged: there were only three Syrian and nine non-Syrian listeners is Ar-2. The support for hypothesis (i) emerging from the inferential statistical results of Eng-2, and amplified by the raw data in Ar-1, is nevertheless replicated here.

Like in Ar-1, the response pattern of Syrians listeners shows a greater tendency towards false acceptances than false rejections. As we shall see, Syrians in Ar-3, too, show a strongly marked propensity in this direction. Here, however, unlike in Ar-1, non-Syrians also conform (marginally) to this pattern.

Also as in Ar-1, within-group inter-listener response variability was fairly low. On non-Syrian stimuli all nine non-Syrian listeners entered at least one 'uncertain' response, while six of the nine—with two Saudis as the exceptions—entered at least one false acceptance (one being the lowest number and five the highest). On the same kind of stimuli, all three Syrian listeners similarly entered at least one 'uncertain' response (four, three and one respectively) and at least one false acceptance (eight, four and two respectively).

On Syrian stimuli, only two of the nine non-Syrians entered at least one false rejection (one Saudi with two, and a second Saudi with one). Interestingly, neither of these Saudi listeners produced any false acceptances of non-Syrian stimuli. This, combined with the performance of the Egyptian and Jordanian listeners in Ar-1 (consult Chapter 9, subsection 9.8.5), perhaps suggests an inverse relationship between individual accuracy on non-Syrian and Syrian stimuli, respectively. None of the three Syrian listeners produced an incorrect rejection of any Syrian stimulus, though one of them entered two uncertain responses.

10.5.6. Stimuli selection for Ar-3

In this chapter, aside from the above survey of raw results to determine the strength of support for hypothesis (i), the other aim associated with the investigation of research objective 1 was determining retention/omission of stimuli for Ar-3. Once more, the guiding principle was maximising differences in response patterns between Syrian and non-Syrian listeners. Further accentuating the two groups' observed divergence in responses to Levantine and non-Levantine stimuli was of particular interest. Stimuli retained for Ar-3, and the specific grounds on which they were selected, were as follows.

All of the non-Levantine stimuli, apart from Kuwait F 23, had occasioned false acceptances or outright uncertain responses by Syrian listeners only. In the expectation that such differences would recur in Ar-3, the two Iraqi stimuli used here were retained for Ar-3, and two Kuwaiti stimuli (Kuwait F 8 and M 17) of the available five were selected because they had elicited the highest number of false rejections among Syrians.

Two Jordanian stimuli (Jordan M 9 and M 22) were retained. Jordan M 9 was selected because, among non-Syrians, it had occasioned the most heterogeneous response pattern of all Jordanian stimuli, and Jordan M 22 because it had provoked no false acceptances by Syrians.

The two Lebanese stimuli were retained because of the close similarities between Lebanese and Damascene Arabic, the expectation being that differences in response patterns among Syrians and non-Syrians would re-emerge, and to a greater extent, in Ar-3.

This total of eight non-Syrian stimuli was augmented by four more. These had been featured in Ar-1 but, because no respondent had falsely accepted them in Ar-1, were excluded from Ar-2. They were chosen at random for Ar-3: two Egyptians (F 2 and M 13) and two Moroccans (M 6 and F 21). The motivation for this decision was to investigate whether responses would be again, as in Ar-1, unanimously correct among a (projected) larger pool of listeners in Ar-3. Egyptian and Moroccan stimuli were chosen over the available Omani or Tunisian stimuli because Egypt and Morocco are both the source of a greater number of asylum applicants than either Oman or Tunisia.

To these 12 non-Syrian stimuli were added, in the interests of balance, all of the 10 available Syrian stimuli. Since all of the Syrian stimuli (except M 36 and M 43) had occasioned false rejections or outright uncertainty among non-Syrian listeners in one or both Ar-1/Ar-2, but no false rejections among Syrians, their inclusion was predicted to further accentuate differences in response patterns between the two listener groups.

10.5.7. Listeners' demonstrated awareness of Syrian dialectology

As was the case with the linguist listeners in Ar-1, comments by the NSNL listeners in Ar-2 indicate strong metalinguistic awareness of cross-border varieties. This revealed knowledge encompasses both the close relatedness of the Levantine dialects and the existence of non-Levantine varieties in Syria.

Syrians and non-Syrians alike left comments expressing such awareness, which appears to have led them to enter 'uncertain' responses to some of the stimuli and incorrect responses to others (mostly false acceptances). As stated of the same tendency among linguists in Ar-1, inaccuracies and outright uncertainty here appear to proceed, in at least some cases, from what non-linguist listeners *do* know rather than what they *do not* know.

As in Ar-1, the comments reproduced below come from two sources: first, those made in connection with judgements of individual stimuli (in associated free-text boxes); and, second, the free-text box at the conclusion of the survey. Comments by Syrian listeners included the following.

- *"Some areas near Palestine have a dialect close to Palestine, and the Albu Kamal area is close to Iraq, and their dialect is close to Iraqi"* (a comment on the stimuli in general)

- *"'It can be for the people of the desert"* (false acceptance of a Kuwaiti stimulus)

- *"It looks a little Iraqi, but it could be the people of the desert in Syria"* (false acceptance of an Iraqi stimulus)

- *"It could be for the people of the desert in Syria"* (an 'uncertain' response to a Kuwaiti stimulus)

- *"Because of the mixing of words, it is very similar, but there is a word that is not Syrian, and it (Kalash) belongs to Iraq, and the Iraqi and Syrian Bedouin languages are very, very similar"* (an 'uncertain' response to a Kuwaiti stimulus)

Comments by non-Syrians in the same vein included the following.

- *"There is a dialectic overlap between Iraq, Syria and Jordan because of the geographical proximity, which makes some dialects actually exist in Syria and are considered Syrian even though their fingerprints indicate that they are from those countries neighboring Syria."* (a comment on the stimuli in general by an Iraqi listener)

- *"Syria, maybe from Daraa"* (an 'uncertain' response to a Syrian stimulus by an Iraqi listener)

- *"It could be the dialect of the countryside of Syria"* (an 'uncertain' response to a Jordanian stimulus by a Saudi listener)

NSNLs' general willingness to accept such marginal cases as Syrian, or at least to enter 'uncertain' responses to them, demonstrates that NSNLs are eminently capable, contra Fraser (2009, 2011, 2019), of bringing caution to bear on their conclusions. The awareness of cross-border varieties demonstrated by the above comments (and those of a similar nature reviewed in Chapter 9) also supports the argument made out in Chapter 4 that non-linguists are not necessarily constrained by 'folk views' suffused with an 'ideology of homogeneism'. This argument is strengthened by the clear bias of the predominantly NSNL listeners towards acceptances, including false acceptances. The latter in particular is the opposite of what would be predicted by the theory of 'homogeneism'.

In addition, Syrian listeners here entered a proportionately larger number of comments of this type than did non-Syrians: there were five comments by Syrians (N = 3) but only three by non-Syrians (N = 9). A similar pattern is observable in Ar-1 (Chapter 9, subsection 9.8.1.7), where the two Syrian listeners entered 20 comments compared to four by the eight non-Syrians. This may be interpreted to mean that Syrians retain greater awareness than non-Syrians of the presence in Syria of non-Damascene varieties of Arabic which also occur in neighbouring countries. If this interpretation is accepted, it may be taken as an explanation of Syrians' greater revealed tendency towards false acceptances and decreased confidence in response to non-Syrian

stimuli, as most plainly illustrated in summative form by Figure 10.7 (and by Figure 9.10 in Ar-1).

It is also notable that, on Syrian stimuli, Syrians demonstrate a greater degree of overall accuracy, certainty (in Likert-scale ratings of 'decision' responses) and confidence (i.e. decision vs. 'uncertain' responses). Perhaps most tellingly, Syrians are less inclined to reject non-Syrian—or, more, specifically non-Damascene—stimuli than are non-Syrians. This may be because of Syrians' more acute awareness of the presence in Syria of non-Damascene varieties of Arabic.

10.6. Research objective 2: results

10.6.1. Organisation

As in Ar-1, research objective 2 involves investigating the validity of hypothesis (ii): speaker-listeners will vary by place of origin in the number and kinds of cues they recognise. Research objective 2 is investigated by comparing the type and number of cues cited by Syrian and non-Syrian listeners, respectively, in response to the question, 'What clues led you to your judgement of the speaker's accent?'

See section 9.9.1, Chapter 9 for a full explication of the salience-diagnostic value framework likewise employed here. For easy reference, Table 10.3 summarises this framework.

Table 10.3: Classification of feature citations by listener response type

Classification	Response type
Diagnostic	Correct
Salient	Correct, incorrect or uncertain

Potentially salient/diagnostic cues present in the stimuli, according to sources consulted on Arabic dialects, were identical to those in Ar-1. Refer to Chapter 9, Tables 9.1 and 9.2 for details. The reflexes in the various dialects of the features, and the locations of their occurrence

in the stimuli, are covered in detail in Chapter 9, subsections 9.9.2 to 9.9.5, and are not repeated here.

Below, results are organised here in broadly the same manner as laid out in subsection 9.9.1. of Chapter 9. In subsections 10.6.2 to 10.6.5, listener citations of individual features, belonging to the four classes of phonetics/phonology, morphosyntax, suprasegmentals and lexis, are specified. Citations of each feature are then examined for the extent to which each seems to have guided listeners' judgements.

Subsection 10.6.6 covers citations of features *not* mentioned in the consulted dialectological literature (there were none). Subsection 10.6.7 (with Tables 10.4 and 10.5), summarises feature citations by Syrian and non-Syrian listeners respectively. Subsection 10.6.8 is an interpretative discussion of observed divergences between features cited by Syrian/non-Syrian listeners, and a comparison with patterns evidenced in Ar-1. The degree of support for hypothesis (iii) inferable from these results is also considered in 10.6.8.

10.6.2. Phonetics/Phonology
MSA /q/

Considering the results of Ar-1, the expectation was that this feature would have considerable salience as a marker of Damascene Arabic, though not perhaps so much in the way of diagnostic value, since not all respondents in Ar-1 had correctly accepted or rejected any stimulus while commenting on it. Owing to variability in the realisation of MSA /q/ in the Lebanese stimuli (specified in section 9.9.2 of the previous chapter) the feature was predicted to be of still lesser value in distinguishing Damascene from Lebanese Arabic.

Four listeners in Ar-2 referred to this feature, a total of five times. A Saudi listener correctly identified two Jordanian stimuli as not Syrian, and to the highest available degree of certainty, while citing in both cases the "pronunciation of the letter qaf". But whether this was a general remark on the pronunciation of MSA /q/ in the two stimuli or referred particularly its occurrence in line 1 or, alternatively, in line 7, is unknown.

An Iraqi listener was uncertain about the provenance of a Jordanian stimulus but commented, "It sounds Jordanian through the word (village) and others", a clear reference to line 7. A Syrian-born listener, also uncertain as to the origin of a Jordanian stimulus, cited the speaker's pronunciation of "the word village". A Moroccan listener, likewise unable to accept or reject a Jordanian speaker as Syrian, remarked on the pronunciation of "village". In each case it seems likely that listeners were referring to realisations of /q/, though it is of course possible that some other property of the word was at issue.

In all instances it is notable that the stimuli involved: (1) were all Jordanian; and (2) could be neither accepted nor rejected by the listener as Syrian. As in Ar-1, MSA /q/ was the feature most generally commented upon, so the prediction of its salience seemed to be broadly supported. Though it evidenced a lesser degree of diagnostic value than it had in Ar-1, it was still superior in this respect to any other feature (see below, Tables 10.4 and 10.5).

MSA/ʤ/

The results of Ar-1 suggested that this feature would be salient for listeners but that listener accuracy, in instances where it was cited, would be mixed—i.e. its diagnostic value would be negligible. In the event only one (Syrian) listener unequivocally seems to have cited it: in connection with a Jordanian stimulus, where the comment was, "The sound and the manner of pronouncing Joha". In this instance, the listener incorrectly accepted the stimulus as 'likely' Syrian. There was a second reference, by the same listener, to this word (*Juha*) in line 1 of the text, in which reflexes of MSA /ʤ/ occur. This reference was in connection with a Lebanese stimulus. Here the listener provided an 'uncertain' response.Overall, MSA /ʤ/ was less salient than it had been in Ar-1 and evinced lesser diagnostic value.

MSA/ð/

This feature was explicitly mentioned three times by two non-Syrian listeners in Ar-1, and proved fairly strongly diagnostic, but it was not explicitly acknowledged by listeners here (however, see subsection 10.6.7, below).

Final imāla

In contrast to Ar-1—in which final imāla was mentioned by two respondents, one of them Syrian, but showed little diagnostic strength—no respondent explicitly cited it here.

10.6.3. Morphosyntax

Adjective negation

In Ar-1 two (non-Syrian) listeners cited adjective negation a total of seven times in correctly identifying two Syrian stimuli—a one-to-one relationship between salience and diagnostic strength. Here it was commented upon a total of six times by three listeners: a Moroccan, in connection with three Jordanian and a Lebanese stimulus, each misidentified as Syrian; an Iraqi with linguistic training, in association with the same Lebanese stimulus and a correct rejection; and a Syrian, also linguistically trained and again in remarking on the same Lebanese stimulus, with an 'uncertain' response.

The Moroccan listener's comments in every instance cited the occurrence of the adjective negation phrase itself without elaboration of what properties had led to the relevant conclusion. It is important to note that in Jordanian Arabic this adjective negation phrase commonly takes the form /miʃkwajsi:n/, (compare to Damascene /mu: mna:ħ/). It would seem, then, that either this Moroccan listener is unaware of the lexico-morphological distinction between the Damascene and Jordanian forms or that recognition of /miʃkwajsi:n/ as a form occurring in Syria demonstrates the listener's knowledge of cross-border varieties.

The Iraqi listener appears to focus on the phonetic properties of the word for 'good'. The comment was, "Mneih: This is a severe inclination found in the Lebanese dialect". This may be a reference to phonetic properties of the final vowel.

While the considerable diagnostic strength of this negation phrase, inferred from Ar-1, was not supported here, there was further evidence of its salience. This is further demonstrated in the results of Ar-3 (Chapter 11).

The preposition *in*

As in Ar-1, this feature was unmentioned by any respondent.

Future marker

Again as in Ar-1, this feature was not commented upon here.

3MSg marker

See above.

3PLSubj marker

In Ar-1 this feature failed to attract comment. Here, a Syrian listener seems to have cited it (as "they know you") in incorrectly identifying a Kuwaiti stimulus as Syrian.

Interestingly, the variant of the marker for 3PLSubj in Damascene Arabic is /jaʕirfuː/, whereas in Kuwaiti Arabic it is /jiʕirfuːn/ (note /–uː/ vs. /–uːn/. The Syrian listener thus misidentified the stimulus in question as Syrian partially on the basis of a stereotypically non-Syrian—or at least non-Damascene/non-Levantine—realisation /jiʕirfuːn/, which is nonetheless dominant in some non-Levantine (especially Bedouin-type) varieties spoken in Syria. Its citation may be a demonstration of the listener's awareness of linguistic diversity in Syria; alternatively, it may have resulted from a simple mishearing or misapprehension of the dominant form of the morpheme in (Levantine) Syrian dialects. Whatever the case, as in Ar-2 this feature evinced both low salience and negligible diagnostic value.

10.6.4. Suprasegmentals

Stress

Stress patterns were not mentioned by any listener.

Vocalic mass/speech rhythm

See above.

Syllable Structure

See above.

Intonation

See above.

10.6.5. Lexis

As in Ar-1, a number of listeners mentioned lexis as diagnostic of Syrian Arabic. These references were made only by Syrian listeners.

A Syrian cited the phrase "Deir Bolk" (*be careful*) in falsely accepting an Iraqi stimulus, as had an Egyptian while correctly accepting three Syrian stimuli in Ar-1. The possibility remains open that these references had to do with something aside from lexis only, as the same phrase is used in the Iraqi, Jordanian and Kuwaiti stimuli. It could also be that this Syrian listener believes that the phrase is only employed in Syria; there were several false acceptances of Iraqi, Jordanian and Kuwaiti stimuli among his/her responses.

A second, linguistically-trained, Syrian listener gave an 'uncertain' response to a Jordanian stimulus while echoing comments in Ar-1 by citing the form "Hadhoul". This word had been mentioned three times by two (non-Syrian) listeners in Ar-1 and was treated as a likely, though not certain, reference to varying reflexes of MSA /ð/. Again this reference, combined with the

'uncertain' response, suggests the listener is aware that cross-border varieties are spoken on the Syrian-Jordanian frontier.

Other patterns of citation of lexical items in Ar-2 were not carried over from Ar-1. The word /kil:iʃ/ (*good*), which occurs in the Iraqi and Kuwaiti stimuli, was mentioned twice here but not at all in Ar-1. A Syrian listener cited "kalash" as non-Syrian in providing an outright uncertain response as to the provenance of a Kuwaiti stimulus (this listener also commented on the "mixing of words" in the stimulus concerned). A second Syrian listener, this one trained in linguistics, similarly rejected "kalash" as a Syrian word while giving an outright uncertain response on the origin of a second Kuwaiti stimulus. Once more, as discussed elsewhere in this chapter and in Chapter 9, these responses suggest that Syrian listeners especially are strongly aware of the presence of non-Levantine varieties of Arabic spoken in Syria.

10.6.6. Comments on features not mentioned in the literature

Other than in the domain of lexis, comments on features *not* mentioned in the literature were entirely absent from Ar-2. This had also been the case in Ar-1.

10.6.7. Specification of features cited, by listener group

Tables 10.4 and 10.5 specify, for Syrians (N = 3) and non-Syrians (N = 9) respectively, the number of citations of individual features ('salient') and the number of correct responses ('diagnostic') associated with them.

Table 10.4: Features cited by Syrian listeners, by salience and diagnostic value

Feature	/q/	/d͡ʒ/	/ð/	Final imāla	Adjective negation	Preposition *in*	Future marker	3MSg	3PLSubj	Suprasegmentals (all types)	Totals
Salient	1	2	0	0	1	0	0	0	1	0	5
Diagnostic	0	0	0	0	0	0	0	0	0	0	0
Totals	1	2	0	0	1	0	0	0	1	0	5

Table 10.5: Features cited by non-Syrian listeners, by salience and diagnostic value

Feature	/q/	/d͡ʒ/	/ð/	Final imāla	Adjective negation	Preposition *in*	Future marker	3MSg	3PLSubj	Suprasegmentals (all types)	Totals
Salient	4	0	0	0	5	0	0	0	1	0	9
Diagnostic	2	0	0	0	1	0	0	0	0	0	3
Totals	6	0	0	0	6	0	0	0	1	0	12

As in Ar-1, in the domain of phonetics/phonology, the various reflexes of MSA /q/ evidence the greatest degree of diagnostic value for non-Syrians. There was only a single (non-diagnostic) citation of MSA /q/ reflexes by Syrians.

As stated above, MSA /q/ was commented upon only in connection with Jordanian stimuli. It was thus apparently diagnostic of Jordanian speech for non-Syrians in some cases. However, among Syrians, it may have prompted uncertainty as to whether the speaker at issue was from an area near the Jordanian frontier (see subsection 10.5.7, above, for further discussion of Syrians' revealed awareness of border varieties).

However, in contradistinction to the results of Ar-1, reflexes of MSA /d͡ʒ/ and /ð/ were not salient for the non-Syrians in Ar-2. Of these two phonemes, only /d͡ʒ/ apparently evinced salience for Syrians, though without any accompanying diagnostic value. Suprasegmentals entirely lacked salience for either listener group, whereas in Ar-1 contrastive stress in Syrian – Egyptian Arabic had proved highly diagnostic, albeit for a single listener (Egyptian stimuli were not featured in Ar-2).

In further replication of the results of Ar-1, few morphosyntactic features evidenced salience (except, minimally, adjective negation and the 3PLSubj marker), and none diagnostic value, for Syrians. For non-Syrians only adjective negation was both salient and diagnostic among features of this class, as it had also been in Ar-1.

10.7. Discussion

Four general observations of particular interest arise from the preceding review of listener comments in Ar-2 and a comparison with the same in Ar-1.

First, in both experiments, the only features cited (i.e. salient), other than lexis, were those attested in the available dialectological literature.

Second, in neither experiment did Syrian listeners apparently make any correct decisions on the basis of these features (i.e. none of the features was diagnostic).

Third, only Syrians in Ar-2 cited lexical cues, but these do not appear to have held much diagnostic value.

Fourth, this range of features appears to have been a great deal more salient and diagnostic for non-Syrian linguists (i.e. the listeners in Ar-1) than for non-Syrian NSNLs (most of the listeners in Ar-2). The non-Syrian linguists in Ar-1 entered 31 comments on these features, of which 25 proved diagnostic. Their NSNL counterparts in Ar-2 entered only nine comments, of which three were diagnostic.

This observation must be treated with caution, since there were 52 stimuli in Ar-1 compared to only 21 in Ar-2, and a greater range of nationalities was represented in Ar-1 than in Ar-2. Notwithstanding this imbalance, all of the comments in Ar-1 were made in connection with either Syrian or Jordanian stimuli, and all except four of the Syrian stimuli used in Ar-1 were also used in Ar-2.

Together these observations suggest that the (predominantly non-linguist) listeners in Ar-2, as well as the Syrian linguist listeners in Ar-1, may have relied to some unknown extent on a range

of cues other than those captured in the reviewed dialectological surveys. These may have been intonational properties (e.g. of the types discussed by Behnstedt 2012), or they may have been fine distinctions at the level of the segment, of the kinds proposed by Nolan (2012) and Klein (1988). It is also possible that they were a concatenation of both. In any case, the inference at this stage is that differences in cue recognition may be present among, on the one hand, Syrians in general/non-Syrian non-linguists and, on the other, non-Syrian linguists.

Again as in Ar-1, with proper acknowledgement of caution owing to the small number of respondents, hypothesis (iii) of the thesis appears to find some support in the results of Ar-2. There is a clear pattern, apparent in the results of both experiments, concerning: (1) the absence of recognised salience/diagnostic value in the reviewed features, as evidenced in comments by Syrians; and (2) the usefulness of the same to non-Syrian linguists in Ar-1, but *not* to the non-Syrian NSNLs in Ar-2. This suggests that, in addition to national background, education may be a determinative factor in cue recognition. Further exploration of the validity of hypothesis (iii) is undertaken in Ar-3.

Chapter 11: Arabic experiment 3

11.1. Introduction

The background to the series of Arabic-language experiments to which this experiment (Ar-3) belongs is detailed in Chapter 9, section 9.1. Listeners were once more presented with short voice samples and asked whether and to what degree of confidence (on a five-point Likert scale) they were prepared to associate the samples with Syria. Once more, the question asked was, 'Is this a Syrian accent?' Ar-3 diverges from the two previous experiments in this series in the number and types of stimuli selected, in the range of countries from which respondents (here, native and non-native speakers of Arabic, linguists and non-linguists) and, most crucially, in being a much larger study than either Ar-1 or Ar-2.

11.2. Ar-3: Hypotheses

In this chapter I first attempt to evaluate via descriptive statistics, and then through inferential statistics, the extent of support for hypotheses (i) and (ii). Hypothesis (i) is that Syrian listeners will be significantly more accurate in identifying Syrian speakers in the stimuli. Hypothesis (ii) is that linguistic education will not reveal a statistically significant effect on differences in accuracy, as predicated by hypothesis (i).

I then endeavour, as in Ar-1 and Ar-2, to evaluate the strength of hypothesis (iii) by examining the number and types of cues cited by the various kinds of listeners recruited for this experiment.

For reasons specified below, Ar-3 is therefore the only experiment in the Arabic series that attempts to assess support for all three of the thesis' hypotheses.

11.3. Ar-3: Research objectives

The research objectives of Ar-3 are identical to those of Ar-1 and Ar-2. Research objective 1 represents the operationalisation of hypothesis (i). It involves presenting audio stimuli featuring Syrian and non-Syrian Arabic speakers to two groups of listeners, also Syrian and non-Syrian, and analysing their respective response patterns to the question, 'Is this a Syrian accent?

An inferential statistical investigation of research question 1—and through it hypothesis (i)—was precluded by the small number of respondents to the two prior experiments in the Arabic series. This limitation does not apply to Ar-3, since a total of 79 listeners was recruited.

Research objective 2 is the operationalisation of hypothesis (iii). The results gathered here—together with those gleaned in Ar-1 and Ar-2—are used to inform the production test proposed in Chapter 12. In Chapters 9 and 10 the investigation of research objective 2 was somewhat limited by the fact that responses were elicited solely from, respectively, Arabic-speaking linguists and Arabic-speaking NSNLs. Here, however, Arabic native speakers in general were recruited, independent of educational profile. The recruitment of listeners of various educational (as well as national) backgrounds means that research objective 2 is more fully explored here than previously.

11.4. Methodology

11.4.1. Task design

Task design and question display differed from that of Ar-2 (Chapter 10, subsection 10.4.1) only in the number and national origins of stimuli used (subsection 11.4.2, below). As in all previous experiments, stimuli were uploaded to Soundcloud, to which they were linked in Qualtrics.

As in Ar-2 (but not Ar-1), all task-related material was translated into Arabic by a native speaker of the language. Once more, this was done to allow participation by listeners who do not speak/read English. Consult Chapter 10, subsection 10.4.1 for the layout of material in Qualtrics.

11.4.2. Listeners

As in Ar-1 and Ar-2, listeners were recruited through contacts at the universities of York and Newcastle. Preliminary questions about listeners' national, linguistic and educational background appeared exactly as they had in Ar-2 (Chapter 10, subsection 10.4.1).

Listeners who completed the survey in full numbered 79. Of native Arabic speakers, 31 were originally from Syria and 43 from various other countries: Lebanon (4), Egypt (5), Jordan (6), Palestine (12), Saudi Arabia (11), Tunisia (1), Sudan (1) and the United Arab Emirates (3). Of native Arabic speakers who declared postgraduate-level linguistic training (classified therefore as linguists), 10 were Syrian and 22 non-Syrian.

The remainder of the native speaker-listeners comprised 21 Syrian and 21 non-Syrian NSNLs. There was one additional NSNL of unspecified origin, with an IP in the United Arab Emirates, whose results could not be counted.

Five non-native speaker-listeners of American nationality also (fortuitously) responded. All were linguists, trained to postgraduate level, and associated with universities in the United States. Table 11.1 specifies the national origins and number of native Arabic-speaking listeners.

Table 11.1: National origins and number of native Arabic-speaking listeners (five non-native speaker linguists not included)

	Syria	Lebanon	Egypt	Jordan	Palestine	Saudi Arabia	Tunisia	Sudan	UAE	Totals
PGling	10	2	1	4	7	5	1	0	2	32
NSNL	21	2	4	2	5	6	0	1	1	42
Totals	31	4	5	6	12	11	1	1	3	74

11.4.3. Stimuli

11.4.3.1 Selection

See subsection 10.5.6., Chapter 10 for the rationale behind stimuli selection for this experiment. The same extract of the *Juha* story was employed as in Ar-1 and Ar-2; consult

Appendix C for transcripts. Stimuli were once more presented in pseudo-random order.

Ar-1 included 52 stimuli from eight countries of origin. Ar-2 featured 21 stimuli from five countries. In Ar-3 there were 22 stimuli from seven countries: 10 of Syrian and 12 of non-Syrian speakers. Of the latter, two stimuli featured Lebanese speakers, and the remainder were from Iraq (two), Jordan (two), Kuwait (two), Egypt (two) and Morocco (two). Table 11.2 shows the number of stimuli and the national origin of each, as used in Ar-3.

Table 11.2: National origins and number of stimuli

Origin	Syria	Lebanon	Egypt	Jordan	Kuwait	Morocco	Iraq	Total
Number	10	2	2	2	2	2	2	22

11.5. Research objective 1: results

11.5.1. Organisation

The following subsections are principally devoted to investigating results connected with research objective 1: How and to what extent response patterns of the five groups—Syrians and non-Syrian linguists/NSNLs and USA-based non-native linguists—varied when confronted with the question, 'Is this a Syrian accent?' As stated in subsection 11.3, this objective relates directly to determining the extent of support for guiding hypothesis (i). In line with this objective, I first examine, though the lens of descriptive statistics, the response patterns of the five listener groups (subsection 11.5.2). I then analyse group accuracy on the basis of inferential statistics (subsections 11.5.3.2 and 11.5.3.3).

In subsections 11.5.3.4 and 11.5.3.5, I investigate listener confidence via inferential statistics. Like in Eng-2, confidence is conceptualised as residing in the demonstrated tendency of various groups towards, on the one hand, outright 'uncertain' responses and, on the other hand, 'decision' responses (to any degree of confidence). The examination of confidence is undertaken in order to assess empirically, in a LAAP-like context, the validity of the assertion that linguists are more willing than non-linguists "...to say 'I'm not sure.'" (Fraser 2019, p. 74).

On the rationale behind the investigation of confidence thus conceived, see Chapters 5 and 8.

In subsection 11.5.4 I evaluate the validity of the argument that confidence and accuracy among NSNLs is poorly correlated (cf. Fraser 2009, Patrick 2010). I do so via a formal investigation of the correlation between confidence and accuracy in the combined responses of all linguist and all NSNL groups, regardless of 'localness'. In this analysis I take into account the various degrees of confidence expressed by listeners on the five-point Likert scale.

Finally, I examine listener comments for what they reveal about the respective groups' awareness of Syrian dialectology (subsection 11.5.5) and review these results in the light of preceding analyses of group accuracy and confidence. This examination parallels those conducted in Ar-1 and Ar-2 (cf. Chapters 9 and 10), to which the equivalent set of findings of Ar-3 are also compared.

11.5.2. Descriptive statistics

11.5.2.1. Group responses to Syrian stimuli

Syrian NSNLs

Figure 11.1 shows Syrian NSNLs' (N = 21) response patterns to the 10 Syrian stimuli.

Figure 11.1: Syrian NSNLs' responses to Syrian stimuli
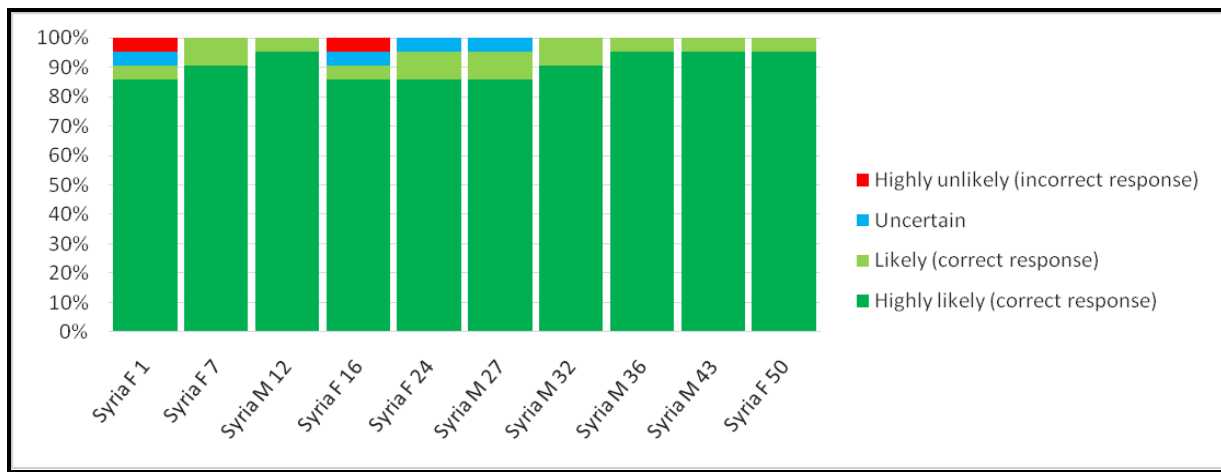
Six stimuli were correctly accepted by all listeners. To these, 90% or more responded with the maximum possible degree of confidence (i.e. 'highly likely'). However, no Syrian stimulus attracted only 'highly likely' responses.

The remaining four stimuli occasioned a more heterogeneous range of responses. Two, Syria F 1 and Syria M 12, each provoked one incorrect ('highly unlikely') and one 'uncertain' response, and two others (Syria F 24 and Syria M 27) each a single 'uncertain' response.

Syrian Linguists

Figure 11.2 shows Syrian linguists' (N = 10) response patterns to the 10 Syrian stimuli.

Figure 11.2: Syrian linguists' responses to Syrian stimuli
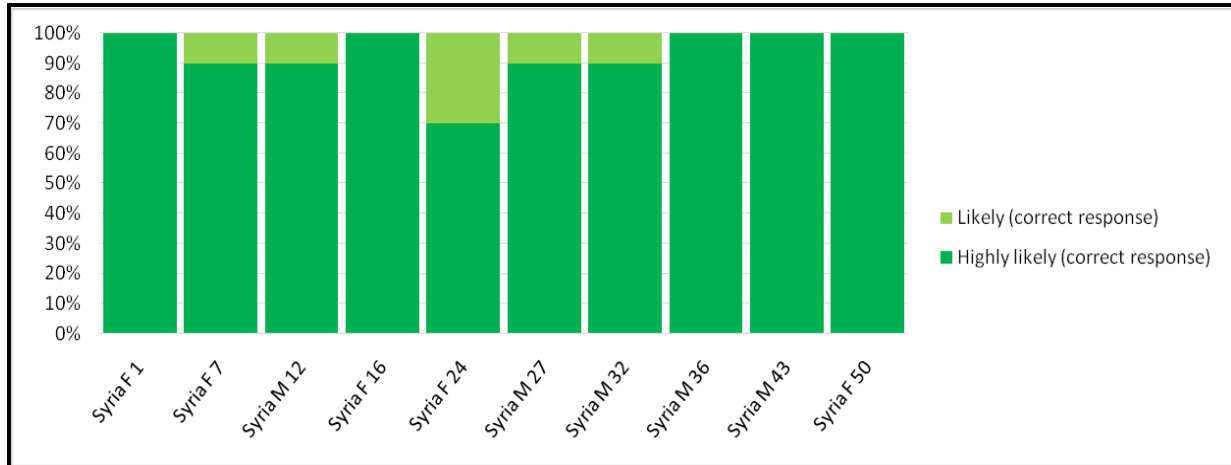
All 10 stimuli were correctly identified by all listeners. Five stimuli occasioned solely 'highly likely' responses. Of the remaining five, four resulted in a single 'likely' response each, while one (Syria F 24) provoked three 'likely' responses.

Non-Syrian NSNLs

Figure 11.3 shows non-Syrian NSNLs' (N = 21) response patterns to the 10 Syrian stimuli.

Figure 11.3: non-Syrian NSNLs' responses to Syrian stimuli
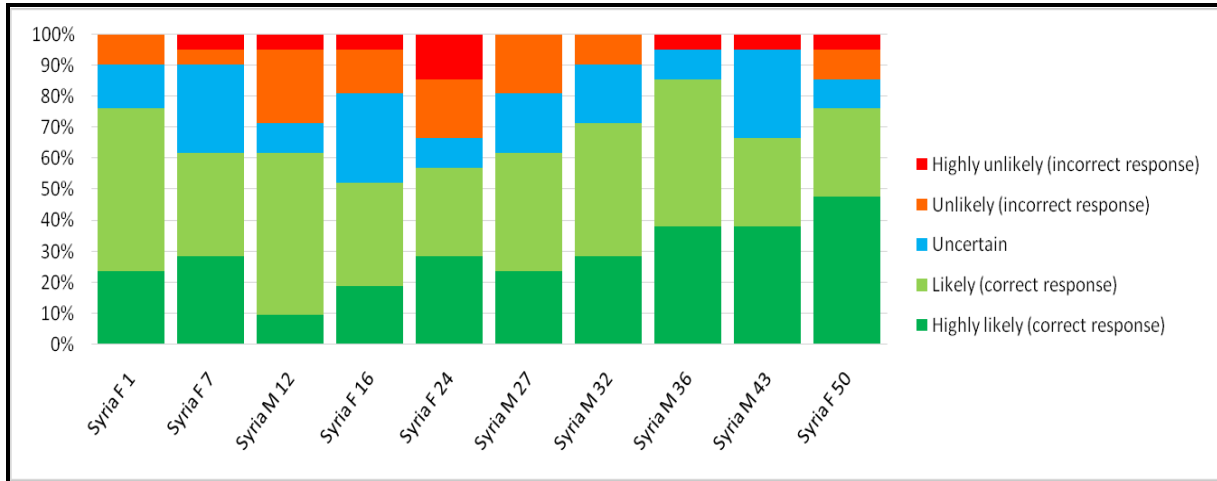
In contrast to the patterns shown in Figures 11.1 and 11.2 for Syrian listeners, no stimulus was correctly accepted by all non-Syrian NSNLs. To only one stimulus (Syria M 36) did correct acceptances exceed 80%, with the balance occasioning between c. 60 and c. 75% correct acceptances. Syria F 16 and Syria M 24 each occasioned a correct acceptance rate of under 60%.

Degree of confidence was generally low. Each stimulus occasioned at least two outright 'uncertain' responses, and there were overall more 'unlikely' than 'highly unlikely' as well as more 'likely' than 'highly likely' responses.

Non-Syrian linguists

Figure 11.4 shows non-Syrian linguists' (N = 22) response patterns to the 10 Syrian stimuli.

Fig 11.4: non-Syrian linguists' responses to Syrian stimuli
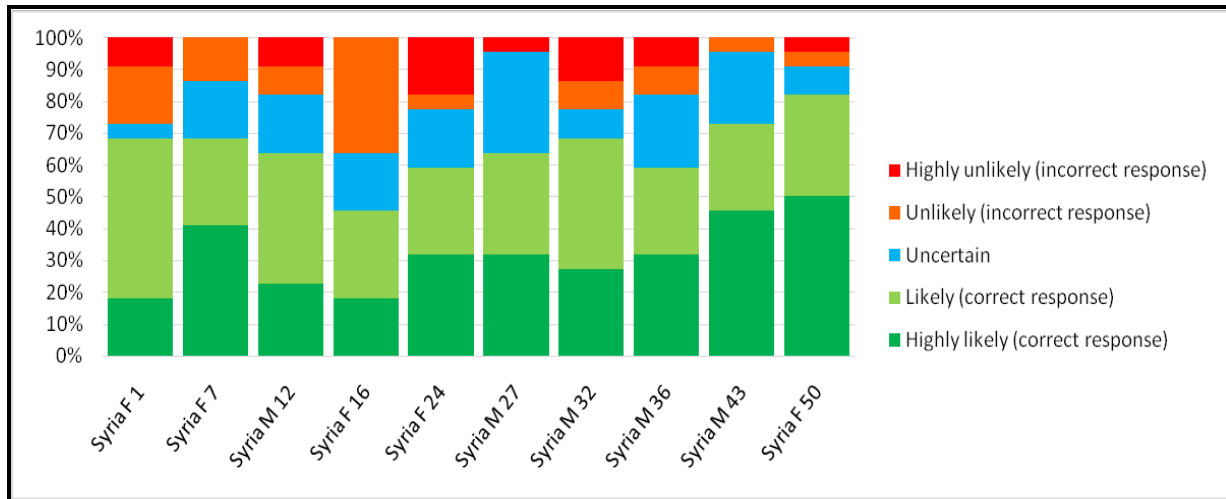
As in Figure 11.3, no stimulus was correctly accepted by all listeners. Again, in only one stimulus (this time Syria F 50) did correct acceptances exceed 80%, with the balance occasioning between c. 60 and c. 75% correct acceptances. This pattern is strikingly similar to that shown in Figure 11.3 but quite unlike those displayed in Figures 11.1 and 11.2. Syria M 24 and Syria M 36 each provoked a correct acceptance rate of under 60%, and Syria F 16 under 50%.

Another evident similarity to the response pattern of non-Syrian NSNLs in Figure 11.3 was the generally low degree of confidence. Each stimulus occasioned at least one outright 'uncertain' response, and there were overall marginally more 'unlikely' than 'highly unlikely' as well as 'likely' than 'highly likely' responses.

Non-native linguists

Figure 11.5 shows non-native linguists' (N = 5) response patterns to the 10 Syrian stimuli.

Figure 11.5: Non-native linguists' responses to Syrian stimuli
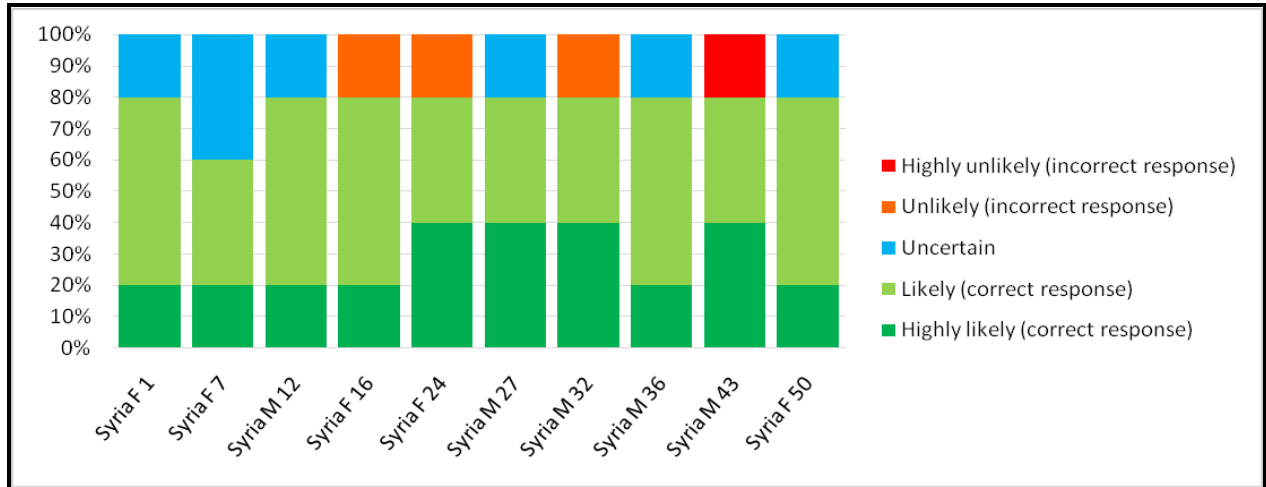
As in Figures 11.3 and 11.4, no stimulus was correctly accepted by all listeners. Non-natives, however, were more accurate than either of the non-Syrian groups: correct acceptances, to all stimuli except one (Syria F 7), amounted to 80%. However, again as in Figures 11.3 and 11.4, degree of confidence was overall low. Six of 10 stimuli occasioned at least one outright 'uncertain' response, and there were overall more 'unlikely' than 'highly unlikely' as well as 'likely' than 'highly likely' responses.

11.5.2.2. Group responses to non-Syrian stimuli

Syrian NSNLs

Figure 11.6 shows Syrian NSNLs' (N = 21) response patterns to the 12 non-Syrian stimuli.

Figure 11.6: Syrian NSNLs' responses to non-Syrian stimuli
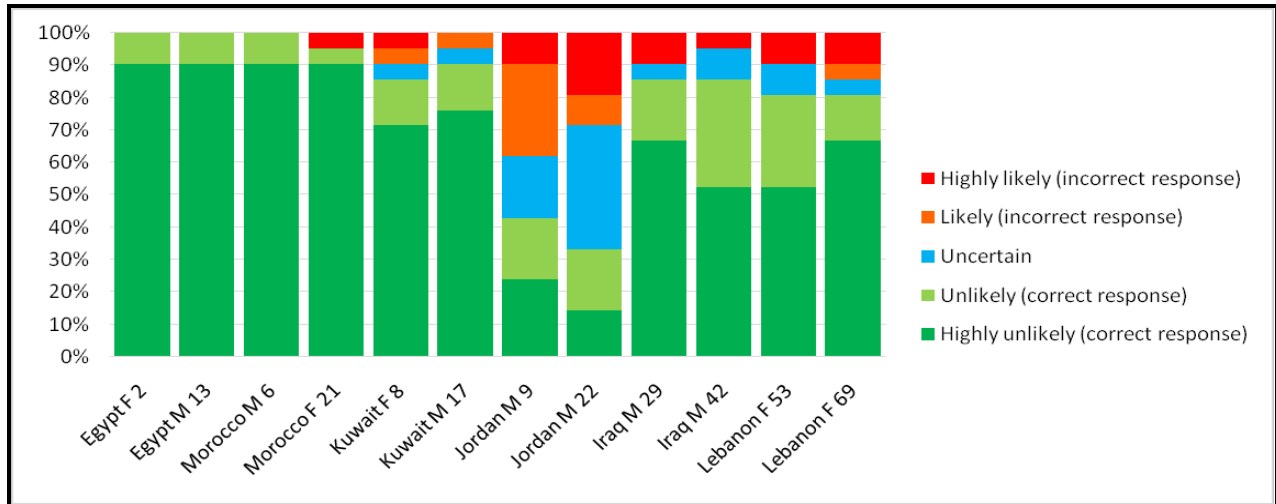
Only three stimuli (the two Egyptians and one of the two Moroccans) attracted 100% correct rejections. Generally, Syrian NSNLs' pattern of responses to non-Syrian stimuli was far more heterogeneous than to Syrian stimuli (cf. Figure 11.1), with many more outright false acceptances and a lower degree of confidence throughout.

Particularly notable is the highly variegated performance of this group on the two Jordanian stimuli: in each case correct rejections constituted less than 50% of the total. There was also a relatively large number of 'uncertain' responses to these two stimuli, especially Jordan M 22. Even so, Jordanian stimuli excluded, correct rejections exceeded 80% in all instances. Interestingly, the two Lebanese stimuli—which, like Jordanian and (Damascene) Syrian, is of the Levantine group—occasioned a similar response pattern to the non-Levantine Iraqi and Kuwaiti stimuli. In short: Syrian NSNLs were less accurate and less certain on Jordanian than on other stimuli, including those from Lebanon.

Syrian linguists

Figure 11.7 shows Syrian linguists' (N = 10) response patterns to the 12 non-Syrian stimuli.

Figure 11.7: Syrian linguists' responses to non-Syrian stimuli
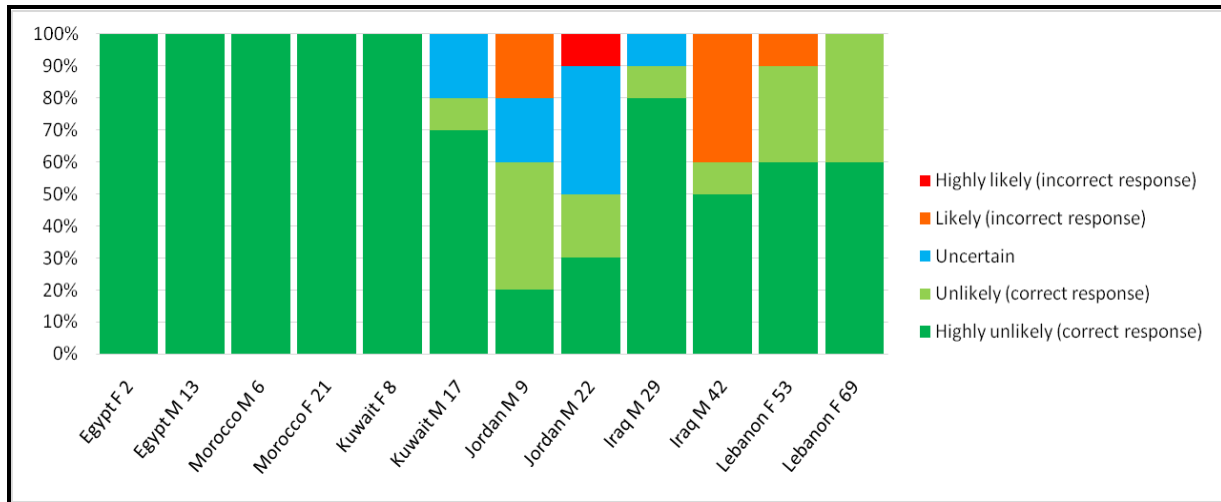
Six of these (the two Egyptians, the two Moroccans, one of the Kuwaitis and one of the Lebanese) attracted 100% correct rejections—unanimously and to the highest available degree of confidence in the case of all but the Lebanese stimulus. Nonetheless, the overall pattern of responses to non-Syrian stimuli was far more heterogeneous than to Syrian stimuli (cf. Figure 11.2), with some false acceptances and a lower degree of confidence (including a number of outright 'uncertain' responses) throughout.

Syrian linguists performed with comparable variability to Syrian NSNLs on the two Jordanian stimuli, Jordan M 19 and M 22. Here, correct rejections comprised, respectively, 60% and 50% of total responses. Similarly, too, there was a relatively large number of 'uncertain' responses, especially to Jordan M 22.

Iraq M 42 also occasioned a large proportion of false acceptances—40% of listeners selected a 'likely' response—and Kuwait M 17 a relatively large number of 'uncertain' responses (20%). However, these and the two Jordanian stimuli, correct rejections exceeded 80% in all instances.

One Lebanese stimulus (Lebanon F 69) resulted in 100% correct rejections, while the other (Lebanon F 53) provoked only 10% false acceptances. Even more than in Figure 11.6, this

relatively low error rate was surprising given the manifold similarities between Lebanese and Syrian (Damascene) Arabic.

Figures 11.6 and 11.7 together suggest that, for Syrian listeners of both kinds, there is a non-trivial perceptual difference between the two types of Levantine stimuli, Jordanian and Lebanese. Lebanese stimuli appear easier to reject than Jordanian stimuli, the latter of which also result in a great many more 'uncertain' responses than do the former.

Non-Syrian NSNLs

Figure 11.8 shows non-Syrian NSNLs' (N = 21) response patterns to the 12 non-Syrian stimuli.

Figure 11.8: Non-Syrian NSNLs' responses to non-Syrian stimuli
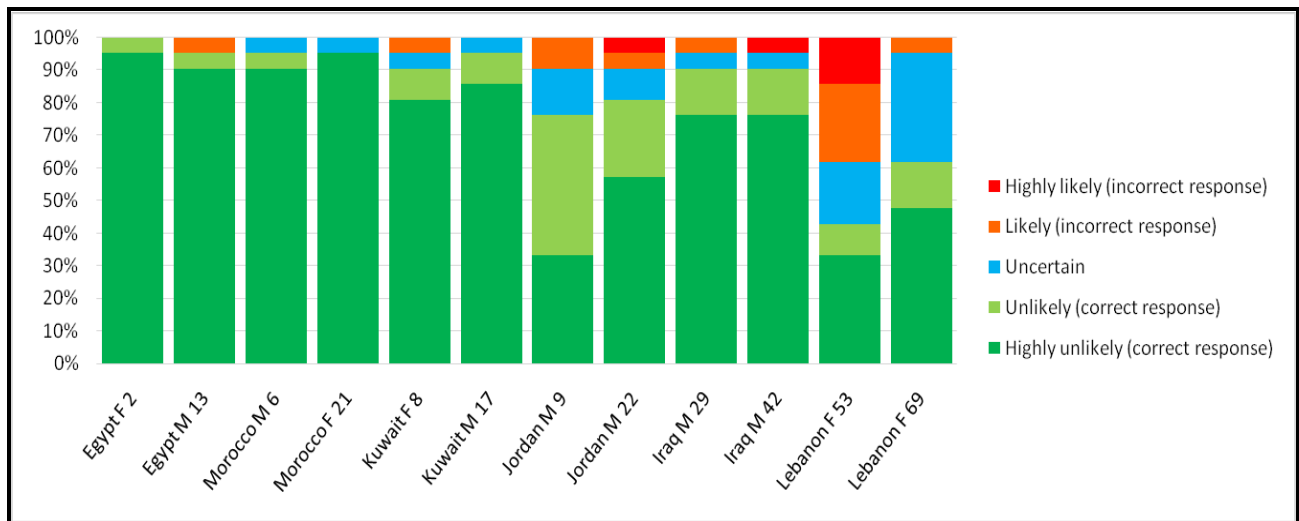
Only one stimulus, an Egyptian, attracted 100% correct rejections. It is immediately noticeable that the heterogeneous pattern evinced in this group's responses is not dissimilar to that shown in their responses to Syrian stimuli, as recorded in Figure 11.3. Here, though, accuracy is greater: 90% or higher on all but Levantine (i.e. Jordanian and Lebanese) stimuli.

Even on Jordanian stimuli, accuracy is high: c. 75% on Jordan M 9 and c. 80% on Jordan M 22, this being a considerably more accurate performance than that of either Syrian group. On Lebanese stimuli, however, accuracy is a good deal lower—c. 40% on Lebanon F 53 and c. 60% on Lebanon F 69. Each is in turn much lower than the percentage of correct rejections attained by Syrian listeners on the same pair of stimuli.

As Figures 11.6 and 11.7 indicate for Syrian listeners, there seems to be a substantial perceptual difference for non-Syrian NSNLs between two types of Levantine stimuli, Jordanian and Lebanese. Here, though, the pattern is reversed: for non-Syrian NSNLs, Jordanian stimuli are easier to reject than Lebanese stimuli, the latter of which also result in a great many more 'uncertain' responses than do the former.

Non-Syrian linguists

Figure 11.9 shows non-Syrian linguists' (N = 22) response patterns to the 12 non-Syrian stimuli.

**Figure 11.9**: Non-Syrian linguists' responses to non-Syrian stimuli
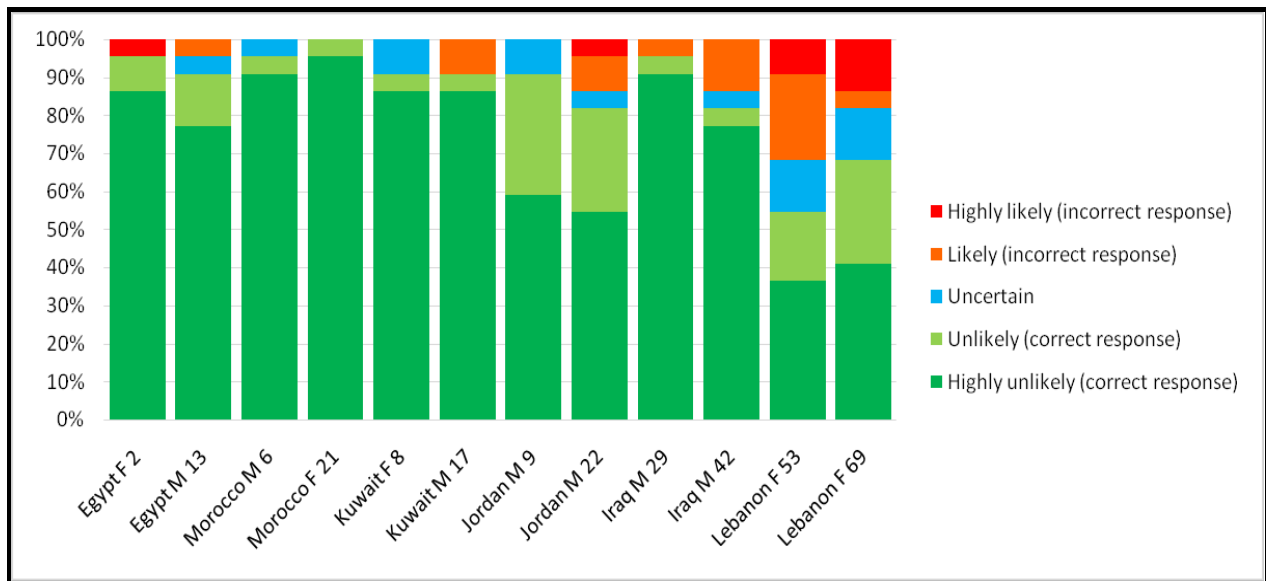
Only one stimulus, a Moroccan, attracted 100% correct rejections. Of note is that the heterogeneous pattern evinced in this group's responses is similar to that shown in their responses to Syrian stimuli, recorded in Figure 11.4.

However, accuracy here is greater: above 80%—and in most cases above 90%—on all except Lebanese stimuli. Accuracy on Lebanese stimuli is higher than that achieved by the non-Syrian NSNLs in Figure 8: c. 55% on Lebanon F 53 and c. 70% on Lebanon F 69, though each is still a lower accuracy ratio than that achieved by either Syrian listener group on the same pair of stimuli.

Again, as Figures 11.6 and 11.7 demonstrate for Syrian listeners and Figure 11.8 for non-Syrian NSNLs, there appears to be a considerable perceptual distance for non-Syrian linguists between the two types of Levantine stimuli, Jordanian and Lebanese. The pattern here clusters with that of non-Syrian NSNLs: for non-Syrian linguists too, Jordanian stimuli are easier to reject than Lebanese stimuli, the latter of which likewise result in more 'uncertain' responses than do the former.

Non-native linguists

Figure 11.10 shows Non-native linguists' (N = 5) response patterns to the 12 non-Syrian stimuli.

Figure 11.10: Non-native linguists' responses to non-Syrian stimuli
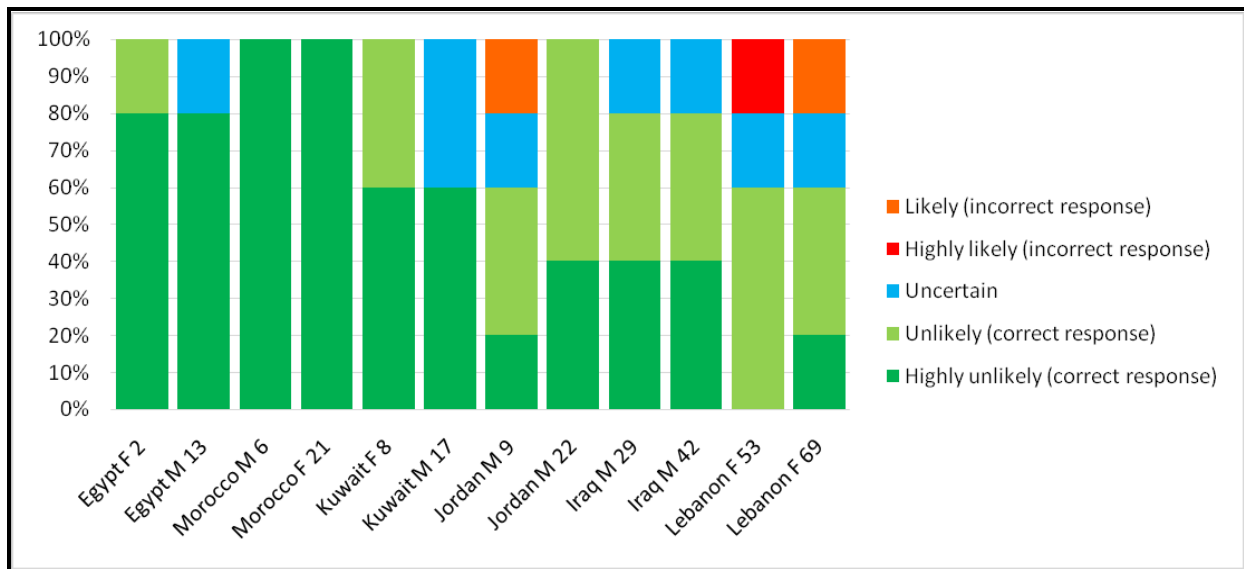
Five stimuli (one of the Egyptian, both of the Moroccan, one of the Kuwaiti and one of the Jordanian) attracted 100% correct rejections. Unlike the patterns shown by Syrian listeners, but similar to those of non-Syrians, non-native linguists evidence a broadly heterogeneous pattern of responses both to Syrian and to non-Syrian stimuli.

However, the overall accuracy of this group is marginally less consistent on Syrian than on non-Syrian stimuli (cf. Figure 5). While five non-Syrian stimuli were correctly rejected by 100% of listeners—and the two Iraqi stimuli each attracted 80% correct rejections, with no false acceptances—both of the Kuwaiti, both of the Lebanese and one of the Jordanian stimuli here

each occasioned only 60% correct rejections. Moreover, the latter three (Levantine) stimuli resulted in a mixture of 'uncertain' responses and false acceptances.

11.5.2.3. Overall group response patterns

This subsection presents and analyses overall group response patterns via two summary figures. The first (Figure 11.11) compares the five groups' response patterns to Syrian versus non-Syrian stimuli. The second (Figure 11.12) is a comparison of each group's combined percentage of correct, incorrect and uncertain responses to all stimuli.

Figure 11.11: All groups' responses to Syrian vs. non-Syrian stimuli
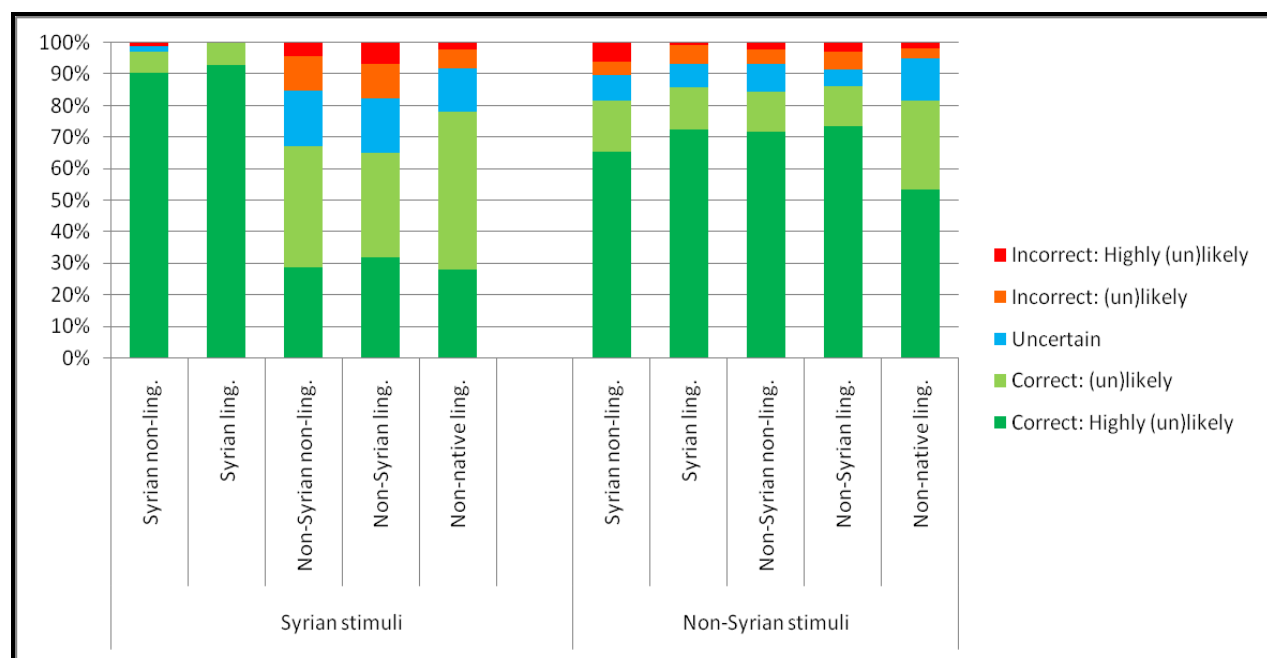
The most striking observation to be drawn from Figure 11.11 is that, on Syrian stimuli, the patterns evinced by Syrian listeners generally cluster together in sharp contrast to those of non-Syrians. Syrian linguists achieved an accuracy rate of 100% and NSNLs 97%, each with c. 90% correct acceptances to the highest degree of confidence. Syrian NSNLs entered c. 2% 'uncertain' responses and Syrian linguists none at all.

Non-native linguists, too, achieved an impressive degree of accuracy, at just below 80%. Non-

295

Syrian native speakers evidence lower accuracy (c. 65%) than non-natives. Non-natives' percentage of 'highly likely' responses (c. 30%) is much lower than that of Syrians but comparable to those of the two non-Syrian native speaker groups. Both non-Syrian native speaker groups entered a roughly similar percentage of 'uncertain' responses (c. 20%), and non-natives a slightly lower percentage (c. 15%).

On non-Syrian stimuli, all five groups evince broadly similar patterns, including an accuracy rate of between 80 and 85%. The principal difference is that non-native linguists demonstrate considerably lower confidence in their correct rejections: c. 30% 'unlikely' against c. 50% 'highly unlikely', compared to, e.g., Syrian/non-Syrian linguists' c. 10% 'unlikely' against c. 70% 'highly unlikely'.

Figure 11.12: All groups' total percentage of correct, incorrect and uncertain responses
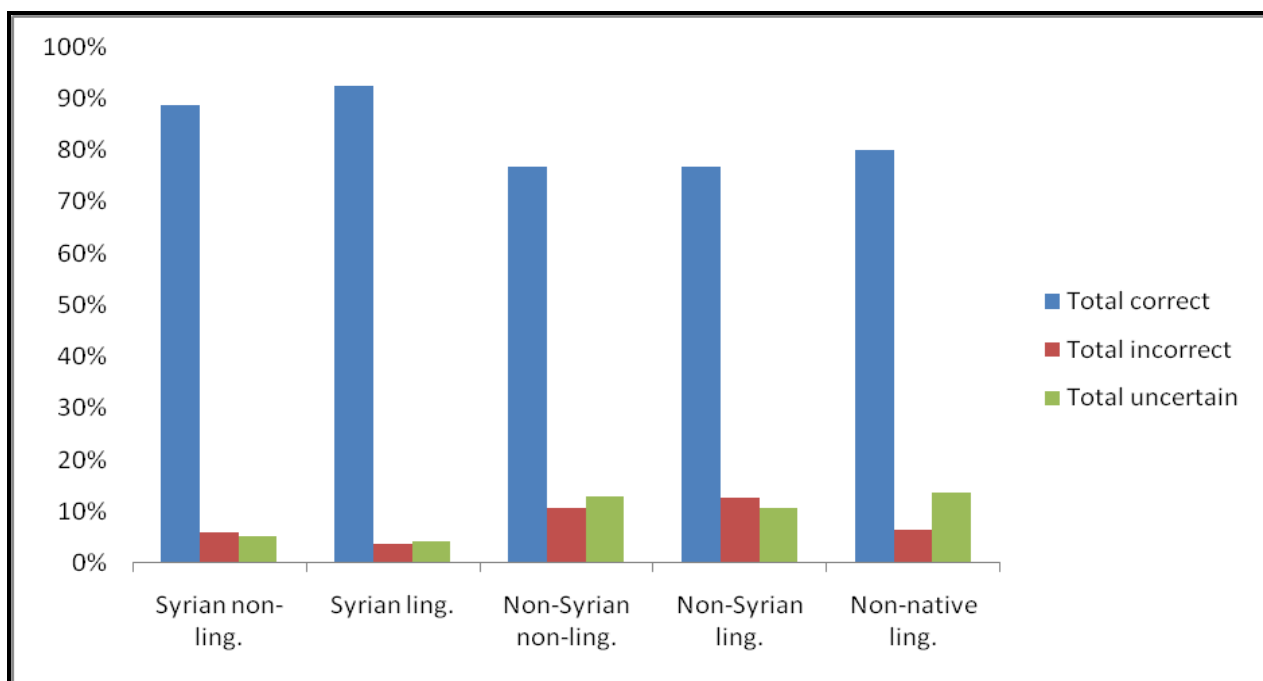
In Figure 11.12, both Syrian groups again cluster more similarly to each other—in correct, incorrect and uncertain responses—than to any of the other groups. Each entered c. 90% correct, c. 5% incorrect and c. 5% uncertain responses. Non-Syrians similarly cluster together (each c. 75% correct, with the balance of responses varying slightly by type between the two groups).

Non-natives pattern between these two clusters on correct responses (80%) but are more similar to

Syrians than to non-Syrians on incorrect responses (c. 5%). They entered marginally more 'uncertains' (c. 15%) than non-Syrian linguists but roughly the same percentage as non-Syrian NSNLs.

Three important overall conclusions, reinforcing those drawn from Figure 11.11, can be reached on the basis of Figure 11.12. The first is that Syrians cluster together in evidencing higher overall accuracy and confidence, as well as higher accuracy, than the three other groups, which in turn cluster together, to varying degrees, on the former two criteria. The second is that non-Syrians, like Syrians, cluster very closely on all three of the above criteria. The third is that non-natives cluster with Syrians rather than non-Syrians on 'incorrectness'.

These results seem to replicate those of Wilson 2009: local NSNLs are both more accurate and more confident than other groups (with the exception of local linguists, to which Wilson did not include an experimental group of analogous type).

11.5.2.4. Discussion

Preliminary to inferential statistical analysis, four main observations of interest emerge from the preceding review of descriptive statistics. The first is that hypothesis (i) appears to be supported: On Syrian stimuli, the Syrian NSNL and linguist groups—in that order—were each more accurate than any of the other three groups, linguist or NSNL. This finding is fully consistent with the descriptive statistical survey of results from the two prior experiments in the Arabic series, in which Syrians were likewise more accurate than were non-Syrians on Syrian stimuli. It also accords with the descriptive statistics examined in connection with the results of Eng-2 (Chapter 8).

The second point of interest is that descriptive statistics provide *some* preliminary support for hypothesis (ii)—that is, Syrians will be more accurate than other listeners in their responses to Syrian stimuli, *independent of education*. Syrian linguists, however, are *marginally* more accurate than Syrian NSNLs, which implies a possible effect for education. Owing to the types of listeners recruited, the response patterns of linguists and NSNLs could not be compared in any

of the previous experiments, except in Eng-2. There, inferential statistical analysis showed education to have a positive effect on accuracy, independent of (sub)national origin.

Third, going beyond the thesis' guiding hypotheses to a discussion of confidence (i.e. 'decision' vs. 'uncertain' responses), it is notable that all three non-Syrian groups, in addition to their lesser accuracy on Syrian stimuli, show a greater tendency than Syrians towards uncertain responses to *all* stimuli; as Figure 11.11 indicates, this is particularly the case on Syrian stimuli. In short, non-Syrians present as less confident than Syrians.

In Ar-1 and Ar-2, descriptive statistics suggested that Syrian listeners were generally more confident than non-Syrians on Syrian stimuli only—a finding consistent with the descriptive statistics reviewed above. In partial contrast, inferential statistical analysis of the results of Eng-2 demonstrated a significant effect on confidence for the identity of the speaker (i.e. Yorkshire origin *only*), among *all* listener groups.

At any rate, the above analysis of the results of Ar-3 would seem to refute the forceful and repeated contention of Patrick (2010) and Fraser (2009, 2019) that education is the determinative factor in (lack of) confidence. Rather, the inference preliminarily made here is that among Arabic speakers 'localness' is decisive: local listeners, regardless of education, are both more confident and more accurate on local stimuli than non-locals.

Fourth, as Table 11.3 indicates below, listener groups varied considerably in their respective percentages of false rejections and false acceptances. As was the case with apparent differences in accuracy and confidence, this distinction seems to depend on the listener's 'localness'. Syrians of both groups entered a negligible percentage of false rejections and, relative to this, a large number of false acceptances. Non-Syrians of all groups tended in the opposite direction: a higher percentage of false rejections than false acceptances. The difference between Syrian and non-Syrian groups on false acceptances, however, is much smaller (cf. Syrian linguists 7%, non-Syrian linguists 8%) than it is on false rejections (cf. Syrian NSNLs 1%, non-Syrian NSNLs 15%).

Table 11.3: percentage of false rejections and false acceptances by listener group

Listener group	False rejections	False acceptances
Syrian NSNLs (N = 21)	1%	10%
Syrian lings (N = 10)	0%	7%
Non-Syrian NSNLs (N = 21)	15%	7%
Non-Syrian ling (N = 22)	18%	8%
Non-native lings (N = 5)	8%	5%

The pattern shown in Table 11.3 is consistent with those observed in both prior experiments in the Arabic series, in which local listeners, linguists or not, also entered higher proportions of false acceptances than false rejections. This pattern runs counter to Wilson 2009 in the sense that, in her experiment, listeners of *all* types entered more false acceptances than false rejections (this was also the case in Eng-1). In the English series, however, all listener groups (except Non-Brit in Eng-2) evidenced the same broad error pattern as that observed by Wilson: a greater proportion of false acceptances than false rejections.

The numbers in Table 11.3 may be interpreted as consequential for LAAP as presently practiced, especially in the light of statistics quoted in Chapter 2. There, it was revealed that, over a 12-month period, 94.5% of applicant claims to Syrian origin were confirmed in reports by the primary-phase LAAP agency Verified, which employs the NSNL + linguist team approach. By contrast, over three successive years, a much smaller percentage of claims were confirmed by LINGUA, which does not employ NSNLs. I also showed that, in Syrian cases, Verified employs NSNLs whose native dialect matches that claimed by the applicant (such dialect-matching applies in all claims examined by the agency).

In Chapter 3 I demonstrated that, in the initial verification task, Verified evaluates the evidence in favour of/against a single hypothesis—i.e. that the applicant's account is true. The strength of this evidence is not, however, weighed against an alternative hypothesis assessing the strength of the evidence that the applicant's account is false. In Chapter 3 I also indicated that, if due account is not taken of the *distinctiveness* of particular features to the language variety at issue, the likelihood that the observations substantiate the applicant's story cannot be properly

assessed. Finally I showed that testing of the evidence for only one hypothesis is thus particularly liable to result in false acceptances.

Given all this, combined with the figures in Table 11.3 (cf. Syrian NSNLs with 10% false acceptances of Syrian stimuli), it is at least possible that some of the 94.5% confirmatory conclusions by Verified reported in Chapter 2 were founded on false acceptances.

However, it must once again be recognised that the data here emerge from listener responses to tightly controlled speech material—i.e. read texts. The extent to which the data can be considered applicable to current LAAP procedure is arguable, considering that a quite different kind of speech material, unstandardised and relatively uncontrolled, results from LAAP interviews.

This objection duly acknowledged, three observations about current team-approach LAAP practice can be made, or re-stated, here. The first is that enhanced cross-checking of initial conclusions—perhaps further to the types already conducted by Verified and the OCILA (cf. Chapter 2)—seems likely to strengthen their reliability. For instance, a claim to Syrian background specifying an origin in the Syrian part of the Horan region, on the border with Jordan, could first be analysed by an analyst from the Syrian Horan and afterwards checked by an analyst from the Jordanian Horan. This procedure may assist in reducing the number of false acceptances by local NSNL analysts, again assuming that the tendencies observed in Table 11.3 are similarly present in real-world LAAP. Second, it is worth re-emphasising once more the potential value of testing the evidence for two hypotheses—and generating likelihood ratios measuring the strength of the evidence for each—in reducing the revealed propensity of local NSNLs towards false acceptances. Third, it may be useful for agencies to test the reliability of their conclusions by means of the supplementary tests whose possible form is outlined in Chapter 12.

In the following subsection I investigate the potential validity of the envisaged perception test. I do so via an inferential statistical analysis of the data, which were examined in the preceding subsection only through the prism of descriptive statistics.

11.5.3. Inferential statistics

11.5.3.1. Procedure

In Ar-3, as in Eng-2, inferential statistical analysis was conducted in *R* (R Core Team 2022), by means of a series of binomial logistic mixed effects models (a form of glmer), built with the lme4 package. Fixed effects were speaker geographical background (two levels: Syrian or non-Syrian), listener geographical/language background (three levels: Syrian native speaker, non-Syrian native speaker, USA non-native speaker) and listener education (two levels: linguistic education or no linguistic education). Participant was included as a random effect. Paired interactions between all three fixed effects were also incorporated into the full model.

In testing the significance of each variable in the full model, a partial model was composed for each variable at issue. These partial models were the same as the full model except that in each instance the variable in question was excluded. Model comparisons were conducted by means of ANOVAs.

Listener responses to the question 'Is this a Syrian accent?' constituted the dependent variable. Two series of dependent variables (DVs) were analysed: one assessing accuracy of response, and the other assessing confidence. Accuracy was modelled as '0' or '1' according to an incorrect or correct response (regardless of degree of confidence as expressed by Likert-scale ratings). Confidence was also modelled as '0' or '1', according to whether the response was an 'uncertain' or 'decision' response, irrespective of accuracy or degree of confidence.

As also in Eng-2, outright 'uncertain' responses (i.e. point three on the relevant Likert scale) were excluded from the analysis of accuracy. This decision was made because of the difficulty in deciding whether listeners' selection of this response type constitutes a simple admission of 'don't know' (i.e. lack competence/familiarity) or caution in identifying the stimulus as Syrian/non-Syrian due to familiarity/competence. The exclusion of 'uncertain' responses means that the y-axes in Figures 11.13 and 11.14 display a higher probability of accuracy than would have resulted from their inclusion.

Each series of DVs was analysed in a separate series of model comparisons. Accuracy is dealt with in subsections 11.5.3.2 and 11.5.3.3, followed by confidence in subsections 11.5.3.4. and 11.5.3.5.

11.5.3.2. Listener accuracy: results

Table 11.4 presents the coefficients of the model of best fit and significance values, as determined by glmer, for accuracy of listener response. It also specifies significance values, where present, for the fixed effects. The intercept encodes two baseline fixed effects: 'Background Non-Syrian [native speaker]' and 'speaker Non-Syrian'. The fixed effect of education did not feature in the model of best fit.

Table 11.4: Coefficients and significance values for glmer analysis of accuracy

```
Fixed effects:
                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                               2.8129     0.2459  11.437  < 2e-16
SpeakerSyrian                            -1.1971     0.2298  -5.209 1.90e-07
BackgroundNon-Syrian (USA)                0.2805     0.8082   0.347    0.728
BackgroundSyrian                         -0.2313     0.3580  -0.646    0.518
SpeakerSyrian:BackgroundNon-Syrian (USA)  0.4859     0.8630   0.563    0.573
SpeakerSyrian:BackgroundSyrian            4.1060     0.7742   5.303 1.14e-07

(Intercept)                               ***
SpeakerSyrian                             ***
BackgroundNon-Syrian (USA)
BackgroundSyrian
SpeakerSyrian:BackgroundNon-Syrian (USA)
SpeakerSyrian:BackgroundSyrian            ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results demonstrate significant effects (by means of predicted probabilities) on accuracy for one independent variable and one pair of independent variables in interaction. The first is speaker origin (i.e. whether the stimulus was Syrian or not). The second specifies an interaction between speaker origin and listener background (i.e. the listener's association with Syria by

citizenship).

Figures 11.13 and 11.14 illustrate the predicted probability of correct response (i.e. accuracy) associated in turn with speaker origin (i.e. whether the stimulus featured a Syrian or non-Syrian speaker) and the interaction between the background of the speaker in the stimulus and that of the listener.

Figure 11.13: Predicted probability of correct response associated with, respectively, non-Syrian and Syrian stimuli

Figure 11.14: Predicted probability of correct response associated with interaction between listener and stimulus type
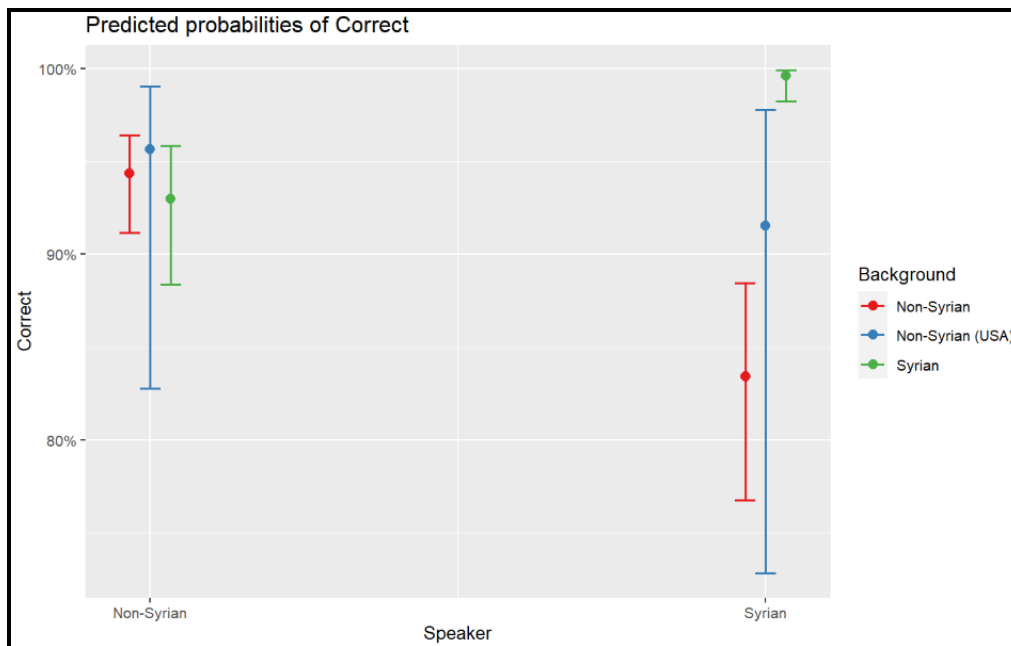
11.5.3.3. Listener accuracy: discussion

As observed in Eng-2, results indicate support for hypothesis (i): local listeners will be most accurate in identifying local speakers. Syrians were significantly more accurate than other groups on Syrian stimuli (p = <.001). Phrased another way, this finding provides support additional to that adduced in Eng-2 for the other-accent effect: speakers find it easier to recognise their own accent. However, while in Eng-2 the least local listeners (of the Non-Brit type) were significantly less accurate than all three British groups, no comparable effect was found here.

Additionally confirming patterns observed in the results of Eng-2, Figure 11.14 demonstrates a broadly negative association between variability in group response patterns to Syrian stimuli and accuracy. Once more, the narrowest range was found among the most accurate group (Syrians). However, unlike in Eng-1, the same trend is not observable among the two less local groups; the least local group here (i.e. non-natives) is the second most accurate but also evidences the greatest variability in its response pattern to Syrian stimuli. This may be

attributable to the small number of non-native listeners (N = 5) in Ar-3.

In a further parallel to the results of Eng-2, there appears to be a *positive* correspondence between accuracy and variability in response patterns to non-Syrian stimuli. The non-native group trends highest on both accuracy and variability, while the Syrian group trends lowest on each. Again, this might have to do with the small number of non-native listeners (N = 5) in Ar-3.

Also, whereas in Eng-2 a significant advantage in accuracy was associated with the fixed effect of education, here it was entirely dropped from the model of best fit in the process of preceding model comparisons. Hypothesis (ii), that local listeners will be more accurate in identifying local speakers irrespective of linguistic education, thus appears to be supported by the results of Ar-3.

Finally, in a reversal of the trend shown in the results of Eng-2, listeners *in general* here were significantly less accurate on Syrian than on non-Syrian stimuli (p = <.001). This result arose despite the fact that Syrians were significantly more accurate on Syrian than on non-Syrian stimuli. This may be taken to illustrate the sheer size of the effect of 'localness' on accurately identifying speakers.

Figure 11.14 shows that variability in the *overall* response pattern to Syrian stimuli was greater than that observable in the case of non-Syrian stimuli. As Figure 11.13 demonstrates, however, Syrian listeners evince extremely low variability in the accuracy of their responses to Syrian stimuli. Most of the variability can thus be accounted for by the response patterns of the two non-Syrian groups.

11.5.3.4. Listener confidence: results

As specified in 11.5.3.1., confidence is operationalised here in binary fashion. The dependent variables are 'uncertain' responses (coded as '0') and 'decision' responses (coded as '1'), regardless of degree of confidence (highly likely, unlikely etc.). The fixed effects are the same as they were in the analysis of accuracy: Syrian vs. non-Syrian stimuli, listener geographical

background (Syrian vs. Syrian) and listener linguistic education.

Table 11.5 shows the coefficients of the model of best fit and significance values, as determined by glmer, for confidence of listener response. As in the above analysis of accuracy, the intercept shown in Table 11.5 encodes two baseline fixed effects: 'Speaker Non-Syrian', and 'Background Non-Syrian [native speaker]'.

Table 11.5: Coefficients and significance values for glmer analysis of confidence

```
Fixed effects:
                                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                                  2.87924    0.23628  12.186  < 2e-16
SpeakerSyrian                               -1.10538    0.22190  -4.982 6.31e-07
BackgroundNon-Syrian (USA)                  -0.59643    0.63568  -0.938    0.348
BackgroundSyrian                            -0.07679    0.35045  -0.219    0.827
SpeakerSyrian:BackgroundNon-Syrian (USA)     0.85168    0.63057   1.351    0.177
SpeakerSyrian:BackgroundSyrian               3.03182    0.58878   5.149 2.61e-07

(Intercept)                                 ***
SpeakerSyrian                               ***
BackgroundNon-Syrian (USA)
BackgroundSyrian
SpeakerSyrian:BackgroundNon-Syrian (USA)
SpeakerSyrian:BackgroundSyrian              ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In an almost exact parallel with results concerning accuracy, the table above demonstrates significant effects (by means of predicted probabilities) on confidence for one independent variable and one pair of independent variables in interaction. The first is speaker origin (i.e. whether the stimulus was Syrian or not). The second involves an interaction between speaker origin and listener background (i.e. the listener's association with Syria by citizenship). Once more, education did not occur as a significant fixed effect in the model of best fit; the same is true of the model of best fit for accuracy.

Figures 11.15 and 11.16 illustrate the predicted probability of a 'decision' response (i.e. confidence) associated in turn with speaker origin (i.e. whether the stimulus featured a Syrian or non-Syrian speaker) and the interaction between the background of the speaker in the

stimulus and that of the listener.

Figure 11.15: Predicted probability of 'decision' response associated with, respectively, non-Syrian and Syrian stimuli
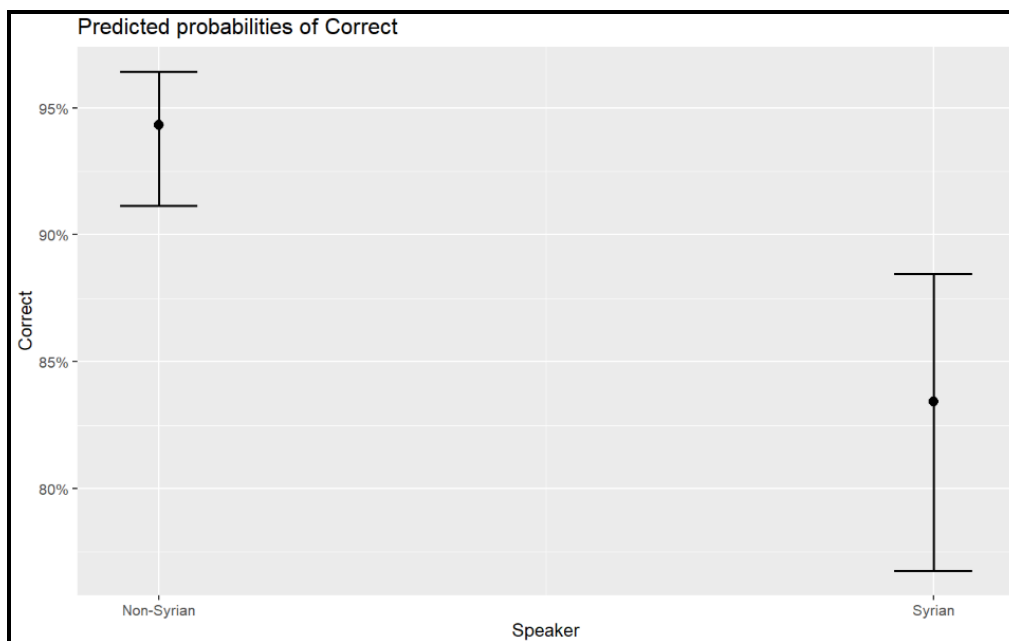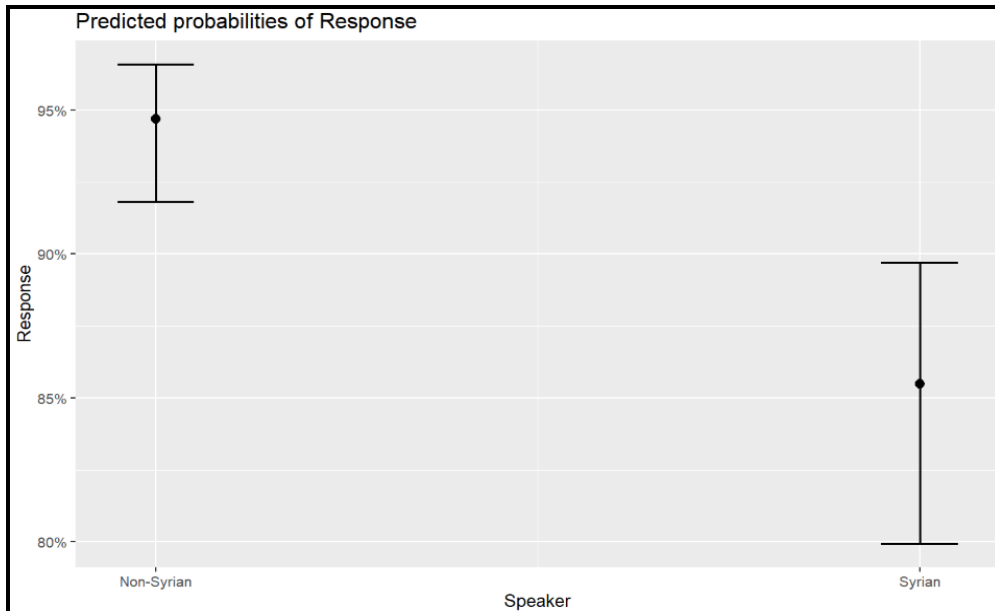

Figure 11.16: Predicted probability of 'decision' response associated with interaction between listener and stimulus type
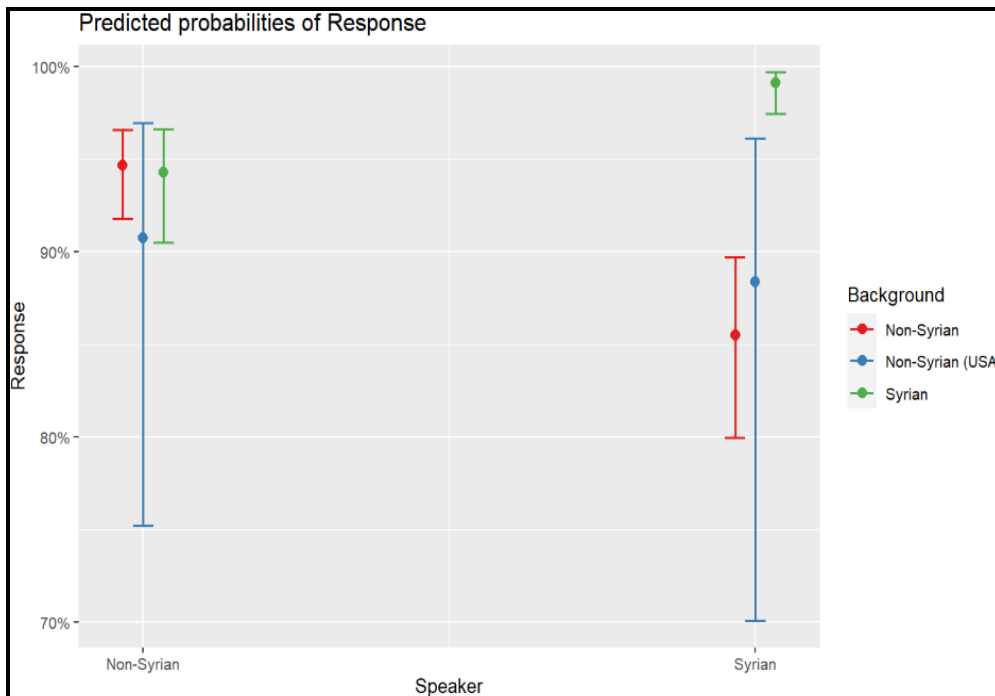
11.5.3.3. Listener confidence: discussion

As in Eng-2, no significant effect of education on listener confidence was located, the variable itself having been eliminated completely from the model of best fit (as was also the case with accuracy). At least in LAAP-like contexts, the replication of these results here casts additional doubt on the validity of claims as to the moderating effect of advanced linguistic training on listener confidence (e.g. in Fraser 2009, 2011, 2019 and Patrick 2010; see Chapter 5).

In a striking commonality with patterns observed in the above results concerning accuracy, Figure 11.16 demonstrates a broadly negative association between variability and confidence in group response patterns to Syrian stimuli. The narrowest range of responses was again found among the most confident group (Syrians). Once more, however, the two less local groups do not conform to this trend: the least local group (i.e. non-natives) was the second most confident on Syrian stimuli but also evidenced the greatest variability in its response pattern. As before, this can probably be attributed to the small number of non-native listeners (N = 5) in Ar-3.

Additionally mirroring patterns observed in the results involving accuracy, Figure 11.16 demonstrates an (even stronger) *positive* association between variability and confidence in group response patterns to non-Syrian stimuli. The broadest range was found among the most accurate group (non-Syrian native speakers) and the narrowest among the least accurate (non-Syrian non-natives). As with accuracy, the wide range observable in non-natives' confidence could well be due to the small number of non-native listeners (N = 5) in Ar-3.

Finally, as is again evident in the results returned for accuracy, a significant effect on confidence of response was located in the interaction between Syrian speakers and Syrian listeners—but only in responses to Syrian stimuli.

11.5.4. Correlation between accuracy and confidence: linguists vs. non-linguists

11.5.4.1 Procedure

I now turn to formally test the correlation between accuracy and confidence among non-linguists and linguists respectively. I do so in order to further investigate the explicit claim that, among non-linguists, accuracy and confidence are poorly correlated (cf. Fraser 2009, Patrick 2010).

Two separate tests were conducted. The first considered the responses of all NSNLs (Syrian and non-Syrian, N = 42) and the second, as a control, the responses of all linguists (Syrian, non-Syrian and non-native, N = 37)

Previously in Ar-3 (and in Eng-2), inferential statistical analyses of confidence were conducted only on a binary basis: 'decision' vs. unsure responses. Here, to permit consideration of *degrees* of confidence and accuracy, each point on the Likert scale was assigned a number in the 'confidence' column. All 'highly likely'/'highly unlikely' responses were coded as '3', with all 'likely'/'unlikely' responses coded as '1', and all 'uncertain' responses as '0'. For accuracy, all correct responses were coded as '1', and all incorrect and uncertain responses were coded as '0'. Responses by each listener to each of the 22 Syrian stimuli were added up and a mean score calculated for both variables, for each listener.

Analysis of the correlation between accuracy and confidence was performed in JASP (JASP Team 2022) via Pearson's r test. This particular test was selected because it is commonly used to examine correlations between the types of raw, continuous data relevant here.

11.5.4.2. Results: NSNLs

Table 11.6 shows the results for Pearson's r test of the correlation between accuracy and confidence in responses by all NSNLs (N = 42, df = 40). Figure 11.17 shows the associated scatter plot and line of best fit.

Results show a strong positive correlation between confidence and accuracy in the responses of NSNLs (r = 0.657), as well as the statistical significance of the result (p = <.001).

Table 11.6: results of Pearson's r test of correlation between accuracy and confidence in NSNLs' responses

			Pearson's r	p
Confidence	-	Accuracy	0.657	< .001

Figure 11.17: scatter plot and line of best fit for NSNLs' responses
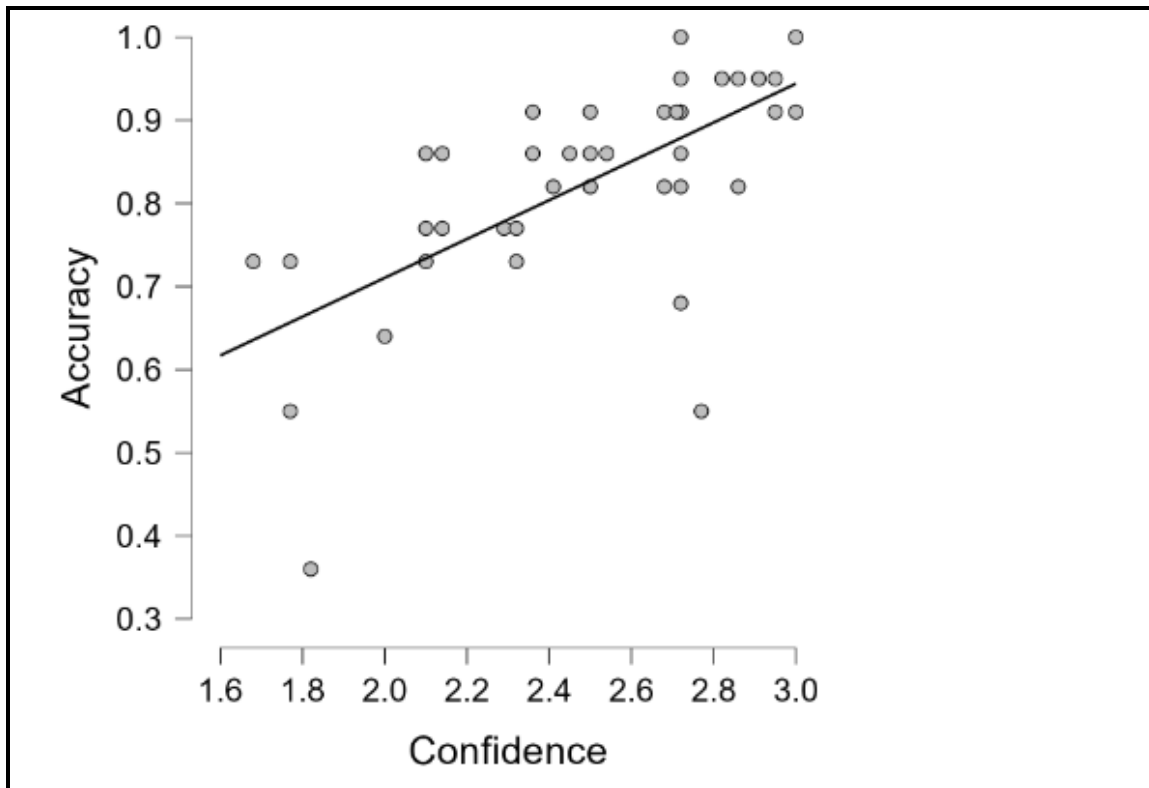
11.5.4.3. Results: linguists

Table 11.7 shows the results for Pearson's r test of the correlation between accuracy and confidence in responses by all linguists (N = 37, df = 35). Figure 11.18 shows the associated scatter plot and line of best fit.

Results show a still stronger positive correlation between confidence and accuracy (r = 0.699) than that returned for the responses of NSNLs, in addition to a similar degree of statistical significance (p = <.001).

Table 11.7: results of Pearson's r test of correlation between accuracy and confidence in linguists' responses

	Pearson's r	p
Confidence - Accuracy	0.699	< .001

Figure 11.18: scatter plot and line of best fit for linguists' responses
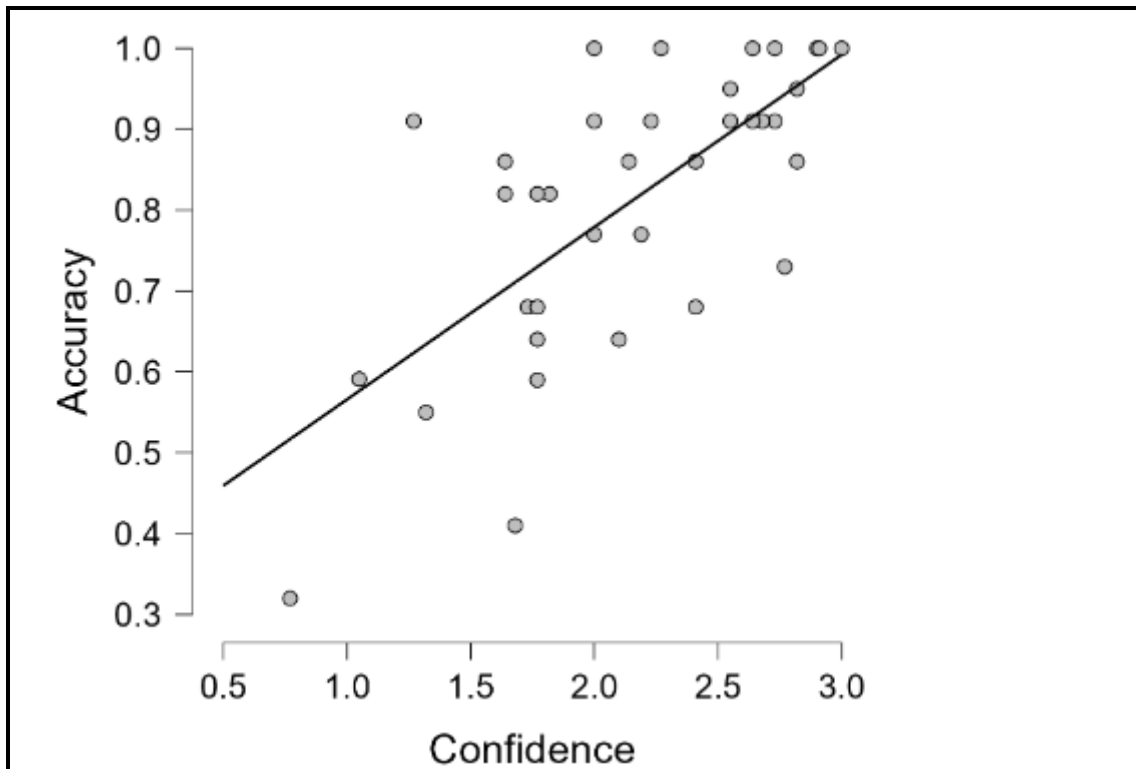
11.5.4.4. Discussion

These results demonstrate that confidence and accuracy are strongly and significantly correlated in the responses of NSNLs, and still more so among linguists, in Ar-3. The former finding stands in opposition to claims advanced in the LAAP literature that NSNLs evidence a poor correlation between the two variables in their responses to LAAP-like tasks. The strong correlation between the variables in the responses of both listener types is interesting, considering that no correlation can be inferred in the model of best fit for either accuracy or confidence, the variable of education being absent from both models (cf. subsection 11.5.3)

The finding that accuracy and confidence are strongly and significantly correlated in the responses of both listener groups furthers the work in this area done by Wilson (2009). However, it contradicts Wilson's finding of no significant correlation of the two variables in the responses of any of her listener groups (cf. Chapter 5).

For the current practice of LAAP, the inference here might be that *both* NSNLs and linguists are reliable judges of language, in the sense that their degree of confidence is closely reflected in the accuracy of their decisions. However, as discussed in 11.5.2.4, the caveat must be entered that the stimuli used in Ar-3 consisted of read texts elicited under closely controlled conditions, whereas in LAAP the speech material assessed is derived from the much less structured speech data occurring in interviews. Furthermore (as covered in section 9.1, Chapter 9), the characteristics of the target variety, Damascene Arabic, are relatively well-known in the Arab world, and contrasts between it and some of the foil varieties—Egyptian and Moroccan—were in general readily apparent to listeners in Ar-3 (cf. the negligible percentage of false acceptances of these two varieties observed in 11.5.2.2). Another consideration is that in LAAP the most problematic cases often involve cross-border or poorly described varieties (Cambier-Langeveld 2018a). In such instances, confidence and accuracy are by no means guaranteed to be as closely correlated as they were in the above analysis.

11.5.5. Listeners' awareness of Syrian dialectology

11.5.5.1. Results

Listeners in Ar-1 and Ar-2, especially those of Syrian origin, demonstrated in their comments strong awareness of the existence of Syrian dialectology. In Ar-3, knowledge of this kind is again reflected in listener comments, and once more they demonstrate cognisance of both the close relatedness of the Levantine dialects and the presence in Syria of non-Levantine varieties.

The large number of comments of the type under discussion precludes their exhaustive reproduction here. As in previous experiments, they have two sources: first, in connection with judgements of individual stimuli (in the free-text boxes associated with each stimulus) and second the free-text box at the conclusion of the survey.

It is notable that the Syrians in Ar-3 were more inclined than non-Syrians to commentary of this broad type. Moreover, as measured by the proportion of individual listeners from each group who left comments, there was less intragroup variability among Syrians, particularly in the case of Syrian linguists), than among non-Syrians.

The Syrian NSNL listener group (N = 21) and the Syrian linguist listener group (N = 10) each entered a total of 16 comments referring to Syrian dialectology. But these comments were distributed more evenly among individuals in the Syrian linguist group: 7 of 10 of linguists, but only 10 of 21 NSNLs, commented in this vein.

Among non-Syrian native speakers, the total number of comments and the number per individual listener were each considerably lower. In the non-Syrian linguist group (N = 22), six comments were left by two individual listeners. In the non-Syrian NSNL group (N = 21), there were six comments by four individuals. In the non-native group (N = 5), meanwhile, two listeners entered seven comments between them.

A minority of these comments were made in response to the general invitation to comment on the tasks in general at the conclusion of the survey, and do not relate to any particular stimulus. There were two comments of this type by Syrian linguists and three by Syrian NSNLs. Two of six comments by non-Syrian linguists and three of six by non-Syrian NSNLs (but none of seven from

non-natives) were of the same kind. The full range of these comments, some of which were originally written in English and others translated from the Arabic, are reproduced below.

Syrian linguists

- *"Some of the dialects I identified as Syrian could also be close to Jordanian and Palestinian Arabic."*

- *"Some areas of Syria may have similar dialects, for example, the dialect of Deir Ezzor and Quneitra may resemble the Iraqi accent a little, and the dialect of Banias, Tartous and the coastal villages of Syria may be similar to the dialects of Tripoli in Lebanon, for example."*

- *"Residents of some areas in Syria speak dialects of neighboring countries, such as the people of the Syrian desert who speak the dialect of the people of Iraq and the people of Houran with a dialect of some of them very close to Jordan."*

Syrian NSNLs

- *"I found it difficult to recognise whether the dialect is Syrian (rural dialect) or Jordanian."*

- *"The people of major cities in Syria do not know the dialects of the people of the villages or the Syrian desert, for example. In several audio clips, I was not sure if the speaker was a Jordanian or an Hourani, for example, or if he was an Iraqi or from Al Jazeera."*

- *"Syria is vast. The part close to Lebanon is influenced by the Lebanese dialect, and the part close to Iraq is influenced by the dialect of the people of Iraq, and the same is the case in the part near Jordan. They are close together."*

Non-Syrian linguists

- *"The countries of the people of the Levant have many dialects close to each other, especially Syria, Lebanon and northern Palestine, and many words differ from one country to another. The Egyptians use it a lot, and likewise the people of Palestine who are in the south use it too, while the people of Levant are Syrians and Lebanese, and some areas in Jordan and northern Palestine quote the word 'Good'"*

- *"The Syrian dialect has a great affinity with the dialects of neighboring countries such as Palestine and Jordan. Therefore, in some passages, it seems difficult to know whether this dialect is Syrian or Levantine (Palestine, Jordan in particular)."*

Non-Syrian NSNLs

- *"In most of the dialects, most of them are similar, in Balad al-Sham [i.e. the Levant], but differences in regions and dialects in the Gulf countries are similar, and so is Morocco. One city can speak more than one."*

- *"Most of the dialects revolve around the region of Syria, so it is very difficult to differentiate them."*

- *"There is a convergence between the Levantine dialects, especially the cities that are on the borders of the neighboring cities of another country. There may be some words used close to each other, or even the manner of pronouncing the words."*

It is noteworthy that comments by Syrians evidence detailed knowledge of linguistic variation within Syria, particularly the existence of varieties straddling the borders with Iraq, Jordan and Lebanon. This is consistent with surveys of listener comments in Chapters 9 and 10. Comments by non-Syrians, meanwhile, focus principally on the *similarity* of the Levantine dialects to each other and the consequent difficulty of distinguishing them.

I turn now to consider the remainder of the comments entered by listeners. These were associated with 'decision' responses—both correct and incorrect (to varying degrees of confidence)—as well as with outright uncertain responses to particular stimuli.

Among Syrian listeners, comments on specific stimuli occur primarily in connection with false acceptances of or outright uncertain responses to Jordanian stimuli. They coincide secondarily with false acceptances or outright uncertain judgements of Iraqi/Kuwaiti stimuli.

Among non-Syrian Arabs the majority of comments materialise in connection with false rejections or uncertain judgements of Syrian stimuli, while among non-native linguists a slight

preponderance are associated with correct rejections of Jordanian and Iraqi stimuli. Of further note is the fact that all three comments by Syrians, on Syrian and Lebanese stimuli combined, occurred in association with correct responses. This is not the case among non-Syrians, among whom a more heterogeneous pattern prevails in comments on these two types of stimuli. Table 11.8 displays the number of comments per stimulus type, per listener group.

Table 11.8: Comments per stimulus type, per listener group

listener group/stimulus type	Syrian NSNLs	Syrian linguists	Non-Syrian NSNLs	Non-Syrian linguists	Non-native linguists
Jordanian	6 incorrect 3 uncertain	1 incorrect 3 uncertain 2 correct		1 correct	3 correct
Iraqi	3 uncertain	2 incorrect 1 correct	1 uncertain		2 correct
Kuwaiti		1 incorrect 1 uncertain			
Syrian		1 correct	1 correct 1 uncertain	1 incorrect 1 uncertain 1 correct	
Lebanese	1 correct	1 correct			2 uncertain

Below I reproduce a selection of these comments. Some are translated from the Arabic, and some were written in English. Note here, as above, the especially acute sense displayed by Syrian listeners of the existence of varieties spoken on both sides of the Jordanian and Iraqi border. This is also evident in the comments of non-native linguists, which—as shown in the Table 11.8—were much less likely to be associated with incorrect acceptances.

Syrian linguists

- *"The use of the g sound instead of q could be an indicator that the person is Jordanian or Palestinian. But the speaker could be from Southern Syria."* (false acceptance of a Jordanian stimulus)

- *"This sounds Syrian, but the use of the verb Ya'irfuun with a full uun instead of oo makes me wonder if this is Palestinian/Jordanian effect or Bedouin effect."* (correct acceptance of a Syrian stimulus)

- *"One of the governorates of Syria speaks the same dialect, which is Dara'a."* (outright 'uncertain' response to a Jordanian stimulus)

- *"Most likely Jordanian, the pronunciation of the letters is heavy, and the word "Hadhul" is used by Jordanians, and it may be from Horan as well."* (correct rejection of a Jordanian stimulus)

- *"This is Lebanese dialect due to the pronunciation of mneeH with an ee instead of aa. The ee is Lebanese than Syrian, although in certain villages in Syria, it is used but there will be other accompanying features."* (correct rejection of a Lebanese stimulus)

- *"Possible dialect from Deir Ezzor or Fenitra"* (incorrect acceptance of an Iraqi stimulus)

Syrian NSNLs

- *"It may be from Dara'a, because the dialect is close to the Jordanian dialect, like the use of the vomited qaf."* (incorrect acceptance of a Jordanian stimulus)

- *"In the word village the pronunciation of the letter qaf is not similar to what Syrians generally use, but it may be from one of the villages in northern Syria."* (outright 'uncertain' response to a Jordanian stimulus)

- *"This dialect is used in southern Syria and northern Jordan, especially the Horan region, in the manner of pronouncing the word village, the word hudoul, and others."* (incorrect acceptance of a Jordanian stimulus)

- *"Either Iraqi or from eastern Syria"* (outright 'uncertain' response to an Iraqi stimulus)

- *"The dialect of the people of Dara'a"* (incorrect acceptance of a Jordanian stimulus)

Non-Syrian linguists

- *"She could be Syrian but not from Damascus, likely to be Palestinian."* (incorrect rejection of a Syrian stimulus)

- *"Not the Damascene accent but close to Dara'a accent or an adjacent country but it is a villager accent. Not Syrian."* (correct rejection of a Jordanian stimulus)

- *"There is a taste of Syrian sound but it could be an urban accent from Jordan." (correct acceptance of a Syrian stimulus).*

- *"The dialects of Bilad al-Sham [i.e. the Levant] are very close to each other, especially the Syrian and Lebanese dialects. It is difficult to differentiate their details."* (outright uncertain response to a Syrian stimulus)

Non-Syrian NSNLs

- *"The speaker talked too fast so it was hard to focus on his dialect. However, he did say 'ʔalashyaʔ' but still some Lebanese dialects might say that too."* (outright 'uncertain' response to an Iraqi stimulus)

- *"The dialect is close to the Syrian accent, and the pronunciation of the letter R is closer to the way it is pronounced by some Syrian regions, where the letter R is thinned."* (correct acceptance of a Syrian stimulus)

- *"Shamiyah* [i.e. Levantine] *dialect, but it is not necessarily Syrian."* (outright 'uncertain' response to a Syrian stimulus).

Non-native linguists

- *"From Jordan or a Syrian Bedouin"* (correct rejection of a Jordanian stimulus)

- *"From Jordan or a Syrian Bedouin"* (correct rejection of a Jordanian stimulus)

- *"Kalash Mo Zainin = Iraqi, unless from Syria close to Iraq"* (correct rejection of an Iraqi stimulus).

11.5.5.2. Discussion

In the equivalent sections of Chapters 9 and 10, a similar survey to the above was conducted of the number and content of comments on Syrian dialectology, by Syrian and non-Syrian listeners

respectively. This prompted two associated conclusions. First, Syrians showed greater awareness of Syrian dialectology than did non-Syrians. Second, comments by Syrians were overwhelmingly associated with false acceptances of or uncertain responses to Jordanian/Iraqi/Kuwaiti stimuli, not false rejections of Syrians (of which there were almost none). This fact, allied with Syrians' greater inclination towards these two response types on non-Syrian stimuli demonstrated by descriptive statistics, suggested that inaccuracy and uncertainty among Syrians might be attributed, at least in part, to what they *do* know instead of what they do *not* know. In other words, Syrians' awareness that some varieties occur on each side of the eastern, western and southern borders of their country means that they have greater difficulty than non-Syrians in correctly rejecting Iraqi, Kuwaiti, Lebanese and Jordanian stimuli.

The same interpretation is tendered in the face of the data reviewed in this chapter. Both descriptive and inferential statistics show that Syrians in Ar-3 replicate, on a larger scale, all of the patterns observed in prior experiments. As in Ar-1 and Ar-2, Syrians here evidence a comparatively pronounced inclination towards commentary in association with false acceptances or uncertain responses to Jordanian, Kuwaiti and Iraqi stimuli, illustrated in Table 11.8. This is most cogently interpreted with reference to Figures 11.6 and 11.7, which show that, among non-Syrian stimuli, both Syrian groups were least accurate/confident on those of Jordanian, Iraqi and Kuwaiti origin, in that order. The combination of this pattern of responses with Syrians' ready commentary on these three stimuli types amplifies the credibility of inferences made in the two prior experiments regarding greater awareness among Syrians of dialectological diversity within Syria (several comments on Jordanian stimuli here and elsewhere mention the Syria/Jordan cross-frontier region of the Horan, as well as Dara'a, a city in the same region).

Worthy of further notice is that Syrian and Lebanese stimuli attracted commentary from Syrians only in connection with correct responses, consistent with Syrians' overall relatively accurate/confident performance on the same. In short, the inference is that Syrians have greater difficulty in rejecting Jordanian, Kuwaiti and Iraqi stimuli than they do in distinguishing 'classic' Levantine urban varieties—i.e. Damascene and (Beiruti) Lebanese—from each other.

By way of comparison, Table 11.8 shows that non-Syrian native speakers entered a relatively large proportion of comments in association with false acceptances of and uncertain responses to Syrian stimuli, and non-native linguists did similarly with Lebanese stimuli. This is consistent with the response patterns in Figures 11.3 and 11.4, for non-Syrian native speakers (and to a lesser degree Figure 11.10, for non-native linguists).

It appears, then, that among native speakers there is a perceptual 'split' between Jordanian and Iraqi Arabic on the one hand (with, e.g., their reflexes of MSA /q/ as /g/ or /q/) and 'classic' Syro-Lebanese urban varieties (with, *inter alia*, their general substitution of [ʔ] for Classical Arabic /q/) on the other. This split appears to affect group response patterns in contrasting ways. Syrians, perhaps in part because of their greater awareness of intra-Syrian dialectology, evidence greater difficulty rejecting Jordanian, Kuwaiti and Iraqi speakers, while non-Syrian native speakers incline to greater degree towards falsely accepting Lebanese and falsely rejecting Syrians. Of course, this leaves open the question of why Syrians would be more accurate in distinguishing a Damascene from a Beiruti accent, since at the segmental phonetic level the two are highly similar. This question, and the tentative but hypothesis of a perceptual split, gleaned also from the results of Ar-1 and Ar-2, might be a productive avenue for future research.

Finally, is notable that Syrian linguists entered the same number of comments as Syrian NSNLs but that these were distributed much more widely among individual listeners (7 of 10 of Syrian linguists, but only 10 of 21 Syrian NSNLs, left at least one comment). This suggests that linguistic training enhances, but is not vital to, knowledge of dialectological variability—or at least the ability to express such. Syrian linguists' slightly superior accuracy on non-Syrian stimuli, as displayed in Figure 11.11, may also imply an advantage in accuracy resulting from enhanced knowledge of this type. As with similar evidence observed in the results of Ar-2, such patterns tend to counter persistent asseverations in the LAAP literature that NSNL judgements are crucially informed by an 'ideology of homogeneism', the more plausible inference being that linguistic training may assist in *describing* but not in*knowing* (tacitly and holistically) one's native language variety (cf. Chapter 4).

11.6. Research objective 2: Results

11.6.1. Organisation

As in Ar-1 and Ar-2, research objective 2 involves investigating the validity of hypothesis (iii) of the thesis: analysis of feature citations will show that speaker-listeners vary by region of origin in the number and kinds of cues they recognise.

Examination of the validity of hypothesis (iii) relates in turn to one of the two practical objectives of the thesis: developing a supplementary production test for asylum applicants. If meaningful differences in cue recognition among the five groups can be discerned here, such differences might be used to shape the test. Its possible form, envisaged to be of the listen-repeat kind pioneered by Moosmüller (2011), is outlined in Chapter 12.

As in the equivalent sections in Ar-1 and Ar-2, this section attempts to discern patterns in the number and types of comments made by the five listener groups in Ar-3—Syrians and non-Syrian linguists/NSNLs and USA-based non-native linguists—in answer to the question, 'What clues led you to your judgement of the speaker's accent?' This question represents the operationalisation of research objective 2.

In Ar-1 and Ar-2, research objective 2 was investigated largely by quantitative means—that is, by classifying and enumerating comments in accordance with a salience/diagnostic-value framework. Herein, 'salience-only' denotes citation by listeners of features in association with *any* response type, and 'diagnostic value' signifies citations occurring *only* in connection with correct responses.

Syrian and non-Syrian groups in Ar-1 and Ar-2 were compared on the number of citations they had made. Citations were divided according to whether they had proved either diagnostic or salient-only. By this method it was found that Syrians generally, as well as non-Syrian NSNLs, found relatively few features salient—and virtually none diagnostic—compared to non-Syrian linguists, who located much greater salience and diagnostic value in the range of features surveyed. These 'reference' features were drawn from dialectological descriptions of Syrian and non-Syrian dialects. A negligible number of features (aside from a few lexical items/chunks) outside this range were cited by either Syrians or non-Syrians.

The relatively small number of listeners in both prior Arabic experiments permitted comprehensive verbatim reproduction of comments and their detailed interpretation in line with descriptions of reference features. However, the larger number of respondents in Ar-3 (N = 79) resulted in a much greater volume and range of comments than either of the previous experiments. Owing to limitations of space, individual listener comments cannot be exhaustively catalogued here as they were previously. Comments from Ar-3 are reproduced in full in Appendix D, mostly in the form of Google translations from the Arabic.

However, a comprehensive quantitative comparison of cue recognition among the five groups, in part deploying again the salience/diagnostic-value framework detailed above, is attempted below.

11.6.2. Methodology used in collating and classifying comments

Comments on cues were collated and classified in the following manner. Any comment that referred to any of the phonetics/phonology, morphosyntax, lexicon or intonation present in the stimuli was counted as a citation, whether it occurred in association with a correct ('diagnostic') or an incorrect/outright uncertain ('salient-only') response. If a given stimulus attracted more than one comment on any of these broad linguistic parameters, each was counted separately as a citation and included in the total.

This methodology yielded a total of 791 feature citations, an average of c. 10 per listener. Of this total, 733 were associated with correct and 58 with incorrect or outright uncertain responses. All were classified under broad primary headings specifying feature types (phonetics/phonology/morphosyntax, lexicon and so on) and secondarily under various subheadings, which are described below.

11.6.3. Range of features cited

11.6.3.1. Phonetic/phonological and morphosyntactic features (mentioned in the consulted sources)

As in prior Arabic experiments, the discrete phonetic/phonological/morphosyntactic units mentioned in the dialectological literature *as well as* being present in the stimuli were designated as potential cues (see Chapter 9, Table 9.1 for details).

One additional feature mentioned in the consulted sources was cited, solely by non-native linguists, in comments in Ar-3. At the outset of the present experimental series, however, it had not been identified as a potential cue. It is described by Eid (2012) as a preverbal particle, /ka/ or /ta/, employed in Moroccan koine when expressing the imperfective aspect. This form occurs in line 6 of the transcripts in Appendix C, in the same syntactic position as (and in form apparently cognate with) Syrian/Lebanese/Jordanian/Iraqi/Kuwaiti /raħ/, Omani /ba/, Tunisian /taw:a/ and Egyptian /ha/. As these are widely characterised in the sources as future markers, there was an argument for including the Moroccan morpheme in the same category. Eid's (2012) description of it as an imperfective marker, though, prompted its categorisation as a morphosyntactic unit in its own right.

Tables 11.9 and 11.10 quantify the segmental phonetic/phonological and morphosyntactic features cited by listeners in Ar-3 *and* mentioned in the sources consulted. Table 11.9 specifies the number of diagnostic and Table 11.10 the number of salient-only feature citations per group. In both tables Syrian and non-Syrian NSNLs' response patterns are highlighted in red font, since it seems likely on its face (given that most asylum applicants are NSNLs) that the pattern of citations by these two groups will be most relevant to the development of the envisaged novel production test. I return to a consideration of some of these features in Chapter 12.

Table 11.9: number of diagnostic segmental phonetic/phonological and morphosyntactic feature citations per group (features mentioned in the consulted literature)

	MSA /q/	MSA /dʒ/	MSA /ð/	Adj. neg.	Final imāla	Future marker	Subject 3PL	3MSG	Moroccan /ka/	Totals
Non-native linguists	25	8	1	17	6	8	7	6	2	80
Syrian linguists	0	7	0	8	1	2	4	0	0	22
Non-Syrian linguists	15	18	0	17	1	3	4	6	0	64
Syrian NSNLs	2	8	0	11	0	5	4	2	0	32
Non-Syrian NSNLs	3	8	1	11	0	0	3	1	0	27
Totals	45	49	2	64	8	18	22	15	2	225

Table 11.10: number of salient-only segmental phonetic/phonological and morphosyntactic feature citations per group (features mentioned in the consulted literature)

	MSA /q/	MSA /dʒ/	MSA /ð/	Adj. neg.	Final imāla	Future marker	Subject 3PL	3MSG	Moroccan /ka/	Totals
Non-native linguists	0	1	0	1	1	1	0	0	0	4
Syrian linguists	1	0	0	0	0	0	0	0	0	1
Non-Syrian linguists	0	0	0	0	0	0	0	0	0	0
Syrian NSNLs	4	0	0	1	0	0	0	0	0	5
Non-Syrian NSNLs	0	0	2	2	1	0	0	3	0	8
Totals	5	1	2	4	2	1	0	3	0	18

11.6.3.2. Other phonetic/phonological features (not mentioned in the consulted sources)

There were a number of fuzzy references to "the sound(s)", "the letters", "the pronunciation", "letters of accentuation", "melody", "rhythm" and "rhyme" and others. These were classified as phonetic/phonological cues and further subdivided under the headings of 'unspecified segmentals' or 'unspecified suprasegmentals'.

A range of further, more specific references to suprasegmental phonetic/phonological features were also made. These, too, were difficult to identify with features specified in the consulted literature. Prompted in part by the advice of a Syrian native speaker of Arabic trained to

doctoral level in Arabic linguistics (Shetewi, pers. comm.), three distinct further categories were therefore composed. The three categories were designated as vowel height (i.e. exclusive of final imāla), vowel length and speech rate.

Under the first, vowel height, were classified references to, e.g., "ta'amarbouta", "breaking the baa", "fraction", "alif" (interpreted as /a/), "kasrah", "tilt" and "inclination". None of these references are well translated from the Arabic. However, according to the linguist native-speaker consultant (Shetewi, pers. comm.), all involve substitution of a short, high, front vowel in some Levantine dialects where a lower vowel would be used in other varieties. The second category, vowel length, encompassed listener references to "prolonged", "stretched" and "extended" vowels. The third, speech rate, permitted inclusion of remarks such as "speed", "pace", "slow", "fast" and "drawn out".

The latter two categories are especially subject to elision: the Arabic of Damascus and Beirut— but not that of Amman—are known as exponents of especially long vowels (Shetewi pers. comm.), while, according to Almbark & Hellmuth (2015), vowel length is associated with speech rate in various Levantine dialects.

There is thus a degree of overlap in the above categories: some comments could as legitimately have been assigned to one category as to another. Those referring in various forms to vowel-raising, for example, could have been assigned to the category of final imāla. The decision not to do so was prompted by information from the linguist native-speaker consultant that vowel-raising occurs outside word-final contexts (Shetewi, pers. comm.).

In fact, in classifying comments generally the exercise of judgement was frequently required, and other researchers may have made different decisions in some or all instances. But the general point was to attempt a quantitative classification of the manner in and degree to which listeners of various kinds perceive dialectological cues. Ignoring such distinctions would have detracted from the aims embodied in hypothesis (iii) of the thesis: determining how the revealed dialectological knowledge of locals and non-locals varies, and to what degree. It would also have impoverished the range of data potentially available to inform the development of the envisaged new test of production, discussed in Chapter 12.

Tables 11.11 and 11.12 quantify the range of segmental and suprasegmental phonetic/phonological features not mentioned in the literature but cited by listeners in Ar-3. Table 11.11 specifies the number of diagnostic and Table 11.12 the number of salient-only feature citations per group. Citations by NSNLs are again highlighted in red font.

Table 11.11: number of diagnostic segmental and suprasegmental phonetic/phonological feature citations per group (features not mentioned in the consulted literature)

	Unspec. seg. phon.	Vowel height	Vowel length	Speech rate	Unspec. supraseg. phon.	Totals
Non-native linguists	7	17	5	0	7	36
Syrian linguists	14	10	3	0	0	27
Non-Syrian linguists	20	4	6	0	0	30
Syrian NSNLs	13	11	5	2	2	33
Non-Syrian NSNLs	9	6	0	9	1	25
Totals	63	48	19	11	10	151

Table 11.12: number of salient-only segmental and suprasegmental phonetic/phonological feature citations per group (features not mentioned in the consulted literature)

	Unspec. seg. phon.	Vowel height	Vowel length	Speech rate	Unspec. supraseg. phon.	Totals
Non-native linguists	0	0	0	0	0	0
Syrian linguists	0	0	0	0	0	0
Non-Syrian linguists	1	1	1	1	0	4
Syrian NSNLs	0	1	0	0	0	1
Non-Syrian NSNLs	0	2	1	3	0	6
Totals	1	4	2	4	0	11

11.6.3.3. Lexical features (not mentioned in the consulted sources)

As was the case in Ar-2, some lexical items/chunks were mentioned frequently enough in Ar-3 as individual units—i.e. not in association with other elements in the text—that they, too, were each designated as discrete cues. For example, references to /hado:l/ (*they*) and its cognates *as a word* were classified separately from references to MSA /ð/ and its reflexes.

Taking /hado:l/ as illustrative of the overall rationale, it seemed reasonable to assume that listeners may have taken as a cue some property other than the (non-)occurrence of MSA /ð/ in the word, even though the opposition /ð/ - /d/ is generally accepted as a kind of shibboleth separating, for example, the Arabic of Damascus from that of and Amman. Similarly, the word for *countryside* in Damascene Arabic is /qarji/, and the initial-consonant with Jordanian /garje/ might have been salient for some listeners. Yet it was impossible to be sure that this contrast was the only feature being cited in cases where the listener mentioned the *word*, not the phone. Vowel length, especially the (non-)occurrence of word-final vowel-raising (i.e.imāla; see Chapter 9), may just as well have motivated some or all citations of this lexical unit. The same thinking governed the composition of a discrete category for the various words (e.g. /mna:ħ/) meaning *good*.

Some lexical items had been mentioned by listeners in either or both of the previous Arabic experiments: /hado:l/ Iraqi and Kuwaiti /k(V)l:iʃ/ and the phrase *be careful* (in Damascene Arabic, /di:rba:lak/). As expected, considering the much larger number of listeners in this experiment, there were many additional lexical cues which were not present in the literature but which had to be accounted for here.

Table 11.13 specifies the form of the lexical features mentioned by listeners in Ar-3, as they occur in various Arabic dialects/dialect groups. It is not an exhaustive accounting of listener commentary on lexis arising from Ar-3. For example, general references to "words", "vocabulary", "terminology" and so on were categorised as 'unspecified lexicon', and this rather nebulous category was among the best-populated of all. This category is catalogued in Tables 11.14 (i) to 11.15 (ii) but not included in Table 11.13, which is intended as a full accounting of explicit references to *discrete* lexical units.

Table 11.13: Lexical features cited by listeners, by Arabic dialect

Syria	Egypt	Iraq/Kuwait	Morocco	Lebanon
/hado:l/ (*they*)	/du:l/ (*they*)			/hado:l/ (*they*)
	/ʔawi/ (*very*)	/k(V)l:iʃ/ (*very*)		
/mna:ħ/ (*good*)	/wiħʃi:n/ (*brutal/bad*)	/ze:ni:n/ (*good*)		
/lam:a/ (*when*)			/faʃ/ (*when*)	
/sħa:bu/ (*his friends*)	/ʔasħa:bu/ (*his friends*)	/ʔasdiqa:ʔa/ (*his friends*)		
/lahiʤtak/ (*accent*)			/djal:ak/ (*accent*)	
/di:rba:lak/ (*be careful*)	/xal:i ba:lak/ (*be careful*)			
/lʔaɣra:dˤ/ (*things/objects*)			biz:af (*things/objects*)	
/qarji/ (*countryside*)				/qarji/ (*countryside*)
/il:i/ (*who*)				/il:i/ (*who*)
	/butu:ʕ/*			

(*described by a non-native linguist listener as a "genitive exponent")

Many of the features listed in Table 11.13 are apparently etymologically distinct forms with similar functions/meanings (e.g. the word for *when*: Syrian /lam:a/ vs. Moroccan /faʃ/). Others exhibit plainly discernable phonetic contrasts. For example, the MSA word for *countryside* occurs in Damascene Arabic as /qarji/, in Kuwaiti Arabic as /qarje/ in Lebanese Arabic as /ʔarje/ and in Jordanian Arabic as /garje/. Still others, such as Syrian /il:i/ (*when*) occur in dialects of the Levant, Mesopotamia and the Gulf. Yet it is possible that concealed within these surface forms, with their readily observable contrasts, there are properties above the segment which are perceptible only to some types of listener (cf. Nolan 2012).

Tables 11.14 (i) and 11.14 (ii) quantify diagnostic lexical feature citations per group; Tables 11.15 (i) and 11.15 (ii) quantify salient-only lexical feature citations per group. Citations by NSNLs are again highlighted in red font.

Table 11.14 (i): number of diagnostic lexical feature citations per group

	Egyptian /ʔawi/	/hado:l/ + variants	/qarji/ + variants	Egyptian /wiħʃi:n/	Egyptian /butu:ʕ/	Totals
Non-native linguists	1	11	3	3	4	22
Syrian linguists	0	2	1	1	1	5
Non-Syrian linguists	3	13	2	6	2	26
Syrian NSNLs	1	14	5	3	3	26
Non-Syrian NSNLs	1	5	9	3	0	18
Totals	6	45	20	16	10	97

Table 11.14 (ii): number of diagnostic lexical feature citations per group

	/lam:a/ + variants	eastern Arabic /il:i/	Iraqi + Kuwaiti /k(V)l:iʃ/	be careful	objects/things	(his) friends	good	accent	unspec. lex.	Totals
Non-native linguists	2	3	6	7	10	3	7	9	2	49
Syrian linguists	3	3	6	5	2	3	0	3	6	31
Non-Syrian linguists	1	5	7	6	7	2	5	6	33	72
Syrian NSNLs	3	0	9	9	9	3	0	3	6	42
Non-Syrian NSNLs	1	0	13	5	17	0	1	3	8	48
Totals	10	11	41	32	45	11	13	24	55	242

Table 11.15 (i): number of salient-only lexical feature citations per group

	Egyptian /ʔawi/	/hado:l/ + variants	/qarji/ + variants	Egyptian /wiħʃi:n/	Egyptian /butu:ʕ/	Totals
Non-native linguists	0	2	3	0	0	5
Syrian linguists	0	1	1	0	0	2
Non-Syrian linguists	0	0	1	1	0	2
Syrian NSNLs	0	3	1	0	0	4
Non-Syrian NSNLs	0	3	2	0	0	5
Totals	0	9	8	1	0	18

Table 11.15 (ii): number of salient-only lexical feature citations per group

	/lam:a/ + variants	eastern Arabic /il:i/	Iraqi + Kuwaiti /k(V)l:iʃ/	be careful	objects/things	(his) friends	good	accent	unspec. lex.	Totals
Non-native linguists	0	0	0	0	0	0	0	0	0	0
Syrian linguists	0	0	1	0	0	0	0	0	0	1
Non-Syrian linguists	0	0	0	0	1	0	0	0	0	1
Syrian NSNLs	1	0	0	0	0	0	0	0	1	2
Non-Syrian NSNLs	0	0	0	1	4	0	1	0	1	7
Totals	1	0	1	1	5	0	1	0	1	11

11.6.3.4. Discussion

The first point to note is that all features, for all listener groups, were more diagnostic than they were salient-only (cf. 733 diagnostic vs. 58 merely salient total citations). Nonetheless, features that do *not* coincide with those mentioned in the literature appear to have been *relatively* more diagnostic for NSNLs—and for Syrian linguists—than they were for linguists (compare table 11.9 with 11.10 and 11.14 with 11.15).

Second, for listeners in general, lexis was the most diagnostic feature category of all, with 339 citations—c. 50% of the total number. Phonetic/phonological and morphosyntactic features

coinciding with those in the literature were the next most diagnostic (225 citations combined, and segmental and suprasegmental phonetic/phonological not present in the literature were the least (152 citations). Both of the above observations are consistent with the broader quantitative comparisons made below, in subsection 11.6.5.

In turning now to briefly consider features associated with incorrect or uncertain responses, I disregard non-specific allusions to segmental phonology ("pronunciation", "accent" and the like) and lexicon ("words" and similar). With these left aside, it appears that such responses by NSNLs were more associated—relative, that is, to all other groups—with *all three* broad categories of feature: lexicon (18 of 29 total citations), phonetic/phonological features *not*mentioned in the literature (7 of 11 citations) and phonetic/phonological/morphosyntactic features mentioned in the literature (13 of 18 citations). In this connection, however, recall that in Ar-3: (1) no effect of education was found on either accuracy or education; and (2) Syrians were more accurate and more confident than all other groups on Syrian stimuli. This suggests that neither quantity nor type of feature citations bears any perceptible relationship to accuracy or confidence. This observation reinforces those made in Chapters 9 and 10 and further amplified in subsection 11.6.5.

It must also be noted, however, that a purely quantitative calculation of individual citations does not reveal the entire picture. For the purposes of developing a supplementary test of production, it is vital to consider the *type of stimulus* (i.e. Syrian or non-Syrian and of what nationality) to which individual citations refer. This factor is examined, in connection with *some* of the above features, in Chapter 12. A full consideration of such is, however, beyond the range of the present thesis.

11.6.4. Overall quantification of feature citations by group

In investigating differences in the response patterns of the five groups in Ar-3, a further matter of interest is comparing the *overall* quantity of feature citations per group. Figure 11.16 shows the raw number of citations of cues by group. Here, unlike above, 'diagnostic' refers to correct responses, and 'salient' refers to any response: correct, incorrect or uncertain.

Figure 11.19: Raw number of salient/diagnostic citations of cues by group
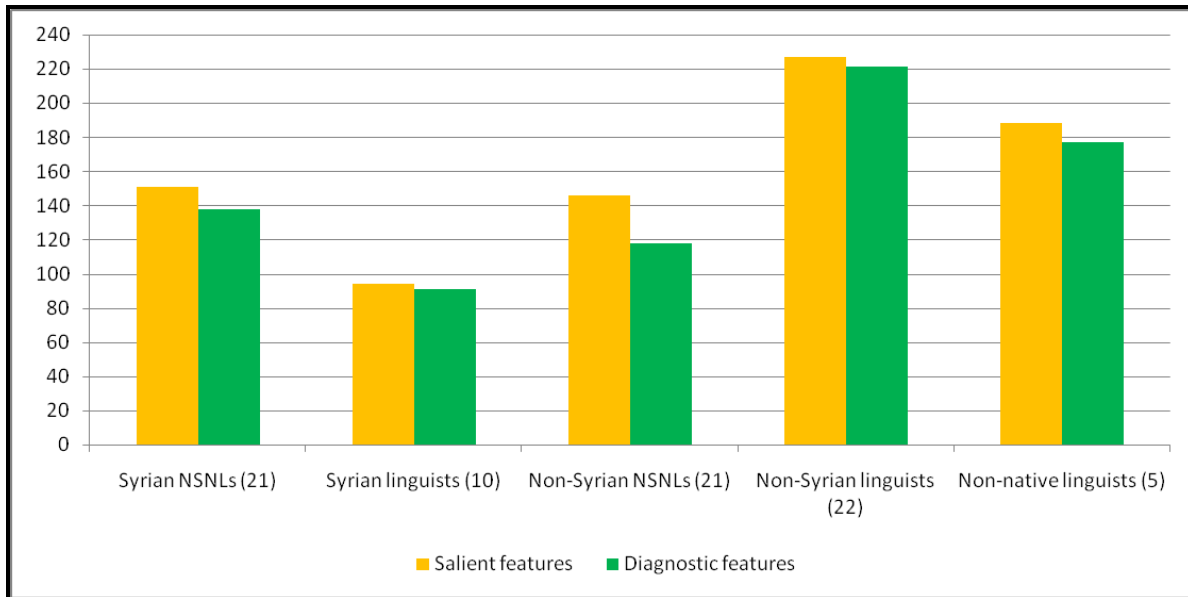
Figure 11.19 prompts the observation that, for all groups, there is much closer relationship between salience and diagnostic value than seen in previous Arabic experiments, in which the diagnostic value of cues was *far* inferior to their salience.

A second ready observation is that the number of both salient and diagnostic features cited by the 22 non-Syrian linguists is considerably greater than that of any other group. However, both figures are lower than those of non-native linguists when the average number of comments per listener is considered. This is illustrated by Table 11.6.

Table 11.16: Mean number of salient/diagnostic feature citations per listener, per group

	Syrian NSNLs (21)	Syrian linguists (10)	Non-Syrian NSNLs (21)	Non-Syrian linguists (22)	Non-native linguists (5)
Salient features	7.2	9.4	6.6	10.8	37.6
Diagnostic features	6.6	9.1	5.4	10.5	35.4

Another observation to be drawn from Figure 11.19 is that the quantitative gap between salient and diagnostic feature citations is narrower among the linguist than the non-linguist groups.

Further, as Table 11.16 also shows, the linguist groups each entered higher average numbers of both salient and diagnostic citations than either of the non-linguist groups.

A relevant factor in considering the possible meaning of the above observations is the distribution of feature citations within and across groups. This is an important because individual listeners can vary widely in the number of features they cite and, in doing so, render group-level comparisons misleading. Figure 11.20 shows the distribution of raw number of comments, per listener, per group.

Figure 11.20: Distribution of raw number of comments, per listener, per group
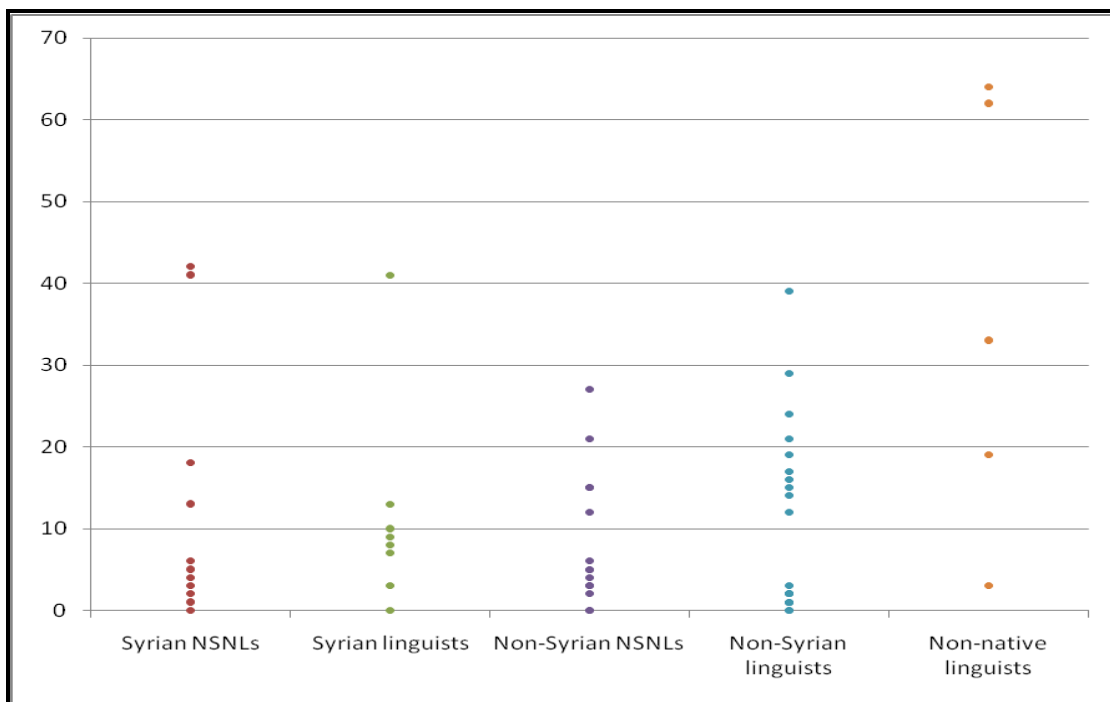
There are three observations of interest to be derived from Figure 11.20. First, while most listeners belonging to the two Syrian groups cited 10 or fewer features, there were rank outliers in each at the upper end of the range: two Syrian NSNLs and one Syrian linguist entered more than 40 citations each. Such outliers naturally skew upwards the group-level patterns in Figure 11.19 and Table 11.16, meaning that the numbers shown there should be treated with caution. This is less so in the case of the two non-Syrian groups, each of which evidences a fairly evenly-distributed pattern, with obvious outliers largely absent at the upper end of the range.

Second, Figure 11.20 indicates that—at the lower end of the range—the two NSNL groups, on

the one hand, and the two linguist native-speaker groups, on the other hand, show a similar pattern to each other. Nine of 21 Syrian NSNLs and 10 of 21 non-Syrian NSNLs cited no features whatsoever, but only three of 10 Syrian linguists and three of 22 non-Syrian linguists did likewise. In this respect Figure 11.20 is reasonably congruent with the results in Table 11.16, which, for the NSNL and linguist native-speaker groups respectively, shows similar mean numbers of feature citations.

Third, Figure 11.20 demonstrates that the distribution of citations among the non-native linguist group is distinct from those seen in the native-speaker groups. Most crucially, the range is much greater: from three at the lower end to 63 and 64 at the upper end. This means that the results of Figure 11.19 and Table 11.16 must again be viewed with circumspection.

Nevertheless, the fact remains that two of five non-native linguists entered a far larger number of citations than any individual native speaker, with the remainder entering at least some feature citations. Combined with the observations discussed above, a plausible interpretation is that non-native linguists generally maintain a greater ability and/or willingness than natives to cite the cues that guide their responses (non-Syrian linguists are a distant second in this regard but with a much more even spread per listener of feature citations). However, inferential statistical results indicate that this kind of 'explicit' knowledge—accumulated via formal, specialist education—yielded no overall advantage in accuracy or confidence.

Given this, the conclusion once again—as expressed also in subsection 11.6.3.4.—is that number of feature citations likely bears little relationship to accuracy. With reference to the theoretical assumptions of the thesis (see Chapters 4 and 5), a further warranted inference is that NSNLs appear to depend to a greater degree than linguists—even those of Syrian origin—on tacit, holistic knowledge, which proceeds from membership of a speech community, unmediated by the kinds of explicit knowledge acquired through specialist education. In other words, paraphrasing Polanyi, it seems that NSNLs indeed do appear to know more than they can say.

11.6.5. Group citation of features within/outside consulted dialectological literature

The next matter of interest is the overall degree to which each group, in responding to the stimuli, relied on features identifiable with those mentioned vs. those *not* mentioned in the surveyed dialectological sources (e.g. Behnstedt 2012, Behnstedt & Woidich 2013, Lentin 2012, Owens 2001, Procházka 2012, Versteegh 2001, Woidich 2012, Hellmuth 2019). The latter category includes the phonetic/phonological, morphosyntactic and lexical types discussed in subsection 11.6.3.

Figure 11.21 shows the overall percentage, by group, of diagnostic feature citations mentioned/not mentioned in the dialectological literature. Figure 11.22 specifies the percentage of salient-only features (i.e. associated with incorrect/uncertain responses), identically sub-classified.

Figure 11.21: Percentage of citations, by listener group, of diagnostic features present/not present in the dialectological literature
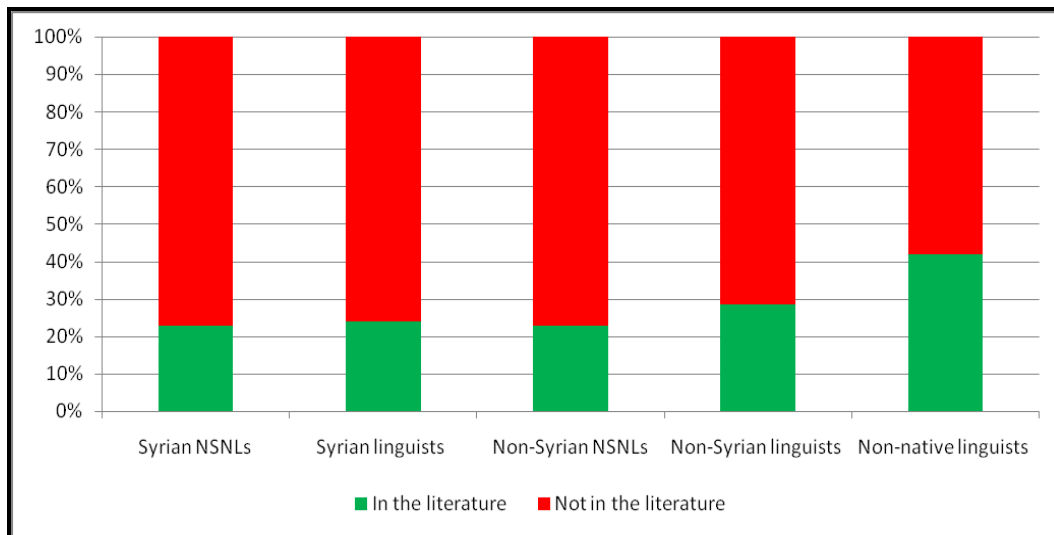
Figure 11.22: Percentage of citations, by listener group, of salient-only features present/not present in the dialectological literature
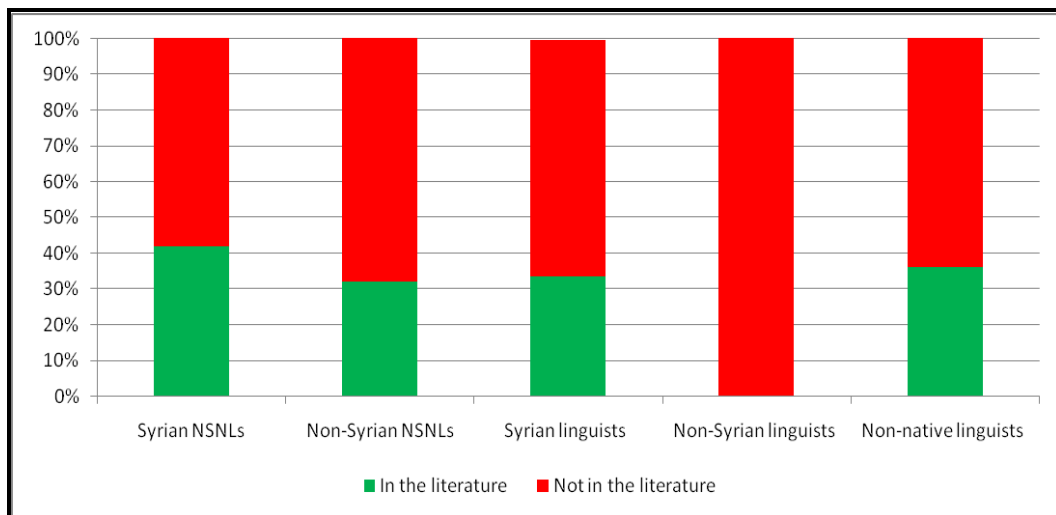
There are two important observations to be gleaned from Figure 11.21. First, in correctly identifying stimuli, all groups relied more on features not present in literature surveyed than on those present; both NSNL groups exhibited the strongest inclination (albeit only slightly stronger than Syrian linguists) in this direction. Second, non-native linguists relied considerably more than any other group on features present in the literature—to nearly twice the extent of Syrian linguists, as well as of Syrian/non-Syrian NSNLs.

From Figure 11.22 it is apparent that *all* the features found salient-only (i.e. leading to incorrect/uncertain responses) by the non-Syrian linguist group were outside the consulted sources on Arabic dialectology. Such features indeed informed the majority of incorrect/uncertain responses among all groups, but the percentage of features present in the literature cited in connection with such responses was higher among all groups than among non-Syrian linguists: > 40% in the Syrian NSNL group, for example.

Together, these observations suggest that features coinciding with those mentioned in the extant dialectological literature on Arabic were apparently of limited (though not negligible) assistance to listeners in reaching correct conclusions as to the origins of the stimuli. Dependence on features outside the literature, however, may also have led listeners to

incorrect or uncertain conclusions.

The fact that all groups—including linguists, though they to a considerably lesser degree than others—relied predominantly on undocumented features is perhaps explained by the fact that the dialectological field worker approaches his/her work with a particular focus. In most cases, it involves collection of data on dialects at the segmental (mainly phonetic/phonological and morphosyntactic) level. This focus naturally shapes the content of the published surveys. But in adopting such a focus the fieldworker is not necessarily representing what native speakers, or even non-native speakers, would find most salient in a given dialect. As noted in 11.6.3.4, listeners seem to have depended on lexical features, which are relatively poorly catalogued in the sources surveyed. This underlines the in-principle legitimacy of Lucas's test of lexicon for LAAP (cf. Chapter 2).

For real-life LAAP one possible consequence of the above is that, when an NSNL analyst is asked to determine the provenance of an accent, he/she is constrained by the literature's primary focus on segmental phonetics/phonology and morphosyntax. This would be especially so in agencies where the format of reports is based largely on segmental features attested in the dialectological sources. In short, it may be the case that the focus and format of LAAP reports do not allow NSNLs' intuitions about the linguistic identity of the speaker to be fully expressed, even under the guidance of linguists. The NSNL is further handicapped in the ability to state systematically, through the use of proper metalanguage (including the IPA), what makes the speaker's accent authentic or not.

A practical corollary of these observations is that, in their reporting of conclusions, LAAP agencies may be well served by not relying too heavily on features mentioned in the existing surveys of Arabic dialectology—a point made repeatedly elsewhere in this thesis. Instead, the reporting of, and greater emphasis upon, a wider range of features might best reflect the perceptions of analysts and better serve the interests of justice. Doing so would avoid the retrofitting of analysts' conclusions in line with the limited range of features found in the literature.

Matras (2018) suggests his databases as one way of avoiding this problem (cf. Chapter 2). He

also alleges that the format of some LAAP reports is an invariant "…template of cited references and diagnostic features…" (2018, p. 73). Locating and describing (as far as possible) the properties of language perceived only by native speakers would be an effective answer to charges levelled at LAAP agencies by their critics. It is partly with this end in mind that the qualitative data reviewed above is made available in Appendix D of the thesis.

11.6.6. Citation of individual features in relation to individual stimuli

Absent from the above survey and discussion of qualitative data is a consideration of how groups may differentially pattern in their citation of individual features in relation to individual stimuli. This question could be evaluated in a number of ways. A dataset of the size gathered here (791 individual feature citations) may reveal a great deal about which individual cues were most frequently cited by which group, in association with which stimuli nationalities and with what degree of accuracy/confidence. Groups' respective dependence on cues within and outside the literature, approached *en masse* in the section above, could be explored at a greater level of resolution.

The limitations of my own knowledge of Arabic mean that the requisite level of analysis is not attempted in this thesis. The data therefore awaits full examination, perhaps in collaboration with a linguist native speaker consultant.

11.6.7. Review of findings in connection with research objective 2

The foregoing enquiries into feature citation patterns have concentrated on intergroup differences from a purely quantitative perspective: which groups are most inclined to cite features of any type? To what extent do groups vary in their reliance on features occurring in the literature vs. those that do not? On the basis of this *en masse* approach to feature citations, a number of telling patterns have emerged, in line with the assumptions underlying hypothesis (iii). First, non-native linguists are by a considerable margin more willing/able than other groups to cite features in connection with their conclusions. Second, NSNLs generally are more likely than any other type of listener not to cite any features whatsoever. Third, all listener groups cite a greater number of features that are *not* mentioned than *are* mentioned in the literature

consulted. Non-natives, however, conform least to this pattern, evidencing notably greater dependence on features coinciding with those in the literature, especially in connection with diagnostic citations.

From these findings, and their relation to the descriptive and inferential statistical analyses of the data (see 11.5.3 and 11.5.4) several inferences can be made. First, listeners' revealed willingness/ability to cite features in association with their conclusions does not appear to bear a strong relationship to accuracy. Secondly, listeners' surprisingly limited reliance on features mentioned in the published dialectological literature suggests both that there may be gaps in these sources and that knowledge acquired via specialist education may play a lesser role than other sources of knowledge, even in the judgements of linguistically-educated listeners (perhaps of the 'tacit' kind). Finally, it is cautiously recommended that LAAP practitioners might better serve the interests of justice by avoiding exclusive dependence on existing dialectological accounts in reaching conclusions as to the linguistic identity of asylum applicants, even considering the relatively well-documented status of Arabic in comparison to other languages.

The final question to be addressed here is the degree of support gleaned for hypothesis (iii) of the thesis:

Analysis of feature citations will show that speaker-listeners vary by region of origin in the number and kinds of cues they recognise.

The above review of findings, and the data on which it is based, suggest some support for this hypothesis. Firstly, non-natives differ substantially from other groups in the quantity of their feature citations. Secondly, non-natives appear to depend more closely than natives on features corresponding to those present in the reviewed dialectological surveys of Arabic. It may be, though, that more interesting and practically informative divergences in intergroup feature citation patterns can be revealed only by proper examination of individual features. Once again: the full range of comments is available in Appendix D to those with the aptitude to do it justice.

In Chapter 12, I suggest, in broad terms, how patterns observable in the citation of individual features in Ar-3 might be used in the composition of a 'listen-repeat' test of language production for LAAP. I also outline the possible shape of a test of perception incorporating

insights drawn from the inferential statistical analysis conducted in this chapter.

Chapter 12: Design and interpretation of novel tests for LAAP

12.1. Introduction

Together, the experimental results detailed in Chapters 7 to 11 constitute an encouraging proof of concept. They indicate that both Yorkshire and Syrian natives are more accurate than non-natives in the 'one of us' task: identifying fellow native speakers. Of particular relevance to LAAP is the fact that the results of Ar-3 (see Chapter 11) show a statistically significant advantage for Syrian speakers in the 'one of us' task. For the current practice of LAAP, the principal finding of Ar-3 is that linguistic education reveals no significant effect on accuracy. This suggests that NSNLs are competent to act as judges of their own language variety.

The proof of concept established, the questions still to be addressed in this thesis are: How can results be translated into novel tests of (1) the linguistic perception and (2) the language production of asylum applicants? In the remainder of this chapter I attempt to address these questions. In 12.2 I suggest how a supplementary perception test might be designed, and in 12.3 how its results might be interpreted. In 12.4 I outline in very broad terms the possible shape of a supplementary test of production.

12.2. Supplementary perception test: possible design and interpretation

Initially in Chapter 1, and subsequently elsewhere, I advanced the idea that a new perception test would differ from the current practice of primary-phase LAAP in three ways. First, the applicant would not be required to speak in taking the test. Consequently, oft-expressed misgivings in the literature—especially as to whether unexpected features of speech emerging from episodes of secondary socialisation could be misattributed to language imitation—might be quelled. Second, also because the new perception test would not involve an interview, no interlocutor would be necessary, which would avert any possibility of accommodation by the applicant to an interviewer's speech. Third, test results would not require interpretation by an

NSNL analyst—or indeed by anybody familiar with the relevant language/variety. Below I demonstrate how this is so.

12.2.1. Perception test: design

The applicant would be directed to answer a similarly simple question to that posed in the experiments in this thesis. One possibility would be a formally identical question, 'Is this a [applicant's claimed or inferred primary linguistic variety] accent?' Another might be 'Does this person speak [applicant's claimed or inferred primary linguistic variety] as you do?' The most apposite phrasing would require investigation in future empirical work.

However the question were to be phrased, the claimed or inferred linguistic variety would be matched as closely as possible to that assessed via the main LAAP task, the speech data for which is derived from an interview. The test design takes account of enduring criticisms, reviewed in Chapters 1 and 2, of the practice of primary-phase LAAP. These centre on two main assertions: first that the interview from which the speech sample is drawn may occasion accommodation by the interviewee to the speech of the interviewer; second that possible evidence of secondary linguistic socialisation or a complex life history, in line with the applicant's claims, is inadequately accounted for or misinterpreted as evidence of language imitation. In the perception test described here, unlike in the interview, the test administrator would not speak to the applicant except in issuing instructions and (if required) in clarifying the tasks involved. Neither would the applicant be required to speak. Because secondary socialisation and accommodation can only be expressed in the applicant's speech, and not in his/her perceptions, the test design both acknowledges and entirely circumvents consideration of the factors from which previous criticisms have arisen. Instead, it concentrates on confirming or casting doubt on conclusions drawn from the interview data as to the applicant's linguistic background.

As reviewed in Chapter 2, the LAAP interview is commonly conducted over video-link (Skype or similar). The supplementary perception test could be administered via the same medium, on

the same day as the interview. To limit the risk of the applicant's being primed by the variety spoken by the interviewer, the supplementary test would be best administered before, not after, the interview.

The test interface would require some development. This lies beyond the range of the present work. It would, however, naturally include a number of audio stimuli featuring the target variety, as well as a number of 'foils'. For maximum audio quality, stimuli should be presented to the applicant via headphones (not speakers) connected to the applicant's device.

As in Ar-3, stimuli would be presented in (pseudo-)random order and played at the applicant's prompting. The interface would show the available range of responses to each stimulus. To cover the possibility that the applicant is (or claims to be) illiterate, available responses could be shown in the interface as icons rather than in writing. One minimal possibility would be a binary yes/no task, with a green tick for 'yes' and a red cross for 'no'. The applicant would mouse-click on the relevant icon to indicate their response to each stimulus. These responses would then be recorded and subsequently interpreted.

Six core questions arise from the above:

(1) How would the necessary samples be gathered, and from whom?

(2) What would be the optimal duration of each sample?

(3) How many times would the applicant be permitted to listen to each sample?

(4) What would be the optimal number (and geographical origin) of 'target' and 'foil' samples, respectively and in total?

(5) How and by whom would test performance be interpreted?

(6) What range of response types would be permitted?

In the remainder of this chapter I address each of these questions in turn.

12.2.1.1. Gathering samples

Samples could be extracted from audio recordings of individuals known to speak the relevant variety—e.g. friends or family members of NSNL analysts who work for LAAP agencies (see further subsection 12.3.2). In this manner a library of samples could be assembled that feature the particular regional variety or varieties spoken by the members of the NSNL analyst's network. The same method could be adopted to gather 'targets' and 'foils'.

A conceptually similar procedure of collecting reference data from informant networks is used in forensic phonetics (Foulkes & Hughes in press; Foulkes & French 2012). Admittedly, this may prove difficult in the case of particular varieties with comparatively few speakers; indeed, Foulkes & Hughes (in press) make the same point with reference to forensic phonetics. Yet, if agencies have in their employ NSNLs capable of assessing a given variety in the primary LAAP task, there is a reasonable chance that the same variety can also be assessed in a supplementary perception test through the activation of networks connected by degrees to the NSNL analyst.

YouTube or existing corpora are other possible sources, for both targets and foils. It is perhaps not very likely that recourse to such sources of speech material would yield samples of narrowly-spoken dialects. As an alternative, widely-spoken varieties such as Damascene Arabic could be used to assess the applicant's judgements of speech at the national rather than regional level. This approach was adopted for the Arabic series of experiments in the thesis— and, as Ar-3 demonstrates, Damascene Arabic is significantly better recognised by Syrians generally (regardless of their regional origin) than by non-Syrians. Whatever the source or type of samples, a suitable library could be amassed, even from multiple different passages featuring the same speaker. This could be accomplished, for example, simply by manipulation of fundamental frequency values.

12.2.1.2. Sample length

In Eng-2 and all three Arabic experiments, c. 10-second samples were sufficient to reveal significantly more accurate performance by local listeners in their respective 'one of us'? tasks. There is no reason in principle to believe that this sample duration would require modification in a similar, real-world test. However, piloting would likely have to be conducted to further validate this assumption in practice.

12.2.1.3. Number of repetitions of each sample

In all the experiments presented in this thesis, listeners were allowed unlimited exposure to each sample. It is doubtful that this would be practical in a real-world test, where time is necessarily limited (cf. Chapter 2, subsection 2.2.3: Verified must produce final reports for the Home Office within five working days). One alternative, of course, would be to allow each sample to be heard only once. However, given the possible occurrence of human error (headphone slippage, interruptions to playback etc.), this would not be optimal.

A better alternative may be to impose an upper limit. A maximum of three to five playbacks seems reasonable, particularly since listeners have been found to be less subject to cognitive processing costs in recognising familiar accents (Sumner & Samuel 2009).

12.2.1.4. Number and geographical origin of samples

In the Arabic-language experiments conducted for this thesis, the number of samples, as well as the ratio of 'targets' to 'foils', varied. In total, there were 52 in Ar-1, 21 in Ar-2, and 22 in Ar-3.

In Ar-1, 10 of the samples featured Syrian targets. The remaining 42 consisted of Egyptian, Moroccan, Tunisian, Omani, Jordanian, Iraqi, Kuwaiti and Lebanese foils. In Ar-2, there were 6 Syrian targets. Foils were from Jordan, Iraq, Kuwait and Lebanon. In Ar-3, there were 10 Syrian targets, and foils were from Egypt, Morocco, Jordan, Iraq, Kuwait and Lebanon.

Egyptian, Moroccan, Tunisian, Omani samples were easily rejected as Syrian by almost all listeners across all three experiments. Jordanian, Iraqi, Kuwaiti and Lebanese foils were less

readily rejected. The former three, especially the Jordanian type, proved most difficult for Syrians to correctly reject, while Lebanese samples were most likely to be incorrectly accepted by non-Syrians. All four varieties are either closely related to Syrian Arabic (Lebanese) or are spoken in countries bordering Syria (Kuwaiti Arabic is the exception, though it is very closely related to the Arabic of Iraq).

Statistically significant differences in accuracy between Syrians and non-Syrians were revealed in Ar-3. No tests of significance were conducted on the results of the other two Arabic experiments, though descriptive statistics showed sharp intergroup discrepancies in response patterns.

For a real-world test, 52 samples would likely be too large a number, especially if the test were administered on the same day as the interview, which might cause fatigue. However, Moosmüller (2011) locates variability in patterns of attention, engendered by task repetition, as a factor negatively affecting the accuracy of non-authentic experimental subjects only (see below, Section 12.4, for a fuller discussion of Moosmüller's findings). For this reason, deployment of a comparatively large number of samples might occasion, or at least accentuate, differences in the accuracy of authentic and non-authentic asylum seekers.

As mentioned in subsection 12.2.1.3, militating against any potential decision to use a large number of samples (e.g. the 52 featured in Ar-1) is the practical consideration of limited time. Overall then, the following two-part proposal on stimuli number seems reasonable.

First, given that a total number of around 20 samples—with an approximately balanced target-to-foil ratio—demonstrated significant differences in local vs. non-local listener performance, there are grounds for thinking that a similar total would be sufficient in a real-world test. Second, whatever the total number of stimuli used, a similarly balanced selection of targets and foils—the latter including mainly neighbouring varieties exclusive of Egyptian Arabic, which showed high discriminability for all listeners in the Arabic experimental series—should be employed.

Excluding Egyptian samples, the total number (20) and ratio of targets to foils used in Ar-3 (10/10) may be a sensible minimum, especially since c. 65% of listeners took under 10 minutes

to complete the survey, and c.35% (of listeners overall) took under five minutes. However, with time limitations in mind, this suggestion should be weighed against decisions made as to the permitted number of repetitions of each sample: more repetitions = fewer samples and vice versa. Another consideration is the fact that there were no consequences for listeners arising from the accuracy of their responses to Ar-3. Further testing would have to be conducted in higher-stakes contexts, possibly incorporating rewards for accuracy (e.g. the awarding of vouchers for accuracy above a certain threshold.

12.2.1.5. Interpretation of test performance and permitted range of responses

If transparent and easily-interpretable statistical instrument(s) were used to assess test results, no ability in (or even familiarity with) the relevant language/variety would be required of the assessor. Neither would any statistical expertise necessarily be required. However, the type of statistical instrument selected may bear on the range of responses permitted in the test. These factors are considered in detail below, in section 12.3.

12.3. Perception test: interpretation

In Chapter 3 I argued that LAAP is fundamentally forensic in its context and—properly but not thoroughly in practice—in its approach to the data. It follows from this that any novel supplementary test would have to be conducted in a forensically rigorous manner; moreover, its results would have to be transparent to the decision maker.

Above I sketched a design for a supplementary perception test that would shift a greater share of the burden of proof onto the asylum applicant, in line with prescriptions enjoined by both the Dutch government and the UNHCR (cf. Chapters 2 and 3). There is, in addition, nothing in the test design suggested above that contravenes the general forensic principles advanced in Chapter 3. Below I suggest output derived from Signal Detection Theory as one way in which the results of such a test might be statistically interpreted in a manner transparent to any layman audience—including LAAP decision makers.

12.3.1. Signal detection theory: overview

The statistical output of signal detection theory (SDT) is used in a variety of domains, including diagnostic medicine, psychology and—in the field of linguistics—sentence acceptability judgements (e.g. Abdi 2009; Stanislaw and Todorov 1998; Dillon & Wagers 2021). According to Stanislaw & Todorov (1999), SDT describes the ability of subjects, human or machine, to distinguish audio stimuli featuring only 'noise' from those featuring 'noise + signal'. Like the experiments in this thesis and the tests based upon them, it is thus a measure of difference in perception. Abdi (2009) explains that the constructs 'signal' and 'noise' should be thought of non-literally in some situations. In these experiments, for example, 'noise + signal' stimuli are equivalent to targets, and 'noise-only' stimuli to foils.

SDT is suitably applied to task designs entailing a binary response—i.e. basically 'yes' or 'no' (Abdi 2009). In this respect it was not applicable to analysing the results of the experiments in this thesis, permitting as they did the expression of judgements on a non-binary scale—including an 'uncertain' alternative. Yet, as I argue below, SDT may still be profitably used in interpreting analogous perception tasks in LAAP (see further discussion in 12.3.2).

SDT quantifies values resulting from the operation of two perceptual factors: the subject's decision *variable* and his/her decision *criterion* (Stanislaw & Todorov 2008). The two factors interact but are independent: the operation of the decision variable results in a 'yes' response when the decision criterion is met and a 'no' response when it is not.

Signal + noise stimuli, when correctly identified by the subject, are termed 'hits' (correct acceptances), whereas 'false alarms' (incorrect acceptances) emerge from noise-only stimuli, to which the subject mistakenly attributes the presence of the signal(s). Thus 'misses' (false rejections) and correct rejections can readily be inferred from hits and false alarms respectively (Stanislaw & Todorov 2008).

To understand the importance of the interaction between the decision variable and the decision criterion, consider that in Ar-3 there were 10 targets and 12 foils. If a listener were to respond 'no' ('This *is not* a Syrian accent') to every sample, the listener would score 12/22 (i.e.

above chance). But this result, arrived at via the operation of the decision variable, could be achieved without even listening to the stimuli (Keating 2005).

SDT casts light on listeners' true sensitivity to the signal by also accounting for the working of the decision criterion: the extent to which listeners are inclined overall to answer 'yes' or 'no'—that is, their response *bias*. Measurement of the two factors, the decision variable (signal sensitivity) and the decision criterion (response bias), elucidate the ratio of hits, false alarms, misses and false rejections encoded in listeners' overall response patterns.

The results of SDT are represented graphically, and are relatively easy to interpret. SDT graphs show two distributions: the left-hand curve represents how values (i.e. 'yes' or 'no') are distributed in noise-only stimuli, and the right-hand curve represents their distribution in signal + noise stimuli.

The difference between the mean scores of these two distributions, which specify the actual mean value of subject responses to all stimuli, is referred to as d-prime (*d'*)—i.e. the statistical measurement of the decision variable, or *signal sensitivity.* The overall tendency of subjects to respond either 'yes' or 'no' to stimuli across stimuli is commonly referred to as β*:* the measurement ofthe subject's decision criterion, or *response bias* (Stanislaw & Todorov 1999).

Vuorre (2017) explains that *d'* is calculated by the formula:

$$d' = \Phi^{-1}(HR) - \Phi^{-1}(FAR)$$

Here, HR is 'hit rate', FAR is 'false alarm rate' and Φ represents the cumulative normal density function; it converts z scores into probabilities. $\Phi{-1}$, meanwhile, converts hit and false alarm rates into *z*-scores. Abdi (2009) states that a simplified formula for deriving *d'* can be expressed as $Z_H - Z_{FA}$, where Z is the z-score, H is hits and FA is false alarms.

According to Stanislaw & Todorov (1999) the natural logarithmic formula for β is:

$$\beta = e^{\left\{ \frac{\left[\Phi^{-1}(F)\right]^2 - \left[\Phi^{-1}(H)\right]}{2} \right\}^2}$$

Here, *e* represents the base number of the natural logarithm.

From the above can be derived the following formula, where *Φ−1* has the same value as in the *d'* formula expressed above:

$$\ln(\beta) = \frac{\left[\Phi^{-1}(F)\right]^2 - \left[\Phi^{-1}(H)\right]^2}{2}$$

In SDT, the more the signal + noise and noise-only distributions overlap, the less sensitive listeners are to the signal(s) and the lower their *d'* score. The further to the left of the intersection of the two distributions the criterion line falls on the *x*-axis, the more 'yes' responses are given and the lower the β value. A more right-inclined criterion line signifies more 'no' responses, and a higher value of β. A 'neutral' criterion line would mean no inclination towards either error type; note that the values on the x-axis are arbitrary (Stanislaw and Todorov 1998).

All this means that, assuming two graphs show identical *d'*, a criterion line (i.e. β) inclined more to the left would result in a greater number of hits and false alarms. This is known as a liberal response strategy. A right-inclined criterion line, meanwhile, would result in a higher number of both misses and correct rejections. This is referred to as a conservative response strategy (Abdi 2009; Stanislaw and Todorov 1998; Dillon & Wagers 2021).Thus, in two graphs showing equal *d'* scores, differing locations of the criterion line along the x-axis would signify different response strategies. These factors are shown in Figure 12.1.

Figure 12.1: modelling of sensitivity and response bias in SDT (Oliver et al 2008)
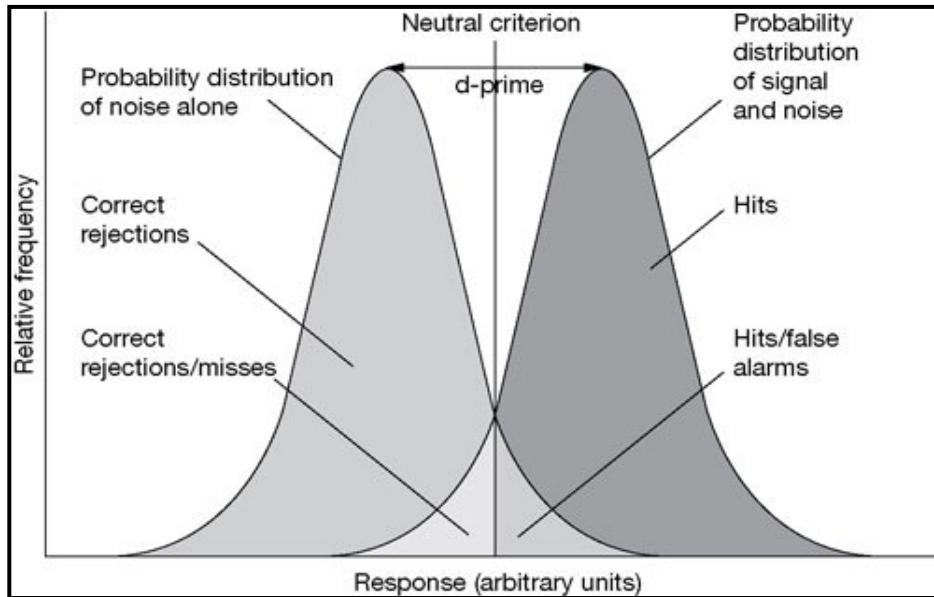
Keating (2005) provides the following values as illustrative of the scalar significance of *d'* values. Total inability to discriminate between signal and noise (such that the two distributions completely overlap) is equivalent to a *d'* of 0. The greatest possible signal sensitivity is represented by a *d'* of 6.93, and accuracy at 69% is reflected in a *d'* of 1.0.

12.3.2. SDT: discussion of suitability to a supplementary perception test

SDT, then, provides two interacting measures of listener perceptions: (1) overall sensitivity to signal-present as against noise-only stimuli; (2) overall bias in responding 'yes' or 'no'. As illustrated by the hypothetical example of the 12/22 score given above, by accounting for the interaction of these two measures SDT can separate genuine from non-genuine attempts at test-taking (e.g. responding 'no' or 'yes' to every stimulus). This makes SDT obviously suitable to the perception tests proposed, since it is quite plausible that non-genuine applicants could attempt to beat the test by giving the same response to every stimulus.

In order to demonstrate the meaning of an individual test-taker's β and *d'* values, gathering reference values for both measures would be necessary. Reference values would allow

comparison of individual test results with those achieved by, on one hand, groups of known *native speakers* and, on the other, groups of known *non-speakers* of the variety in question.

Naturally, the reliability of such reference values would depend on the gathering and analysis (via SDT) of responses from sufficiently large and representative samples of natives and non-speakers, as well as a distribution of results that adequately separate the two groups (this is likely to be gradient rather than absolute, which underlines the point made below as to the establishment of a meaningful statistical threshold for test performance).

Like the samples themselves, reference values for native speakers of the variety in question could be gleaned from the results of volunteer native-speaker test-takers from NSNLs' networks of trusted family members and friends. Reference values for non-speakers could be compiled from the test results of known non-speakers of the relevant variety.

In illustration of how this might work, Figures 12.2 and 12.3 simultaneously demonstrate in numeric and graphic form how the Syrian and non-Syrian NSNLs in Ar-3 differed overall in their response patterns, as analysed via SDT. For the purposes of demonstration, outright uncertain ratings are included as inaccurate responses—i.e. false rejections of Syrian stimuli and false acceptances of non-Syrian stimuli. This at least has the virtue of consistency; however, it naturally skews results in favour of correct responses on both types of stimuli, thus also skewing the *d'* and β values.

The above caveats entered, Figures 12.2 and 12.3 indicate clearly distinct values on both SDT measures: *d'* (2.79 for Syrians and 1.46 for non-Syrians—i.e. greater sensitivity among Syrians) and β (0.26 for Syrians and 1.5 for non-Syrians—i.e. a more liberal response strategy among Syrians). These are graphically represented by both the distance between the distributions and the distinct relative positions of the criterion line. Also represented are differences in the red-hatched and green-hatched areas; these show the probability of false alarms and hits respectively ([Elvers n.d. a](#)).

In subsection 12.3.1 I gave a somewhat technical account of the calculation of *d'* and β. Putting the matter in layman's terms, the value of β is the ratio of the height ('ordinate') of the signal + noise curve at the criterion line to the height ('ordinate') of the noise-only distribution at the

criterion line. The *d'* value is the distance between the respective averages of the left and right curves, with the average calculated from the proportion of hits to false alarms.

Figures 12.2 and 12.3: SDT output for Syrian and non-Syrian NSNL responses in experiment Ar-3 ([Elvers n.d. b](#))
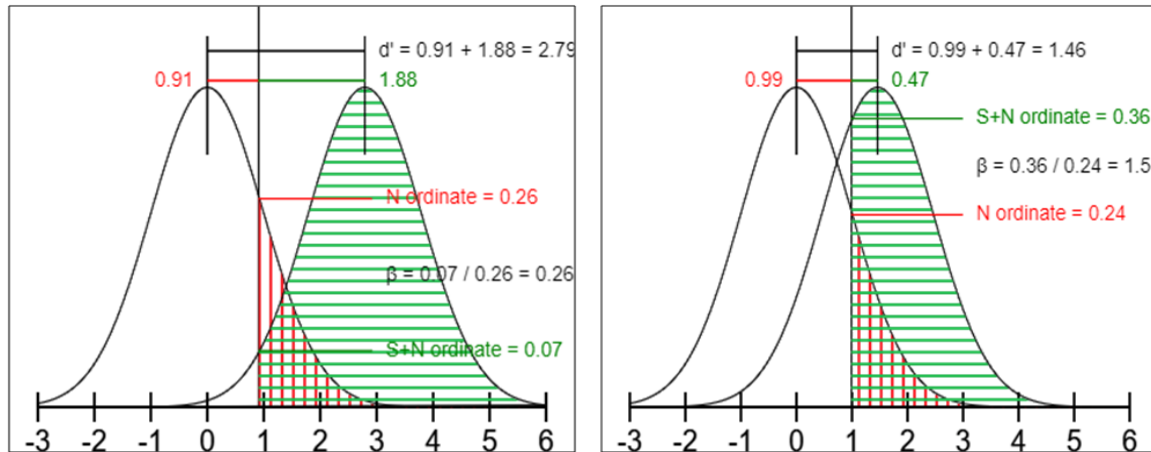
The above values and graphs conceptually represent baseline (or threshold) *d'* and β values for native speakers and non-speakers of Damascene Arabic, against which individual applicants' performance might be measured. The problem, of course, is that some authentic applicants will fall under the threshold for natives and some non-applicants above it. Therefore, in the case above—and for all varieties potentially assessed in a new perception test—measurements of the *likelihood* of native speakers and non-speakers achieving particular *d'* and β values would have to be established. As this task lies outside the range of the present thesis, I leave it aside for now.

To further illuminate how SDT measures might work in practice, I now return to the hypothetical example of the 12/22 asylum applicant test score, given in subsection 12.3.1. Let us assume that the applicant entered no outright uncertain responses to the stimuli in the test, and that the number of Syrian and non-Syrian stimuli is the same as in Ar-3 (i.e. 10 and 12 respectively). Under these conditions, a test-taker scoring 12/22 would evince a low *d'* value— i.e. the left and right distributions would be relatively close together—but the particular *d'* value would be dependent on the *types* of responses entered. The β value would also vary according to response types.

For example, if nine responses out of a possible 10 to Syrian stimuli were correct acceptances ('hits'), but only three responses out of a possible 12 responses to non-Syrian stimuli were correct rejections (i.e. nine 'false alarms'), the SDT graph and the two values would appear as in Figure 12.4.

Figure 12.4: 12/22 test score with 9/10 correct acceptances and 3/12 correct rejections (<u>Elvers n.d. a</u>)
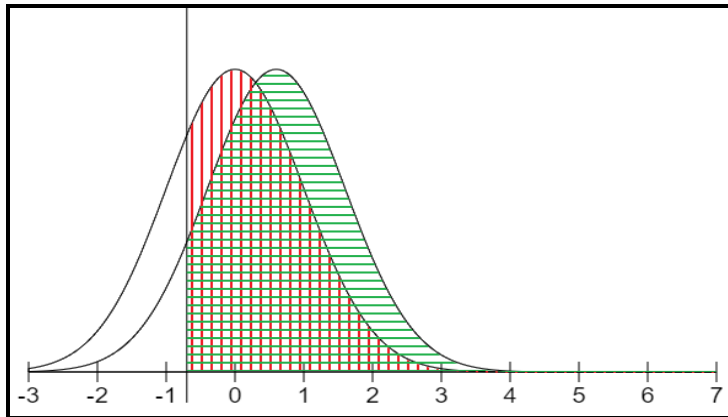
Here, the *d'* value is 0.6, and the β value is 0.549 (the latter does not properly align with the x-axis; but recall, cf. subsection 12.3.1, that the x-axis specifies arbitrary values).

Let us now suppose that, under identical conditions, the applicant instead achieved only three correct acceptances but nine correct rejections (i.e. only three 'false alarms'). In this instance, the SDT graph and the two values would present as in Figure 12.5.

Figure 12.5: 12/22 test score with 3/10 correct acceptances and 9/12 correct rejections (<u>Elvers n.d. a</u>)
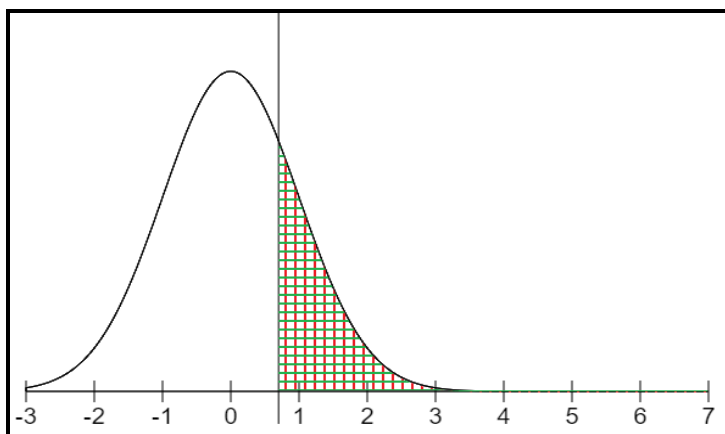
The *d'* value here is 0, and the β value is 1.

Figures 12.4 and 12.5 thus illustrate distinct signal sensitivities (*d'*) and response strategies (β). The test assessor may conclude, preliminary to considering computed statistical likelihoods (see above), that Figure 12.4 is more likely to represent a true Syrian than is Figure 12.5. Such a conclusion would be warranted because in 12.4 the *d'* is higher, and the β is lower, than those shown in 12.5; this comports with the comparison of reference values of *d'* and β, for Syrians and non-Syrians respectively, in Figures 12.2 and 12.3.

The transparent combination of numeric and graphic information shown by SDT calculations might aid interpretation of test results by non-specialists in statistics and non-speakers of specific language varieties—LAAP practitioners and decision makers alike. In real-world LAAP cases, SDT graphs and statistics like those above could be compared with the conclusion arrived at on the basis of the primary-phase LAAP interview data and then presented to the relevant decision maker. The report as a whole would, of course, best be presented to decision makers along with a brief explanation of the meaning of the *d'* and β values, the manner of their visual representation and the calculation of the thresholds and likelihoods for each kind of test-taker.

SDT values, and the associated figures, have the additional benefit that they can be easily generated and checked via online applications created by Greg Elvers (Elvers n.d. a, b).These applications were used to create the figures above. Eshed Margalit (Margalit n.d.) has likewise created an online SDT calculator though which graphs and data can be displayed. In both cases, one simply enters a number, between 0 and 1, for the ratios of hits and false alarms entered by the test-taker(s) and follows the steps to generate the values and figures. The required ratios can, of course, be easily derived from percentage scores—for instance, 0.9 = 90%. Figure 12.6 shows Elvers' interface.

Figure 12.6: interface of Elvers n.d. b
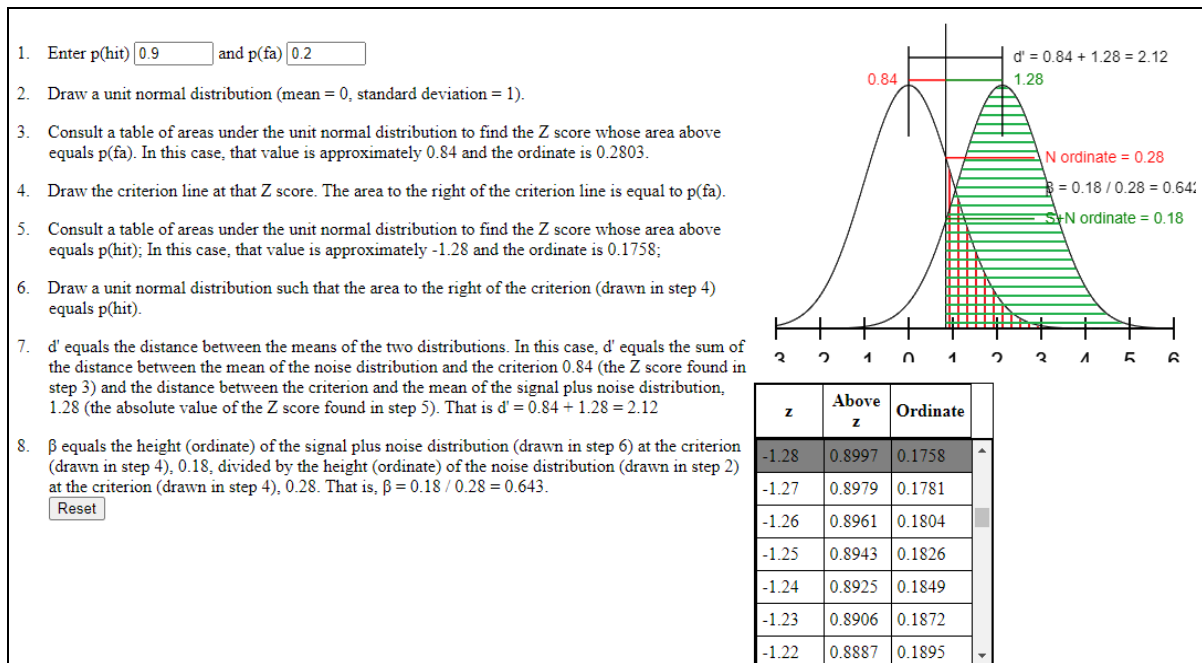
General statistical software programs—e.g. R (R Core Team 2022) and JASP (JASP team 2022)—also allow the calculation of SDT measures and the output of figures, though doing so requires a fairly extensive knowledge of statistics and basic programming. The deployment of specialist statistical expertise would, for most LAAP agencies, mean either training extant personnel or engaging additional expertise. This would enhance the rigour of the test but sacrifice much of its accessibility.

As specified in subsection 12.3.1, one drawback of SDT is that it is relevant only to tasks involving a yes/no response. This limitation is what prompted the elision of decision and 'unsure' responses in creating Figures 12.2 and 12.3. This is not optimal, considering the ambiguous status of 'uncertain' responses: are they best considered as resulting from competence or from *lack of* competence in the decision task at hand? (cf. discussion in Chapter 5). The exclusive deployment of SDT in interpreting the results of a new perception test would effectively mean limiting the range of responses available to the applicant: only binary decisions could be permitted. This would eliminate the possibility of statistical comparisons of confidence vs. decision responses, independent of accuracy (as was done in Eng-2 and Ar-3) and evaluations of scalar confidence, such as those embedded in Likert scale ratings. It would also

mean allowing the measurement instrument to dictate the design of the test, which seems at the very least counter-intuitive.

The assessment of 'uncertain' vs. 'decision' responses was shown in Ar-3 to yield significant intergroup differences in confidence—in particular, between Syrians and non-Syrians. With this in mind, it may be wise to evaluate separately this dimension of listener perception. Perhaps a meaningful statistical threshold could be established through analysis of response patterns (i.e. ratio of 'uncertain' to 'decision' responses) in tests taken by a reference population, which could be measured against the test results of individual asylum applicants. Again, though, likelihoods of various scores being achieved by native speaker and non-speakers would have to be calculated. Once more, I leave consideration of this possibility to future work.

Whatever the merits and drawbacks of SDT for present purposes, it can at least be claimed that proposing its use as a statistical tool in a new test of perception represents an in-principle advance on current LAAP practice. However, refinement of both test measurement and the design of the tests themselves must for the time being await empirical treatment in future studies.

12.4. Supplementary production test: a broad outline

The second kind of supplementary test this thesis set out to develop involves the applicant's production of language. This test was to be informed by an analysis of perceptual data, derived from Ar-3, of a qualitative type: responses to the question, 'What clues led you to your judgement of the speaker's accent?'

In Chapter 11, I reviewed and quantitatively analysed differing group patterns revealed by responses to this question. I acknowledge that I have insufficient expertise in the Arabic language to examine intergroup patterns of citation of individual features. Rather, a close examination of such patterns, by persistent and properly qualified researchers, was suggested as a possible source of material for an audio text to be used in a supplementary test of production. The envisaged test is of the listen-repeat type prefigured by Moosmüller (2011).

In the remainder of this chapter I suggest a general design for such a test. Firstly, I give a general outline of the listen-repeat tasks proposed by Moosmüller (2011). Secondly, I describe her empirical model of listener behavior (Moosmüller 2010, 2011). Finally, and on the basis of the preceding material, I suggest how a limited selection of individual features cited by listeners in Ar-3 might be used to shape a supplementary listen-repeat task for use in LAAP.

12.4.1. Moosmüller's listen-repeat task

Moosmüller (2011) details five case studies involving four target languages (Viennese German, the Crioulo of Guinea-Bissau, Igbo, and Mandinka). The subjects in each case are what Moosmüller terms 'partly cooperative speakers'—i.e. suspected non-local speakers of these languages, who are equivalent to suspected non-authentic asylum claimants in LAAP.

All perform a task in which they attempt to repeat, in succession, multiple series of audio recordings featuring authentic speakers of the target language/language variety. The example given is of three series of 20 short, presumably non-dyadic, sentences from a mock telephone call.

The repetition procedure occurs several times, at intervals, with the relevant reference recording being played at the commencement of each. Four repetitions of all three series of utterances is suggested as an example. The duration of intervals between repetition series ('rounds') varies widely in the experiments, from five to 45 minutes. In any case, Moosmüller (2011, p. 186) states that the intervals are timed such that the suspected imitator "...will not remember the way he or she has pronounced any given utterance in any of the previous rounds". The guiding assumption, then, is that imitators become less proficient in their imitations as the repetition rounds progress.

Moosmüller finds, on the basis of the results of the above case studies, that partly cooperative listeners (herein 'imitators') exhibit some or all of the following behaviours in the course of performing a listen-repeat task.

"1. Articulatory precision increases with a decrease of attentiveness, also with respect to clear-cut segment boundaries.

2. Application of substitutions takes place when attention is high.

3. Processes and substitutions do not follow phonetically motivated principles.

4. Processes which are not part of the phonology of the language are applied.

5. Variability is unsystematic and inconsistent."

(2011, p. 199)

Some clarification is required here, especially as some of these behaviours run counter to common-sense expectations (e.g. one would predict articulatory precision to increase rather than decrease along with greater attention). Moosmüller explains how each of these behaviours manifests itself in practice.

Firstly, increasingly precise articulation along with decreasing attentiveness occurs as the imitator goes through the repetition rounds. In practice, this means that the imitator "...progressively speaks the way he usually does" (Moosmüller 2011, p. 192).

Secondly, 'application of substitutions' refers to the pronounced tendency of imitators to substitute non-native-like for native-like phones. Moosmüller (2011, pp. 192-3) found that fewer substitutions occurred in the final two rounds of repetitions, when attention was lower, than in the initial two, when it was higher.

Finally, Moosmüller's behaviours 3, 4, and 5—the failure of 'substitutions and processes' to conform to 'phonetically motivated principles', as well as the application of processes not

native to the target language and the irregularity of variation—all refer to related behaviours. Imitators tend to use phonemes either that do not belong to the target language/language variety itself or that would not be used in a particular context by a genuine native speaker. Again, these behaviours occur with increasing frequency in later repetition rounds, when the imitator is likely to forget the substitutions essayed in earlier rounds; they also become less consistent and more arbitrary (Moosmüller 2011, p. 194).

In sum, then, Moosmüller proposes: (1) a general design for a test which could be readily adapted to LAAP; (2) a range of behaviours that might expose imitators in attempting such a test. A further consideration, however, is the composition of suitable audio texts. On this subject, Moosmüller is relatively vague. What kind of audio text might induce in imitators the types of behaviours detailed above? In answering this question, it is helpful to refer to a second contribution by Moosmüller (2010).

Moosmüller (2010) examines authenticity judgements made by native speakers of the Viennese dialect when confronted with non-local (but native German) speakers imitating the dialect. Her finding is that none of the imitators were judged to be authentic. She attributes this fact to imitators' over-generalisation of a single feature: the dark lateral in the Viennese dialect. For sociolinguistic reasons this feature is avoided by local speakers in some contexts, but it is highly salient to non-locals as a stereotypical marker of the Viennese dialect and was thus over-generalised in their (imitated) utterances.

12.4.2. How might a listen-repeat task be designed for LAAP?

Moosmüller's finding was therefore that a *single* salient feature can be used to identify inauthentic speech. This is directly applicable to the development of audio texts for a listen-repeat task in LAAP, based on the qualitative data gathered from the Arabic experiments in this thesis. Individual features cited could be inserted in the audio text in contexts where they would not occur in the speech of genuine natives. The features chosen might be best drawn from citations associated with false rejections of Syrian stimuli by non-Syrian NSNLs, who are

analogous to the kind of applicant who might attempt to imitate Syrian Arabic in LAAP. In this connection, the data gleaned from Ar-3 suggests a number of candidate features for a test with Damascene Arabic as the target. I preface the paragraphs below with the observation that the small numbers of individual feature citations specified (e.g. a mere three citations of /ʔalu:lu/) must be considered against the fact that non-Syrian NSNLs left a total of only 36 comments in association with false rejections.

In three separate instances (cf. Chapter 11, Table 11.10), non-Syrian NSNLs cited /ʔalu:lu/ (*they said*) in incorrectly rejecting three distinct Damascene stimuli (Syria F16, M36 and F50). No other category of listener cited /ʔalu:lu/ in association with a false rejection. Precisely what property of the pronunciation of /ʔalu:lu/ occasioned these citations is unknown. However, it is possible that non-native NSNLs' stereotypical perceptions of the variety were violated by the occurrence of an initial consonant (/ʔ/ instead of /q/) which they did not expect to hear. Note that /q/ and /ʔ/ occur in structured (not free) variation in Damascene Arabic, /q/ mainly in certain MSA 'loans' (e.g. Damascene /qarji/, *countryside*).

In fact, non-Syrian NSNLs cited the word /qarji/ (cf. Chapter 11, Table 11.15 (i)) in incorrectly rejecting two separate Syrian stimuli (Syria F16 and F50). This further implies that stereotypical expectations of the distribution of /ʔ/ were not met, albeit in the opposite direction. There are, of course, other possibilities—e.g. violated expectations of the phonemic shape of the final vowel and/or of suprasegmental properties such as vowel length. Ignoring these alternatives for the present, the suggestion arises that in the audio text /q/ could be widely substituted for /ʔ/ and vice versa. In /qarji/, initial /ʔ/ could be inserted in place of /q/, while initial /q/ could be inserted instead of initial /ʔ/ in/ʔalu:lu/.

If (claimed Syrian) test-takers were directed simply to listen to and repeat the audio text, a range of imitator behaviours identified by Moosmüller might be triggered—but only in non-authentic applicants. For example, the prediction would be that, prompted by the substitution of initial /ʔ/for /q/ in /qarji/, non-natives might 'parrot' the use of initial /q/ in the initial round(s) of the listen-repeat test. Genuine Damascene Arabic native speaker-listeners, meanwhile, *should* be able to recognise initial /ʔ/ as anomalous in /qarji/. Moosmüller's

typology of imitator behaviour further predicts that, in later repetitions of the audio text, non-locals would produce a confusion of forms for the initial vowel—perhaps a mixture of /ʔ/, /q/ and the reflex used in their own dialect—but locals' production of /q/ would remain constant. The former pattern of behaviour would potentially result in imitators being recognised as such.

Based on the data gathered in Ar-3, suprasegmental features may be another candidate for substitution. Non-Syrian NSNL listeners cited speech rate in incorrectly rejecting three separate Syrian stimuli (Syria F 1, M 12 and M 27). Moreover, a frequent subject of commentary by non-natives generally was (their perception) that Syrian speech is typically slow; vowel length was also repeatedly mentioned (cf. Chapter 11, subsection 11.6.3.2). Shetewi, a Syrian native speaker linguist, confirms that Syrian speech is widely stereotyped as slow (pers. comm.), while Almbark & Hellmuth (2015) explain that more than one study has found that vowel length and speech rate are related in Levantine dialects.

Proceeding on this basis, vowels could be synthetically lengthened to degrees impermissible even in relatively slow Damascene speech and inserted in the audio text. According to Moosmüller's typology, the resulting pronunciations would not likely be repeated by true local speakers, but among imitators they might occasion misguided attempts at imitation of the deceptively modelled 'Damascene' vowel lengths. As repetitions proceeded, imitators might revert to vowel lengths consistent with those of their own dialect, or even exaggerate them further. True native speakers of Damascus Arabic would, however, maintain the local norm—i.e. shorter than modelled but still long enough to be recognisably Damascene.

Synthetically lengthened vowels could be fruitfully incorporated into chunks of language—e.g. (semi-)fixed expressions such as Damascene /hado:l mu: mna:ħ/ (*they are not good*). This language chunk was the subject of more citations than any other feature. It was mentioned in 68 separate comments (of a total of 791), more or less evenly distributed among all four listener groups, and overwhelmingly in association with correct acceptances of Syrian stimuli. This indicates that it is a generally well-recognised stereotype of Syrian (or at least Damascene) Arabic. Notwithstanding its evidently high diagnostic value, the occurrence in it of /o:/, /u:/ and /a:/ makes it an obvious potential candidate for the insertion of artificially lengthened vowels.

An alternative to a listen-repeat task of production would be a *second* supplementary test of

perception. This could work as follows. Rather than having the applicant listen to and repeat the audio text in multiple rounds, the applicant would be asked to identify anomalies in the text, in which would occur the same kinds of substitutions/insertions as suggested above. Authentic Syrian applicants (but not non-authentic applicants claiming Syrian background) might be expected to identify, for example, incongruous tokens of /q/ where /ʔ/ would normally occur in Damascene Arabic.

On one hand, this second kind of perception test would have the possible disadvantage that it would draw to some extent on the applicant's command of linguistic metalanguage. This, as the comments reviewed in Chapters 9 to 11 show, cannot be assumed equal in all cases. On the other hand, the proposed second perception test would have the advantage of not requiring that the audio text be played multiple times, making it relatively quick to administer. In addition, it would perhaps engender more easily quantifiable results than the production test. However, to allow adequately rigorous assessment of results, the same difficulties would arise as with the perception task outlined in Section 12.3. Baseline test scores would have to be established, as well as a series of likelihoods for the accomplishment of various scores outside the baseline, for native speakers and non-speakers respectively. In other words, regardless of whether the production test or the second type of perception test was selected for further development, further empirical work of a statistical nature would be required.

The above suggestions on the design of a supplementary production/perception test are advanced both sketchily and provisionally, on the basis of experimental data involving closely-controlled speech in a well-known language variety. For these reasons it may be that the individual feature citation data gleaned from Ar-3 turns out to be less productive than anticipated as a source of audio text material for a real-world test. Alternatively, the data may prove uninterpretable for use in the projected manner. They remain available, however, for examination by researchers in possession of the necessary expertise to attempt such a task (see Appendix D).

12.5. Chapter summary and conclusion

In this chapter I have made preliminary, tentative suggestions as to the design and interpretation of supplementary perception and production tests, the development of which was the ultimate aim of this thesis. In both design and interpretation, such tests would require considerable elaboration in future work. I have expressed some doubt about the use of SDT as a tool of interpretation for the model perception test—in particular, the fact that SDT analysis is limited to quantifying yes-no responses. Neither have I treated in depth the question of establishing statistical likelihoods for various test scores. I have also pointed out that proper analysis of listener comments, which might be used to shape the audio text for the test of production, would require the expertise of a competent native-speaker linguist.

These limitations acknowledged, I conclude this chapter by expressing the hope that researchers with the requisite expertise in statistics and/or (Syrian) Arabic will take up and advance the suggestions offered in this chapter. The composition of workable real-life tests of the projected types is a worthy aim which would enhance the fairness, rigour and transparency of LAAP.

Chapter 13: Conclusions

In this chapter I first summarise the degree of support, adduced from the experimental results described in Chapters 7 to 11, for the hypotheses advanced in Chapter 6 of the thesis. I next resubmit results of the additional investigation of listener confidence in the same chapters, as to which no formal hypothesis was entered in Chapter 6. Following this, I briefly reiterate the possible implications of the preceding findings for the practice of LAAP, drawing on Chapters 9 to 11 and 1 to 3. I then sum up the limitations of the results. Finally, I suggest some overall conclusions and possible directions for empirical work further to that undertaken in this thesis.

13.1. Support for hypotheses

Hypothesis (i):

'Local' speaker-listeners will show greater accuracy in recognising 'local' voice samples in comparison to speaker-listeners of other ('non-local') varieties.

To test this hypothesis, in all five experiments listeners were called upon to respond to the question, 'Is this a [local] accent?' The hypothesis finds support in the results of all five experiments. This is somewhat equivocally so in the purely descriptive statistical analysis of the small (N = 28) pilot experiment described in Chapter 7 (Eng-1), which found that Yorkshire B&R listeners were less accurate on Yorkshire stimuli than Yorks Res listeners. In all other experiments, however, results showed robust support for hypothesis (i).

Of particular relevance to LAAP are the results of Ar-3 (Chapter 11), with Damascene Syrian Arabic as the target. Inferential statistical analysis indicated a significant interaction between Syrian listeners and Syrian stimuli, demonstrating that Syrians were significantly more accurate on Syrian than non-Syrian stimuli and that their performance on this stimulus type was significantly more accurate than that of non-Syrians.

Eng-2 was the largest experiment (N = 197) of the five described in the thesis. As specified in Chapter 8, this study also showed a significant interaction between Yorkshire B&R listeners and Yorkshire stimuli, indicating significantly greater accuracy on the part of locals. As outlined below, however, a significant effect was also returned for linguistic education.

Ar-1 and Ar-2, described in Chapters 9 and 10, were minor studies (N = 10 and N = 12 respectively) in which solely descriptive statistical analysis was employed. Nevertheless, on this basis notably greater accuracy by Syrian listeners on Syrian stimuli was located.

Hypothesis (ii):

The predicted differences in accuracy will hold irrespective of whether speaker-listeners are trained in any branch of linguistics.

This hypothesis is supported by the results of all three experiments in the Arabic series but not by the results of the two English experiments.

The Syrian native-speaker linguists in Ar-1 and the Syrian NSNLs in Ar-2 both achieved greater accuracy on Syrian stimuli than either the non-Syrian linguists in the former or the non-Syrian NSNLs in the latter. In Ar-3, no effect for education was returned.

No equivalent comparison could be made of results across Eng-1 (which included NSNL listeners only) and Eng-2. In Eng-2, linguistic education had a significant effect on overall accuracy on Yorkshire stimuli. However, this may have been attributable, at least in part, to the imbalance in the Non-Brit listener sample (the least accurate group on Yorkshire stimuli), which included non-linguists only.

Hypothesis (iii):

Analysis of feature citations will show that speaker-listeners vary by region of origin in the number and kinds of cues they recognise.

The findings here are complex and, by design, apply only to the Arabic experiments. Nonetheless, they appear to show moderate support for hypothesis (iii).

In Ar-1 and Ar-2 a simple quantitative comparison of Syrians and non-Syrians found that the former were less inclined than the latter to cite features coinciding with those described in the dialectological literature. The same range of interpretations was tendered in response to similar patterns encountered in Ar-3. It was inferred from this that local NSNLs may be especially reliant on tacit, holistic knowledge (cf. Chapter 4). This kind of language knowledge is not concordant with the necessarily segmental bias of dialectological accounts. Neither is it accounted for in the format of LAAP reports, which overwhelmingly prescribe reliance on the dialectological literature.

Further, when features were cited by Syrians in Ar-2—virtually none were mentioned in Ar-1— they rarely proved diagnostic (i.e. they were frequently associated with incorrect responses). Instead, Syrians tended to cite features in association with incorrect acceptances of non-Syrian stimuli (especially those from Jordan, Iraq and Kuwait). This was interpreted as indicating greater knowledge among Syrians of the complex dialectology of Syria. A similar pattern was evident in Ar-3.

Detailed quantitative analysis of feature citations in Ar-3 further showed that all listeners cited features outside the reviewed dialectological literature more often than those mentioned in it. The proportion to which this propensity was present, however, varied among different types of listeners. Syrians generally (along with non-Syrian NSNLs) demonstrated the most pronounced inclination towards this pattern, and non-native linguists the least, though this finding was treated with caution owing to the particularly small number of non-native linguist listeners (N = 5).

13.2. Additional investigation: listener confidence

An investigation of listener confidence was prompted by claims in the LAAP literature that NSNLs are more confident than linguists in their responses to accent identification tasks and that in them accuracy and confidence evidence a poor correlation. The review of relevant empirical findings conducted in Chapter 5 failed to prompt a conclusion as to the validity of these assertions, such as might have permitted the construction of a formal hypothesis. However, the absence of a hypothesis did not preclude an empirical investigation of confidence among the various groups.

In investigating the claim of greater NSNL confidence in LAAP-like tasks, inferential statistical analyses of confidence were conducted on the results of both Eng-2 and Ar-3. Confidence was operationalised in binary fashion—i.e. 'decision' vs. outright 'uncertain' responses on a five-point Likert scale. In neither analysis did education reveal any effect on confidence. Moreover, in Ar-3, Syrian listeners were *generally* more confident (as well as more accurate) on Syrian stimuli than were other listeners.

In interrogating the claim of a poor correlation between accuracy and confidence among NSNLs, a formal analysis of the correlation between accuracy and confidence, via Pearson's r test, was performed on the results of Ar-3. This analysis took comprehensive account of listener ratings on a five-point Likert scale. It showed that confidence and accuracy were strongly and significantly correlated in the responses of *all* NSNL listeners—and still more so among *all* linguist listeners.

These results were together taken as an empirical refutation of the general argument that there is an appreciably different relationship between confidence and accuracy, among NSNL and linguist listeners respectively, in LAAP-like tasks. This rebuttal was interpreted as having implications for the practice of LAAP.

13.3. Implications for the practice of LAAP

The above finding, of no evident relationships between confidence and accuracy contingent on linguistic education, prompted the inference that LAAP agencies are justified in employing NSNL analysts, at least to make judgements as to the linguistic identity of applicants claiming Syrian (Damascene) origins. This inference is reinforced by the aforementioned results showing significantly superior accuracy by Syrians on Syrian stimuli, again with no accompanying effect for linguistic education (cf. section 13.2).

On the basis of these empirically validated proofs of concept, in Chapter 12 I proposed two supplementary tests for LAAP. The first, a test of perception, would involve a similar in-principle task to that of Ar-3 (i.e. listening to target and foil stimuli and responding to the question 'Is this a [X] accent?' or similar), except that the test would be taken by NSNL asylum applicants whose origin is in doubt. Superior accuracy on the part of Syrians in Ar-3, regardless of education, suggests at the very least that true Syrian NSNL applicants would perform better than non-Syrians in such a test. In Chapter 12 I outlined a possible design for this test, which would involve the applicant neither speaking nor being spoken to, thus avoiding the secondary socialisation problem and simultaneously acknowledging and bypassing criticisms of the data drawn from LAAP interviews and their interpretation by NSNL analysts (cf. Chapters 1 to 3).

In Chapter 12 I next sketched two possible designs for a second supplementary test. The first involves production: the applicant would first listen to an audio text and then be asked to repeat it multiple times. I proposed that the audio text could involve anomalous (to true Syrian listeners) substitutions of various phonetic/phonological, morphosyntactic and lexical features. I also submitted an alternative: an additional test of *perception*, in which the applicant would have to verbally identify the anomalous substitutions in the audio text. The individual features essayed as potential targets for substitution were drawn from listener comments in Ar-3. In either version of the test, applicants' differing competence in implicitly or explicitly identifying feature substitutions might help to confirm true Syrians and expose fakers.

In Chapter 11 I suggested that LAAP agencies and the interests of justice may be best served by not relying excessively on features described in the dialectological literature. This tentative

recommendation was based on the general complexion of listener comments drawn from Ar-3 (see 13.2, findings in support of hypothesis (iii)), as well as varying group response patterns to different kinds of stimuli.

In Chapter 2 I reviewed statistics showing that applicants claiming Syrian origins had their claims validated by the LAAP agency Verified at a rate of c. 95%. Verified depends on NSNL judgements, under the supervision of linguists in a team approach. By contrast, over three successive years, claims were confirmed at a much lower rate by the Swiss government agency LINGUA, which does not employ NSNLs. In Chapter 3, I demonstrated that Verified Verified evaluates the evidence for/against a single hypothesis (and not two competing hypotheses) in the initial LAAP verification task. I also showed that this practice is conceptually liable to lead to false acceptances.

As summarised in 13.1, in connection with hypothesis (iii), Syrian NSNL listeners in Ar-1 and Ar-2 evidenced a pronounced tendency towards feature citations in association with false acceptances of Jordanian, Iraqi and Kuwaiti stimuli. Mirroring this pattern, analysis of descriptive statistics, most notably in Ar-3, demonstrated a marked proclivity among Syrians towards false acceptances of these three types of stimuli, especially those featuring Jordanian speakers. This tendency was not present among non-Syrians, who were more inclined to false rejections, as well as false acceptances of Lebanese speakers.

Reinforcing these patterns, the conclusion derived from analysis of inferential statistics in Chapter 11 was that Syrian NSNLs in present-day LAAP are likely to be significantly more competent than non-Syrians (linguists included) in correctly accepting speakers of their own variety, and not significantly less competent in correctly rejecting non-Syrians. However, based on analysis of descriptive statistics from all three Arabic experiments, the additional inference must be that Syrian NSNLs (and Syrian linguists) may falsely accept, at a non-trivial rate, applicants from Jordan (especially), Iraq and Kuwait claiming to be Syrian. Furthermore, it is possible that this tendency is amplified by the practice of testing the evidence for a single hypothesis—i.e. the apparently endemic failure in LAAP to consider the respective strength of the evidence, expressed via likelihood ratios, for two competing hypotheses.

These factors together were advanced in Chapters 3 and 11 as a possible explanation for the very high rate at which applicant claims to Syrian origin were accepted by Verified. The suggestion, also made in Chapters 3 and 11, was that explicit testing of the respective evidence for two hypotheses in turn might be of particular help in reducing the risk of false acceptances in primary-phase LAAP. In Chapter 11 I further proposed that cross-checking of conclusions between Syrian and Jordanian/Iraqi/Kuwaiti NSNL analysts, in cases where a Jordanian, Iraqi or Kuwaiti origin is possible, may assist in reducing the likelihood of false acceptances. I also suggested that the adoption of supplementary tests by LAAP practitioners might be of additional assistance in this connection, providing as they would an extra layer of validation through which initial conclusions could be either confirmed or called into question.

13.4. Limitations

The first general limitation (see, e.g., discussion in Chapter 11) concerns the reliability of results for making inferences as to current LAAP practice. This has to do most obviously with the type of stimuli used in the Arabic experiments: they were read texts in which potential variability of forms was closely controlled. Yet the speech data assessed in real-world LAAP is drawn from the context of an interview, where such control is absent. Furthermore, all target stimuli featured Damascene Arabic, a well-known variety. In real-world LAAP, however, cases in doubt often concern claimed origins in border areas, where lesser-known and lesser-described Arabic varieties might be spoken.

Nonetheless, this limitation applies *only* to the reliability of inferences made as to results here in relationship to current practice. There is no demonstrable reason that stimuli used in *supplementary* tests should not employ read stimuli, with Damascene Arabic as the target. Significant differences between Syrians and non-Syrians in accuracy (and confidence) were located in the results of in Ar-3. It is difficult to imagine that similar results would not obtain in supplementary perception tests using the same types of stimuli, with Syrian/non-Syrian test takers.

A further limitation on the applicability of experimental results to current LAAP practice concerns direct comparison of the respective performance of local NSNLs and non native linguists (cf. Chapter 11). Although this comparison arises frequently in the LAAP literature (cf. Chapters 1 and 2) I have not emphasised it here, for two reasons. The first is that the ultimate aim of the thesis is elsewhere: developing novel supplementary tests. The second is the small number of non-native linguist listeners (N=5) in Ar-3.

A second general type of limitation has to do with the space available in a project of this type, and a third with my own expertise. The limits of space in the thesis, and of my own statistical knowledge, meant that I was unable to elaborate on how results of a supplementary perception test might be measured. Owing to my unfamiliarity with Arabic, I could not proceed any further in the development of a supplementary production (or, alternatively, second perception) test than the rough outline given in Chapter 12. From admission of these limitations naturally follow suggestions for further research.

13.5. Suggestions for future research

Plainly, then, the proposed supplementary tests require further development by parties with relevant expertise. In this connection, opportunities for future research are manifold.

The perception test for Arabic sketched in Chapter 12, and founded on inferential statistical results described in Chapter 11, could be further developed in replicated studies employing stimuli of other languages/linguistic varieties of interest to LAAP (e.g. regional varieties of Kurdish, Persian and Pashto). For all languages investigated, piloting of potential perception tests would have to be conducted with the cooperation of LAAP practitioners (cf. the overview in Chapter 1). Further investigation of the general applicability of SDT as a measurement tool for this test would be necessary, as would the calculation of likelihoods for various response patterns. This would have to be undertaken by researchers with considerable statistical expertise.

A second potential avenue of research involves further development of the proposed test of production—or, alternatively, second test of perception—also outlined in Chapter 12. Further assessment of how individual feature citations interact with stimulus type is especially required. This would crucially involve the expertise of native-speaker linguists with specialist knowledge of the languages/linguistic varieties in question. Here, as in the experiments conducted for this thesis, the focus must be squarely upon what native speakers know about their own language.

Appendices

Appendix A: Task layout in Qualtrics for Eng-1 and Eng-2

What is your residential background in the United Kingdom?

○ 1. Citizen of the United Kingdom and have lived in the country my whole or most of my life

○ 2. Have lived in the United Kingdom for MORE than 12 consecutive months at some stage

○ 3. Have lived in the United Kingdom for UNDER 12 consecutive months at some stage

○ 4. Have never lived in the United Kingdom

Were you born in Yorkshire and/or have you at some stage lived in Yorkshire for more than five years consecutively?

○ Yes

○ No

○ N/A (I'm not from Yorkshire and have never been a long term resident)

If you answered 'Yes' above and are not originally from the place in Yorkshire where you live now, how long have you lived in your present area of residence? (e.g. 'I was born in London but have lived in Leeds since I was five.')

Which part of Yorkshire are you from/have you mainly lived in?

○ North Yorkshire

○ South Yorkshire

○ West Yorkshire

○ The East Riding of Yorkshire

○ N/A (I'm not from Yorkshire and have never been a long term resident)

If you are from the United Kingdom, which part are you originally from? Please tick 'N/A' if you chose alternative 5 for the preceding question.

○ Northeast England

○ Northwest England

○ English East Midlands

○ English West Midlands

○ Southeast England

○ Southwest England

○ London

○ Wales

○ Northern Ireland

○ Scotland

○ Other

○ N/A

In the box below, please enter any additional comments you wish to make about the tasks and/or your responses.

Sample 2: Is this a Yorkshire accent?

jimmyt
2

SOUNDCLOUD

0:29

▶ 62

Cookie policy

○ Yes

○ No

○ Don't know

If your answer above was 'No', what kind of accent do you think the speaker has?

Appendix B: The Mercian-Northumbrian split

The close historical kinship between Yorkshire and Lancashire dialects was an important motivation for their use in both English experiments. However, the degree to which they are distinguishable from each other in the present day caused some concern—a problem amplified by the ever-wider use of GNE in the two counties. There is, in addition, surprisingly little literature on the topic. The theory of the Mercian-Northumbrian split, though, seemed in need of investigation before proceeding with the English experimental series.

Wakelin (1977, p. 102) proposes that an isogloss boundary runs through approximately the middle of Yorkshire from the Humber and into Lancashire, as far west as the estuary of the River Lune, at Lancaster. Varieties found to the south and west of this line are considered descendants of the dialects spoken in the Anglo-Saxon kingdom of Mercia. Those to the north and east are said to be modern successors of the dialects of the kingdom of Northumbria. All of the Lancashire stimuli in this study, and all of the Yorkshire stimuli except two (Harrogate and, less clearly, Skipton) are according to this classification 'Mercian'. The theory of a Mercian-Northumbrian split suggests that listeners would have great difficulty distinguishing between West Yorkshire accents and those of historical (southern) Lancashire, since all belong in the Mercian zone.

Other authorities take a sceptical view. [The Yorkshire Dialect Society](#) points out that the isoglosses defining the split have never been clearly explicated and that the boundary in any case is in flux, in a generally northerly direction. Wakelin (1977) himself states that as of his writing northern features had been ceding territory to 'north midlands' variants since the early years of the 20th century.

The Mercian-Northumbrian split (the Yorkshire Dialect Society)
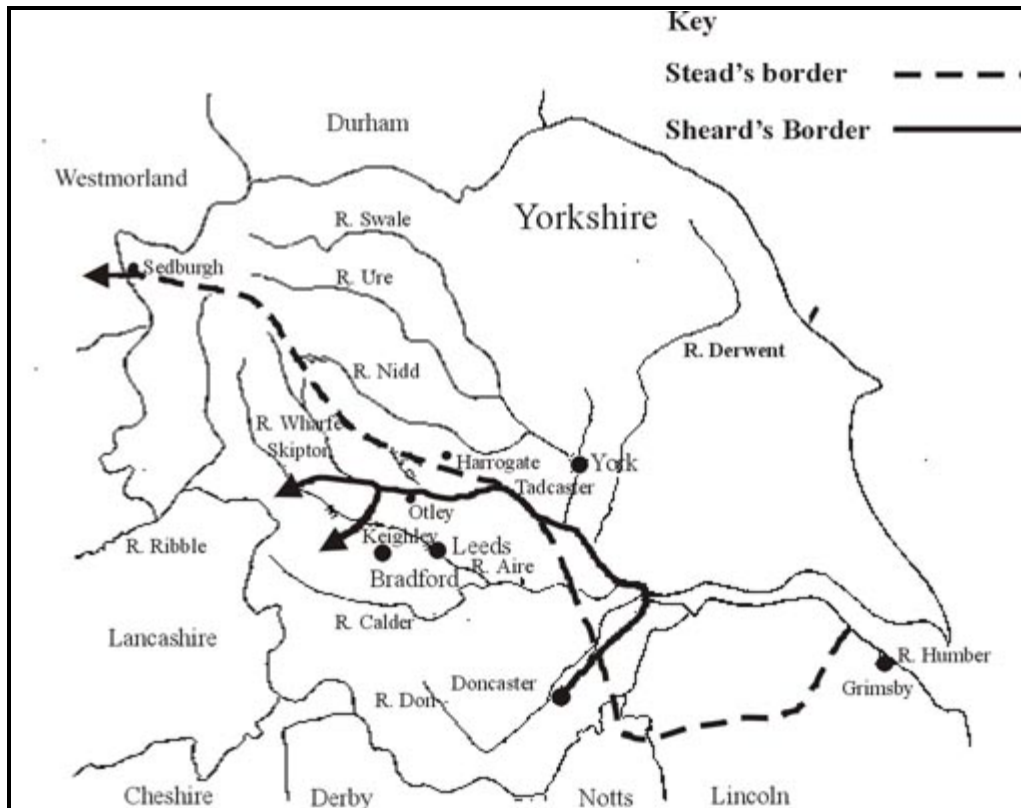
Wells (1982, p. 350) disputes the modern-day relevance of the split, while accepting its validity in historical dialectology. He argues that the generally perceived contemporary boundary between northern and midlands dialects runs between the River Severn and The Wash. Though Wells' objection is weak in the sense that it invokes not linguistic evidence but public perception, there is sufficient doubt about the empirical basis and the continuing vitality of the proposed split that it seemed safe to disregard it for the purposes of the English experimental series. Doing so permitted consideration of Yorkshire/Lancashire dialects—geographically rather than historically defined—as potentially suitable target/foil varieties for the experiments. But the question of how distinct Yorkshire and Lancashire dialects are from one another remained unresolved. The decision was made, then, to proceed in investigating the Yorkshire-Lancashire distinction, allowing the question to be resolved empirically.

Appendix C: Scripts of the *Juha* story, from stimuli used in the Arabic experimental series (Almbark, pers. comm.; Khattab pers. comm.)

Egypt

line 1	ʔasħa:bu ʔalu:lu	اصحابه قالولو	his friends told him
line 2	xal:i ba:lak ja guħa minil baj:aʕi:n butu:ʕ lmadi:na	خلّى بالك يا جحا من البياعين بتوع المدينة	be careful Guha from the sellers in the city
line 3	du:l wiħʃi:n ʔawi	دول وحشين أوي	they are very bad
line 4	ʔaw:il ma yismaʕu lahigtak	أول ما يسمعوا لهجتك	when they hear your accent
line 5	w yiʕrafu ʔinnak minil ʔarja:f	و يعرفوا انك من الارياف	and know that you are from the countryside
line 6	ha ɣallu ʕali:k lħa:ga	هيغلو عليك الحاجه	they will increase the prices for you

Iraq

line 1	ga:lu:la ʔasdiqa:ʔa	قالوله أصدقائه	his friends told him
line 2	di:rba:lak ja dʒuħa minil baj:aʕi:n il:i bilmadi:na	دير بالك يا جحا من البياعين اللي بالمدينة	be careful Juha from the sellers in the city
line 3	ðo:la kul:iʃ mu: ze:ni:n	هذولا كلش مو زينين	they are very bad
line 4	lam:an jismaʕu:n lahidʒtak	لما يسمعون لهجتك	when they hear your accent
line 5	w jiʕirfu:n ʔin:ak min ilqarja	و يعرفون إنك من القرى	and know that you are from the countryside
line 6	raħjɣal:lu:n ʕale:k lʔaʃja:ʔ	رح يغلون عليك الأشياء	they will increase the prices for you

Jordan

line 1	galu:lu sħa:bu	قالولو صحابو	his friends told him
line 2	di:rba:lak ja dʒuħa minil baj:aʕi:n il:i filmadi:ni	دير بالك يا جحا من البياعين اللي في المدينة	be careful Juha from the sellers in the city
line 3	haðo:l miʃ kwajsi:n	هذول مش كويسين	they are very bad
line 4	lam:a jismaʕu lahidʒtak	لما يسمعوا لهجتك	when they hear your accent
line 5	w jiʕrifu ʔin:ak min ilgarje	و يعرفوا انك من القرية	and know that you are from the countryside
line 6	raħ jɣal:lu ʕale:k lʔaʃja:ʔ	رح يغلو عليك الأشياء	they will increase the prices for you

Kuwait

line 1	rabʕa ga:lo:la	ربعه قالوله	his friends told him
line 2	di:r ba:lik ja dʒuħa min il:i bi:ʕu:n filmadi:na	دير بالك يا جحا من اللي يبيعون في المدينة	be careful Juha from the sellers in the city
line 3	haðo:la mu: ze:ni:n kil:iʃ	هذولا مو زينين كلش	they are very bad
line 4	ʔaw:il ma yismaʕu:n lahdʒitik	أول ما يسمعون لهجتك	when they hear your accent
line 5	w jiʕirfu:n ʔin:ik minl ʔarja:f	و يعرفون إنك من الأرياف	and know that you are from the countryside
line 6	raħ jɣal:on ʕale:k lʔaʃja:ʔ	رح يغلون عليك الأشياء	they will increase the prices for you

Lebanon 1*

line 1	ʔalu:lu sħa:bu	قالولو صحابو	his friends told him
line 2	ntibiħe ʒiħa mnil bij:eʕi:n ile bilmadi:ne	رد بالك يا جحا من البياعين اللي في المدينة	be careful Juha from the sellers in the city
line 3	hajdo:la mne:ħ	هاذو ما راهم حلالف	they are very bad
line 4	lam:an jismaʕu lahiʒtak	كيف يسمعوا لهجتك	when they hear your accent
line 5	w jaʕirfu ʔin:ak minil ʔarje	و يعرفوا انك من القرية	and know that you are from the countryside
line 6	raħ jɣal:o ʕle:k liɣra:dˤ	توا يغلو عليك الحاجات	they will increase the prices for you

Lebanon 2*

line 1	ʔalu:lu sħa:bu	قالولو صحابو	his friends told him
line 2	di:r be:lak je:ʒiħa mnil bij:eʕi:n ile bilmadi:ne	رد بالك يا جحا من البياعين اللي في المدينة	be careful Juha from the sellers in the city
line 3	hejdo:l məʃ mne:ħ	هاذو ما راهم حلالف	they are very bad
line 4	lam:an jismaʕu lahiʒtak	كيف يسمعوا لهجتك	when they hear your accent
line 5	w jaʕirfu ʔin:ak mnil qarje	و يعرفوا انك من القرية	and know that you are from the countryside
line 6	raħjɣal:o ʕle:k liɣra:dˤ	توا يغلو عليك الحاجات	they will increase the prices for you

Morocco

line 1	galulu sħabu	قالولو صحابو	his friends told him
line 2	ħade: rasak ja ʒħa min m:aljin limdina	حيضي راسك يا جحى من ڧالين المدينة	be careful Juha from the sellers in the city
line 3	welad lħra:m	ولاد الحرام	they are very bad
line 4	faʃ kajsmaʕu l:ahʒa djal:ak	فاش كيسمعوا اللهجة ديالك	when they hear your accent
line 5	w kajʕarfuk min r:if	وكيعرفوك من الريف	and know that you are from the countryside
line 6	ka ɣal:ju ʕlek biz:af	كيغليو عليك بالزاف	they will increase the prices for you

Oman

line 1	ga:lu:la rabʕa	قالوله ربعه	his friends told him
line 2	xalli ba:lik ja: dʒuħa min tidʒa:r lmadi:na	خلي بالك يا جحا من تجار المدينة	be careful Juha from the sellers in the city
line 3	haðo:la wa:jid xa:jsi:n	هذولا وايد خايسين	they are very bad
line 4	lam:a jisimʕu:n lahdʒatk	لما يسمعون لهجتك	when they hear your accent
line 5	w jiʕirfu ʔinnak min se:ħ	و يعرفوا انك من السيح	and know that you are from the countryside
line 6	baɣal:u:n ʕale:k lʔaʃja:ʔ	بيغلون عليك الأشياء	they will increase the prices for you

Syria

line 1	ʔalu:lu sħa:bu		قالولو صحابو	his friends told him
line 2	di:rba:lak ja ʒuħa minil baj:aʕi:n il: ibilmadi:ni		دير بالك يا جحا من البياعين اللي بالمدينة	be careful Juha from the sellers in the city
line 3	hado:l mu: mna:ħ		هدول مو مناح	they are very bad
line 4	lam:a jismaʕu lahiʒtak		لما يسمعوا لهجتك	when they hear your accent
line 5	w jaʕirfu ʔin:ak min ilqarji		و يعرفوا انك من القرية	and know that you are from the countryside
line 6	raħ jɣal:u ʕale:k lʔaɣra:dˤ		رح يغلو عليك الأغراض	they will increase the prices for you

Tunisia

line 1	qalu:lu sħa:bu		قالولو صحابو	his friends told him
line 2	rid: be:lik ja ʒħa: minl baj:a:ʕi:n il:i filmadi:na		رد بالك يا جحا من البياعين اللي في المدينة	be careful Juha from the sellers in the city
line 3	haðu: ma: ra:hum ħle:lif		هاذو ما راهم حلالف	they are very bad
line 4	ki:f jisimʕu lahiʒtik		كيف يسمعوا لهجتك	when they hear your accent
line 5	w jaʕrfu ʔin:ik minl qarja		و يعرفوا انك من القرية	and know that you are from the countryside
line 6	taw:a jɣal:u ʕale:k lħaʒe:t		توا يغلو عليك الحاجات	they will increase the prices for you

* Lebanese stimuli were not featured in Ar-1; I thank Professor Ghada Khattab for checking my transliterations of these two stimuli.

Original numbering and naming of stimuli in the IVaR database:

Egypt:

37 = egca-sto-f1-part 1-a

34 = egca-sto-m1-part 1-a

28 = egca-sto-f2-part 1-a

18 = egca-sto-m6-part 1-a

13 = egca-sto-m5-part 1-a

2 = egca-sto-f3-part 1-a

Iraq:

49 = irba-sto-f1-part 1-a

45 = irba-sto-f2-part 1-a

42 = irba-sto-m4-part 1-a

33 = irba-sto-m1-part 1-a

29 = irba-sto-m2-part 1-a

3 = irba-sto-f3-part 1-a

Jordan:

35 = joka-sto-f1-part 1-a

22 = joka-sto-m2-part 1-a

19 = joka-sto-m3-part 1-a

14 = joka-sto-f2-part 1-a

9 = joka-sto-m1-part 1-a

4 = joka-sto-f3-part 1-a

Kuwait:

47 = kwur-sto-f1-part 1-a

30 = kwur-sto-m2-part 1-a

23 = kwur-sto-f2-part 1-a

17 = kwur-sto-m4-part 1-a

8 = kwur-sto-f3-part 1-a

5 = kwur-sto-m1-part 1-a

Morocco:

51 = moca-sto-f2-part 1-a

46 = moca-sto-f3-part 1-a

40 = moca-sto-m1-part 1-a

21 = moca-sto-f1-part 1-a

20 = moca-sto-m3-part 1-a

6 = moca-sto-m2-part 1-a

Oman:

41 = ombu-sto-f2-part 1-a

39 = ombu-sto-f4-part 1-a

31 = ombu-sto-m2-part 1-a

25 = ombu-sto-m4-part 1-a

15 = ombu-sto-f6-part 1-a

11 = ombu-sto-m3-part 1-a

Syria:

50 = syda-sto-f1-part 1-a

43 = syda-sto-m3-part 1-a

36 = syda-sto-m2-part 1-a

32 = syda-sto-m4-part 1-a

27 = syda-sto-m5-part 1-a

24 = syda-sto-f5-part 1-a

16 = syda-sto-f2-part 1-a

12 = syda-sto-m1-part 1-a

7 = syda-sto-f3-part 1-a

1 = syda-sto-f4-part 1-a

Tunisia:

52 = tuns-sto-f1-part 1-a

48 = tuns-sto-f3-part 1-a

44 = tuns-sto-m2-part 1-a

38 = tuns-sto-m3-part 1-a

26 = tuns-sto-m4-part 1-a

10 = tuns-sto-f2-part 1-a

Appendix D: Comments from Ar-3

Feb 28, 2021 11:48 PM Saudi 11 (Linguist)

Maghribi dialect, the pronunciation and the special verbal rhythm in it is Moroccan (6 Moroccan highly unlikely) (2 & 0)

Jan 5, 2021 10:22 AM Egyptian 5

The method of pronunciation there is no Ritem (1 Syrian, uncertain)

Lack of rhyme in pronunciation (12 Syrian, uncertain)

Dec 21, 2020 4:47 PM Syrian 28 (linguist) I was born in Syria. I lived in St. Lucia (1987-1994) and the USA (2003-present).

This is Egyptian Arabic, e.g. the use of xalliibaalak instead of dier and lHaaga instead of lHaaja. (2 Egyptian highly unlikely)

This sound like Gulf Arabic, e.g. yGalluun 'leek l?ashyaa? would be said in Syrian Arabic as yGalloo 'leek l?ashyaa. (8 Kuwaiti highly unlikely)

The use of the g sound instead of q could be an indicator that the person is Jordanian or Palestinian. But the speaker could be from Southern Syria. (9 Jordanian likely)

This is Egyptian Arabic, using xalliibaalakyaGoHa...ha-yGalluu 'aleeklHaaga. So the use of g in place of j and the use of lexical terms such as xallii instead of dier. (13 Egyptian highly unlikely)

This sounds like Gulf Arabic, although some Bedouins in Syria may sound like this. (17 Kuwaiti unlikely)

This sounds like Gulf Arabic, but there are some Bedouins in Syria who may sound similar. (22 Jordanian unlikely)

This sounds Syrian, but the use of the verb Ya'irfuun with a full uun instead of oo makes me wonder if this is Palestinian/Jordanian effect or Bedouin effect. (24 Syrian likely)

The dialect sounds like Iraqi dialect, although certain parts of Eastern Syria may sound the same. (29 Iraqi - uncertain)

This dialect sounds like Iraqi dialect, but some parts of Eastern Syria or Bedouins may sound like that too. (42 Iraqi unlikely)

This is Syrian and most likely Damascene; all words are pronounced the way a Damascene person would say them. (50 Syrian highly likely)

This is Lebanese dialect due to the pronunciation of mneeH with an ee instead of aa. The ee is Lebanese than Syrian, although in certain villages in Syria, it is used but there will be other accompanying features. (53 Lebanese highly unlikely)

This is Lebanese because of the pronunciation of mneeh with ee instead of aa and the way dierbaalak is pronounced as dierbeelek. (69 Lebanese highly unlikely)

Some of the dialects I identified as Syrian could also be close to Jordanian and Palestinian Arabic.

Jordanian 4 (linguist) Dec 16, 2020 7:47 PM

A word of your companions (1 Syrian highly likely)

It is an Egyptian accent from the pronunciation of the letter J (2 Egyptian highly unlikely)

This is a Moroccan dialect of "Dialak" and "Bazaf." (6 Moroccan highly unlikely)

Pronunciation of the word "your accent" and the extension in vowels indicate that it is Syrian (7 Syrian highly likely)

Gulf dialect by pronunciation (8 Kuwaiti highly unlikely)

Pronunciation of the letter Qaf and the lack of extension of the vowels (9 Jordanian highly unlikely)

Tide method (12 Syrian highly likely)

The word "brutality" (wiHshi:n) is Egyptian (13 Egyptian highly unlikely)

The word "Mo Manah" (16 Syrian highly likely)

The word "Kalash" is not Syrian but Iraqi (17 Kuwaiti highly unlikely)

The word "" is not good enough (22 Jordanian unlikely)

The term "Mo Manah" denotes that he is Syrian (27 Syrian highly likely)

"The tails of Kalash Mo Zainin" is not Syrian (29 Iraqi highly unlikely)

Who pronounces the word "Juha" and extends the word "your mind." (32 Syrian highly likely)

The method of extending the vowel and pronouncing the letter j (36 Syrian highly likely)

The tide is in the vowels and pronouncing the letter c (43 Syrian highly likely)

The way of pronouncing the letter alif in the word "your mind" indicates that it is Lebanese (69 Lebanese unlikely)

There are words that were used to easily indicate the speaker's accent. If the speech were standardized, the degree of difficulty would be greater

Dec 16, 2020 1:47 PM Jordanian 3 (linguist)

Hedla (1 Syrian unlikely)

Goha, accent, need, c, pronunciations g (2 Egyptian highly unlikely)

North African accent, dialect, Balzaf (6 Moroccan highly unlikely)

rah, lima, hadual (7 Syrian highly likely)

They know, they hear (8 Kuwaiti highly unlikely) (NB referring to yismau3u:n vs yisma3u)

This is, not, good. It is mostly a Jordanian accent (9 Jordanian highly unlikely)

hadual, lah, yali (12 Syrian likely)

Egyptian accent, the pronunciation of gym, such as g (13 Egyptian highly unlikely)

Higloo (ighalu) (16 Syrian likely)

A quarter, they boil ربع، يغلون (17 Kuwait highly unlikely)

North African accent, your dialect (21 Moroccan highly unlikely)

Hadul, a Jordanian dialect (22 Jordanian highly unlikely)

Hadul, Juha, Mo (24 Syrian highly likely)

Hedoul (27 Syrian highly likely)

Kalash (29 Iraq highly unlikely)

Alulua, Yali, Hadul (36 Syrian highly likely)

Iraqi dialect (Kalash Mo Zainin) (42 Iraq highly unlikely)

rah, yali (43 Syrian highly likely)

Hadul, Mo, Alua (50 Syrian highly likely)

Pay attention (break the Baa), Rah, Hdul (53 Lebanese highly likely)

Lebanese dialect, Juha (fractured jim (69 Lebanese highly unlikely)

Dec 16, 2020 10:58 AM Jordanian 2 (linguist)

She is trying to copy the Syrian accent. The dʕ ض sound is heavily pronounced. (1 Syrian highly unlikely)

She could be Syrian but not from Damascus, likely to be Palestinian. (16 Syrian unlikely)

Not the Damascene accent but close to Dara accent or an adjacent country but it is a villager accent. Not Syrian (22 Jordanian highly unlikely)

There is a tase of Syrian sound but could be an Urban accent from Jordan. (24 Syrian likely)

Very close to Syrian though the speaker has a soft sweet sound but lacks the Syrian sound. (27 Syrian uncertain)

I think Lebanese but not from Beruit (32 Syrian highly unlikely)

Palestinian speaker (36 Syrian highly unlikely)

Lebanese or Jordanian Urban speaker (43 Syrian unlikely)

I think Palestinian (69 Lebanese highly unlikely)

Dec 13, 2020 4:53 PM Egyptian 4

Moroccan, because I did not understand it, and because of the word Bazaf (6 Moroccan highly unlikely)

Because the word Kalash is Tunisian, almost not Syrian (8 Kuwaiti highly unlikely)

Payne is the Lebanese accent, because it has Imaala (24 Syrian uncertain)

Moroccan dialect because the (s) are clear (29 Iraqi highly unlikely)

Some Sham(e)? words (36 Syrian highly likely)

69 Egyptian (69 Lebanese highly unlikely)

Dec 13, 2020 3:58 PM Palestinian 9 (linguist)

Syrians say they are friends, not friends (1 Syrian unlikely)

The Syrians do not speak or speak a lot, and the letter Qaf is spoken by Alif, not q .. The dialect is clear that it is Egyptian (2 Egyptian highly unlikely)

alsuwriiyn ma biahikuubialzayf .. bas altawanisat aw almagharibatbiahkuubialzayf (6 Moroccan highly unlikely)

The language is Gulf. The people of the Gulf are going to Zainin. The Syrians are talking about Manah (8 Kuwaiti highly unlikely)

The Syrians do not talk, they are not good .. they say it is not possible (9 Jordanian highly unlikely)

Mo Manah .. evidence that they are Syrians (12 Syrian likely)

The Egyptians use countries in their speech instead of Hadul ... and awy instead of a lot ... and a monster instead of a man almisriiynbiustakhdimuu fi kalamihim dual badalhadwal .. w 'uwybadalktir ..wahshinbadal mw manah (13 Egyptian unlikely)

The word Mounah is like the words of the Syrians (16 Syrian likely)

The dialect is clear, different and very close to the Gulf .. The Syrians speak with the Alif, not q (17 Kuwaiti highly unlikely).

Syrians use the word good, they say good (22 Jordanian unlikely)

The dialects of Bilad al-Sham are very close to each other, especially the Syrian and Lebanese dialects. It is difficult to differentiate their details (27 Syrian uncertain)

Syrians do not go to Zainin ... they say Mnah (29 Irqai highly unlikely)

The word "Mawah" and "cold" may be Lebanese, Palestinian, or Jordanian .. The Levant is very close to their dialects (32 Syrian likely)

Allah is close to the Syrian, especially if he talked to Mo Manah (36 Syrian uncertain)

Syrians don't talk to Zainin, they talk about Manah instead (42 Iraqi highly unlikely)

It is very clear that it is Syria .. Biju Mu Manah (50 Syria highly likely)

Mnah does not tell the way it is told in Syrian (69 Lebanese uncertain)

The countries of the people of the Levant have many dialects close to each other, especially Syria, Lebanon and northern Palestine, and many words differ from one country to another. The Egyptians use it a lot, and likewise the people of Palestine who are in the south use it too ... while the people of Levant are Syrians and Lebanese, and some areas in Jordan and northern Palestine quote the word "Good"

Syrian 21 Dec 12, 2020 11:46 PM (identified all Syrians as being from Damascus) I was born in Damascus, I lived in Cyprus for 3 years from 2007 to 2011, then I lived in Britain for 10 years from 2011 until now

The dialect of the Maghreb Balzaf (6 Moroccan highly unlikely)

Daraa (Jordanian 9 likely)

The dialect of the people of Daraa (22 Jordanian highly likely)

Palestinian 7 Dec 12, 2020 12:53 PM

An Egyptian accent, clearly, because of the substitution of the gram with the g, as in the English language (2 Egyptian highly unlikely)

It does not at all resemble the Syrian dialect, and it is a Gulf dialect likely to be Kuwaiti because of the nun in the last verbs (NB referring to yismau3:n vs yisma3u), the use of the classical Arabic gem (i.e. jim) and the word Kalash (8 Kuwaiti highly unlikely)

The word Mu calm in hadith does not resemble other dialects and is closer to the Syrian accent (12 Syrian highly likely)

Egyptian and perhaps clearer than the previous voice, the Arabic speaker has the ability to define the Egyptian hue for several reasons, such as the letter g, for example (13 Egyptian highly unlikely)

It is difficult to explicitly describe the reasons for this Syrian accent (16 Syrian highly likely)

The word Balzaf is used in Algeria, not in Syria (21 Moroccan highly unlikely)

It is the same as the kasrah at the end of the words with a rather long length of speech (36 Syrian highly likely)

The word Kalash (42 Iraqi highly unlikely)

The way to break the last letter in addition to the complete sentence is complete and clear that it is Syrian (50 Syrian highly likely)

Lebanese dialect due to the use of the word village by replacing the qaf with a sound (53 Lebanese highly unlikely)

I am not sure about some of the answers, so I did not give reasons, whether the dialect was Syrian or not, as I think that the Arab can differentiate the Syrian dialect in particular without giving a clear reason. Our Lord, this dialect is similar to the accent of a Shami series, or that is similar to the accent of a Syrian series, meaning that the listeners 'Arabic ears are accustomed to it. He can give clear reasons. Another thing I think is that there is an Egyptian accent in the voices. I think the first voice does not resemble the usual Egyptian accent. Perhaps the voice of the voice lives outside Egypt or it is not Egyptian in the first place. May I have miscalculated, but it is not like the smoothness of the usual Egyptian dialect. The questionnaire is very interesting, even if I have time. The more reasons I have been given the more success.

Palestinian 6 Dec 12, 2020 12:07 PM

Dialect from Lebanon (1 Syrian uncertain)

The dialect is Egyptian and its speed is an Egyptian accent (2 Egyptian highly unlikely)

The dialect is Moroccan, from the word Bazaf, and Dial is a dialect from Tunisia, Morocco and Algeria (6 Moroccan highly unlikely)

A Palestinian Jordanian dialect (7 Syrian unlikely)

Iraqi dialect of Klaj (i.e. Gulf) (8 Kuwaiti highly unlikely)

Dialect from Palestine in Gaza (9 Jordanian unlikely)

The dialect is Palestinian from Gaza (12 Syrian unlikely)

From the Levant (16 Syrian unlikely)

The dialect is from Gaza (22 Jordanian unlikely)

The dialect is close to Lebanese, not Syrian (22 Syrian unlikely)

An Iraqi language from Kilg Mo Zain (kullish mu: ze:ni:n) (42 Iraqi highly unlikely)

Shamiyah dialect, but it is not Syrian, of course (43 Syrian uncertain)

The Lebanese dialect is the same as the word village (2arye) (53 Lebanese unlikely)

The dialect is close to the Palestinian dialect (69 Lebanese unlikely)

In most of the dialects, most of them are similar, in Baladr al-Sham, but differences in regions and dialects in the Gulf countries are similar, and so is Morocco. One city can speak more than one.

Palestinian 5 (linguist) Dec 12, 2020 11:48 AM

Both the way they are and the tone of their voice (8 Kuwaiti highly unlikely)

Voice tone and use of letters (50 Syrian likely)

Jordanian 1 (linguist) Dec 12, 2020 11:41 AM

The Syrian dialect uses the prolonged pronunciation (24 Syrian highly likely)

Palestinian 4 (linguist) Dec 12, 2020 11:08 AM

Jordanian (1 Syrian uncertain)

Jordanian (7 Syrian highly unlikely)

How to pronounce words (16 Syrian likely)

How to pronounce words (24 Syrian likely)

Palestinian 3 (linguist) Dec 12, 2020 11:07 AM

They said (2alu:lu) (1 Syrian likely)

Juha (guHa) (2 Egyptian highly unlikely)

Many words like Bzaf (6 Moroccan highly unlikely)

Speech speed and mastery of the text (7 Syrian highly likely)

Not good (i.e. mish kwaysi:n) (9 Jordanian unlikely)

He told him, take care of you (2alu:lu… di:rba:lak) (12 Syrian likely)

Juha (guHa) (13 Egyptian unlikely)

The village, they said (16 Syrian uncertain)

Deira (17 Syrian highly unlikely)

Bzaf (21 Moroccan highly unlikely)

Hdhul village (hadho:l, garye) (22 Jordanian unlikely)

Hadol, City (24 Syrian unlikely)

 the village (qaryi) (27 Syrian likely)

Kalash (29 Iraqi highly unlikely)

His companions, they told him (2alu:lu SHa:bu) (36 Syrian uncertain)

Your friends (2aSdiqa:2a) (42 Iraqi highly unlikely)

Juha (43 Syrian likely)

Mo Manah (50 Syrian likely)

Looks Lebanese, from the city (IMALA) (53 Lebanese unlikely)

Juha, Hedol (jiHa, haydo:l (69 Lebanese unlikely)

The Syrian dialect has a great affinity with the dialects of neighboring countries such as Palestine and Jordan. Therefore, in some passages, it seems difficult to know whether this dialect is Syrian or Levantine (Palestine, Jordan in particular)

Dec 12, 2020 11:02 AM Palestinian 2 (linguist)

Pronunciation (32 Syrian, highly likely)

Hydol (69 Lebanese, unlikely)

Dec 12, 2020 10:26 AM Palestinian 1 (linguist)

Vocabulary and pronunciation (1 Syrian, highly likely)

Words from the Maghreb (6 Moroccan, highly unlikely)

Vocabulary and pronunciation (7 Syrian, highly likely)

Vocabulary (9 Jordanian unlikely)

Pronunciation (12 Syrian, likely)

It looks Palestinian (16 Syrian, unlikely)

Pronunciation and vocabulary (24 Syrian, likely)

Some words and how to spell them (29 Iraqi, likely)

Lack of smooth talk (32 Syrian, uncertain)

Some vocabulary such as Manu Mallah (36 Syrian, likely)

Pronunciation and vocabulary (43 Syrian, likely)

Pronunciation (50 Syrian, likely)

How to pronounce it to the village (53 Lebanese uncertain)

Lebanese word of breaking (69 Lebanese unlikely)

Dec 12, 2020 9:39 AM Syrian 20

Hashin, the need is used in Egypt (2 Egyptian, unlikely)

Bzaf (6 Moroccan unlikely)

Iraqi dialect, the word Kalash (8 Kuwaiti unlikely)

The Egyptian dialect of the gym pronounces ج (13 Egyptian unlikely)

Mo Manah (50 Syrian highly likely)

Dec 12, 2020 8:57 AM Syrian/Saudi 19

Iraqi. Pronouncing the words is different than Syrian (Kuwaiti 17 unlikely)

The sound and the way they pronounced some words helped me to recognise that it's not Syrian dialect. There are some letters in particular been pronounced differently in other Arab countries than Syria such as ج ق . There are some words that have been used in the recordings helped me to know the origin of this dialects, Morrocan, Iraqi, Lebanese, Egyptian and Algerian. Finally, I found it difficult to recognise whether the dialect is Syrian (rural dialect) or Jordanian.

Dec 12, 2020 12:10 AM Saudi 9 (linguist)

Words like Deir Balk (Syrian 50 highly likely)

From some words .. Do not go and take care of your mind .. (Lebanese 53 likely)

Dec 11, 2020 10:16 PM Syrian 17 (linguist)

One of the governorates of Syria speaks the same dialect, which is Daraa (9 Jordanian uncertain)

Egyptian accent, they change the letter J with their accent (13 Egyptian highly unlikely)

Moroccan use Balzaf (21 Moroccan highly unlikely)

As one of the governorates of Syria speak the same way (22 Jordanian uncertain)

The first voice of a Syrian speaker who uses the term "Hadul" (27 Syrian highly likely)

Accent, correct pronunciation and use of headl (32 Syrian highly likely)

This Iraqi dialect is where they use the word (Kalash) (42 Iraq highly unlikely)

A Lebanese accent, not a Syrian one, as the Lebanese break the letter alif in their manner of speaking. (Lebanese 69 highly unlikely)

Dec 11, 2020 9:47 PM Syrian 16

Levantine dialect (7 Syrian highly likely)

The dialect of the people of Daraa appears (9 Jordanian likely)

Dec 11, 2020 9:29 PM Syrian 15 (linguist)

This sounds Jordanian but I am not sure if people living at the border speak this accent (22 Jordanian uncertain)

Dec 11, 2020 8:34 PM Syrian 14 (linguist)

Possibly a Jordanian or Palestinian accent. The accent, and how to use terms such as "not" (9 Jordanian unlikely)

Possible Iraqi or Syrian from other cities (17 Kuwaiti uncertain)

It is possible to have a Syrian dialect without the cities of Daraa or Deir Ezzor (42 Iraqi likely)

Possible from one of the Syrian villages or cities (53 Lebanese likely)

Lebanese accent tilting the winds (69 Lebanese unlikely)

Sorry, I could not provide evidence, but any Syrian did not intend to know the dialect from the first two words by nature

Dec 11, 2020 8:22 PM Syrian 13 (linguist)

The word Rafqatu and "Alolo". And others indicate that it is Syria (1 Syrian highly likely)

The dialect is purely Egyptian, the word rural instead of village, the pronunciation of the letter Jm, and the word havai instead of objects (2 Egyptian highly unlikely)

Maghrebi dialect, too, whose words are not understood by the Syrians in the first place (Morocco 6 highly unlikely)

Pronunciation of Syrian letters and words too (7 Syrian highly likely)

The dialect is mostly Emirati and some words were very clear and not used in Syria, such as "Rab'a" and "Kalash" (8 Kuwaiti highly unlikely)

Mostly Jordanian, and she may have come from Houran in Syria (9 Jordanian unlikely)

Syria, too, like Dayr Balik and Moumnah, and the pronunciation of the letters (12 Syria likely)

Egyptian dialect words such as "wahshin", "ha-ha-ha," and the manner of pronouncing the letters (13 Egypt highly unlikely)

Syrian dialect of the words of Ghurad, Yili, Higlo, and the manner of pronouncing letters and words (16 Syria highly likely)

"Rubak" is not Syrian words and terms at all in a Gulf dialect, such as "Mo Zainin Kalash" (17 Kuwaiti highly unlikely)

A Maghreb accent, not a Syrian at all, and most of the words are incomprehensible to the Syrians (21 Moroccan highly unlikely)

Most Jordanian, the pronunciation of the letters is heavy, and the word "Hadhul" is used by Jordanians, and it may be from Houran as well (22 Jordanian unlikely)

The method of stretching some letters and some words and terms, such as the word Yili and take care of your mind and break the gym (Syrian 24 likely)

There are also Syrian terms such as "Deir Balk" and "Mo Manah (Syrian 32 likely)

A Kuwaiti Gulf dialect is often "Kalash Mo Zainin". It is not said in the Syrian dialect, and the manner of pronouncing the sheep is completely different. (Iraqi 29 highly unlikely)

Also, most of the words were spoken in a Syrian accent, but this person was a little closer to classical Arabic, but the Syrian accent is still clear (32 Syrian likely)

It is clear that it is Syria from words, terminology and pronunciation of letters (36 Syrian highly likely)

Mostly an Iraqi dialect, Kalash Mo Zainin, is not Syrian (Iraqi 42 highly unlikely)

Words like objects instead of things and willy instead of that and the pronunciation of Juha to break the gym and Sahabo instead of his friends. And the method of pronouncing and stretching letters in general (43 Syrian highly likely)

The way letters and many words are stretched from the Syrian dialect as well (50 Syrian highly likely)

A Lebanese accent, often like the word "attention, not take care" (Lebanese 53 unlikely)

A Lebanese dialect that is slightly different from the Syrian dialect, but an apparent difference (Lebanese 69 unlikely)

Dec 11, 2020 8:10 PM Syrian 12

Hashin is an Egyptian word (2 Egyptian highly unlikely)

Bazaf is a non-Syrian word (6 Moroccan highly unlikely)

Kalash Iraqi (17 Kuwaiti highly unlikely)

Bzaf (21 Moroccan highly unlikely)

The word Kalash (42 Iraqi highly unlikely)

Haidol Lebanese (69 Lebanese unlikely)

Dec 11, 2020 7:43 PM Syrian 5, United Arab Emirates 1984-2003 (linguist)

Moroccan dialect: Dial-Bazaf (6 Moroccan highly unlikely)

The Emirati or Gulf dialect: Mo Zainin – Kalash (8 Kuwaiti highly unlikely)

Iraqi dialect: Kalash (17 Kuwaiti highly unlikely)

The dialect is Moroccan: Bazaf-Diyalak (21 Moroccan highly unlikely)

The Jordanian dialect is Hadhul-Rahe-Al-Qarya (IMALA)-not good (22 Jordanian highly unlikely)

Possible dialect from Deir Ezzor or Fenitra (42 Iraqi likely)

The Lebanese dialect: Hydol - Mesh Menah – Jeha (69 Lebanese highly unlikely)

Some areas of Syria may have similar dialects, for example, the dialect of Deir Ezzor and Quneitra may resemble the Iraqi accent a little, and the dialect of Banias, Tartous and the coastal villages of Syria may be similar to the dialects of Tripoli in Lebanon, for example.

Dec 11, 2020 7:43 PM Egyptian 3 (linguist)

Moroccan dialect Lexical items (Morocco 21)

 The pronunciation of some words like the last word in the sentence (36 Syrian highly likely)

Dec 11, 2020 6:17 PM Lebanese 2

The Egyptian dialect is the word "J" in Juha (2 Egyptian highly unlikely)

The word Qaf in his speech (6 Moroccan highly unlikely)

Pronunciation (7 Syrian highly likely)

The Iraqi dialect is similar, especially to Kalash (8 Kuwaiti highly unlikely)

The word Qaf in the word village (9 Jordanian highly unlikely)

The word Hadol is repeated (16 Syrian likely)

The word Kalash (17 Kuwaiti highly unlikely)

Al-Qaf in his speech is similar to the Bedouin dialect (21 Moroccan highly unlikely)

Elongating the words (24 Syrian uncertain)

The word Qaf in his speech (29 Iraqi highly unlikely)

The stretched out slot on the gym in the word your accent (36 Syrian highly likely)

Mo Zainin's word (42 Iraqi highly unlikely)

Kasrah is on the gym in Juha, and the letter qaf is not pronounced in the village (53 Lebanese highly unlikely)

Take care and the word Manah, broke the thousand instead of pronouncing it (69 Lebanese highly unlikely)

Dec 11, 2020 6:14 PM Syrian 10

Bzaf (Moroccan 6 highly unlikely)

Kalash Mo Zainin (29 Iraqi highly unlikely)

The word Kalash (42 Iraqi highly unlikely)

Dec 11, 2020 6:04 PM Syrian 9 Syria from 1988-2013 Spain 2013-2015 Britain 2015-2020

Your name is Balzaf as you know (6 Moroccan highly unlikely)

Not (9 Jordanian unlikely)

Deira Kalash (17 Kuwaiti highly unlikely)

Hedol Mo Manah (24 Syrian highly likely)

Mo Zina (i.e. mu: ze:ni:n) (29 Iraqi highly unlikely)

Mo Zainin (42 Iraqi highly unlikely)

No Manah Hedol (43 Syrian highly likely)

Headul Belk (53 Lebanese highly unlikely)

Dec 11, 2020 6:03 PM Syrian 8 I was born in Damascus and I now live in the UK

It is clear that the Egyptian dialect, the Syrians, pronounces the gym clearly, while the Egyptians pronounce it as g in English (2 Egyptian highly unlikely)

Syrians do not say Kalash Mo Zeinin, and it is clear that the dialect is Iraqi (29 Iraqi highly unlikely)

Dec 11, 2020 6:01 PM unknown 1

Damascene dialect (1 Syrian highly likely)

Egyptian dialect .. the word countries .. need (Egyptian 2 highly unlikely)

The word Mo Manah (7 Syrian highly likely)

rbeh..kilsh (8 Kuwaiti highly unlikely)

Not good (9 Jordanian unlikely)

Moomnah word (12 Syrian highly likely)

kalimatalgharad..dimishaq (16 Syrian highly likely)

The word Yallak..Balzaf (21 Moroccan highly unlikely)

Mo Manah (27 Syrian highly likely)

Mo Zainin..Iraqi dialect, perhaps (29 Iraqi unlikely)

The word "Bayain" .. (53 Lebanese unlikely)

Dec 11, 2020 5:31 PM Syrian 7 I was born in Syria and lived there for 22 years. I moved to Britain seven years ago

It may be Syrian, but in the clip it looks like someone who feigns the Syrian accent (24 Syrian likely)

I wish you success and achieve the desired benefit from the study and the referendum, but I have a note which is that the people of major cities in Syria do not know the dialects of the people of the villages or the Syrian Badia, for example .. In several audio clips, I was not sure if the speaker was a Jordanian or an Hourani, for example, or if he was an Iraqi or who Al Jazeera Thank you ...

Dec 11, 2020 5:06 PM American 5 I was born in the United States and have lived in Jordan for four months and in Morocco for more than five years

- The tilt of the marbouta - the name of the sign of Hedul - Mo for the negation - the adjective of Manah (1 Syrian likely)

She is Egyptian: Btawa, Hashin, Strong (2 Egyptian highly unlikely)

Qaf is like the Egyptian gym - Kalash meaning very - the adjective of Zenin - lack of the present tense (8 Kuwaiti unlikely)

From Jordan or a Syrian Bedouin (9 Jordanian unlikely)

He is Egyptian: a monster, for the future (13 Egyptian highly unlikely)

She might be Palestinian, she has the tilt of the tilted Ta, but she uses H for the future like Egyptians (16 Syrian unlikely)

Qaf is like the Egyptian gym - Kalash with a very meaning - the adjective of Zenin - lack of present tense - a thing instead of objects (17 Kuwaiti uncertain)

Dial, bazaf, Moroccan pronunciation (21 Moroccan highly unlikely)

From Jordan or a Syrian Bedouin (22 Jordanian unlikely)

Pronouncing al-Qāf as al-hamza - inclination of the bound taa - the name of the sign of Hadhoul - not of negation - adjective of Manah (27 Syrian likely)

The pronunciation of the qaf as the hamza - the inclination of the bound ta - the name of the sign of Hedul - Mo for the negation - the adjective of Manah (32 Syrian likely)

Pronouncing al-Qāf as al-hamza - inclination of the bound taa - the name of the sign of Hadhoul - not of negation - adjective of Manah (36 Syrian likely)

- Pronouncing the qaf as the hamza - the inclination of the bound ta - the name of the sign of Hedul - Mo for the negation - the adjective of Manah (43 Syrian likely)

You can be from Lebanon because of the tilt of a thousand in directions (53 Lebanon unlikely)

You may be Lebanese or from western Syria, I hear the word Hadul tilted (69 Lebanon uncertain)

Dec 11, 2020 2:56 PM American 4 I was born in the United States and lived in Egypt for three months in 2009 and one year between 2012 and 2013

The pronunciation of the qaf is a hamza, the pronoun becomes "ah" o, "rah" for the future (1 Syrian likely)

The pronunciation of the gym "g", Egyptian vocabulary such as "btoa" and "dole", the Egyptian accent (2 Egyptian highly unlikely)

Pronouncing the qāf "g", using Moroccan vocabulary such as "rask", "fash", "bezaf" (6 Moroccan highly unlikely)

"Hedol" sounds like a non-Syrian to me, but I'm not sure (7 Syrian uncertain)

The pronunciation of the qaf "g", the pronunciation of zaalisnania, the pronunciation of the jim "djima", the plural noon, the use of the pronoun "-k" for the masculine (8 Kuwaiti highly unlikely)

Pronounce "g", the negation with "mish" (9 Jordanian unlikely)

The pronunciation of the qaf is hamza, "Rah" for the future (12 Syrian likely)

The pronunciation of the gym "g", Egyptian vocabulary such as "btoa", "dole" and "wahshin", the Egyptian noun (13 Egyptian highly unlikely)

The gym point "Djima", the pronunciation of the rhyme is a true rhyme in "village", glorifying a thousand (16 Syrian unlikely)

The pronouncement of "g", the pronunciation of zaalisnania, the conjugation of the verb in the plural (17 Kuwaiti highly unlikely)

The pronunciation of ta`amarbouta, the use of Moroccan vocabulary such as "dial" and "bezaf" (21 Moroccan highly unlikely)

The pronunciation of the qaf "g", the pronouncement of zaalisnania, the negation with the "mush" (22 Jordanian highly unlikely)

The tilt of the bound taa, the use of "yalli" instead of "who", the future verb with "reh", supplying the last syllable of the word ("Lahjtaak") (24 Syrian highly likely)

Tilted Ta-tied, "yali", "your accent" (27 Syrian highly likely)

The pronunciation of qāf "g", the pronunciation of zaalisnania, the verb conjugation of the plural "nun" (29 Iraqi highly unlikely)

The inclination of the bound ta, the conjugation of the future verb with "Rah", the supplying of the last syllable of the word (for example, with "Lahjtaak") (32 Syrian highly likely)

Tilted Ta 'Marbouta, the supply of the last syllables ("Juha", "Lahjtaak") (36 Syrian likely)

Pronouncing qāf "g", the conjugation of the verb in the plural nun (42 Iraqi highly unlikely)

The pronunciation of the qaf is a hamza, the inclination of the ta`amarbouta, "rah" for the future, "the hijtak" (and the characteristic that distinguishes the last syllable may be something related to the melody and not just the length) (43 Syrian highly likely)

Tilt the bound, "Rah" for the future (50 Syrian likely)

Tilting a thousand in "Manah", using "for whom" instead of "for what" (53 Lebanese unlikely)

The alif tilt, the bound y pronunciations of the ja, "hedol" (69 Lebanese highly unlikely)

Dec 11, 2020 2:48 PM Syrian 6

Word endings and lettering (1 Syrian highly likely)

The Egyptian dialect is due to the Egyptian pronunciation of the gym and Egyptian vocabulary such as Khali Balk, Al Hajjah, Al Rif (2 Egyptian highly unlikely)

An Algerian accent due to the use of words such as Dialak, Bazaf and Letters (6 Moroccan highly unlikely)

Lettering (7 Syrian highly likely)

The Gulf dialect is due to the use of the words Raba'ah, Kalash, Zainin, Hathol, in addition to the letters of accentuation. (8 Kuwaiti highly unlikely)

It may be from Daraa, because the dialect is close to the Jordanian dialect, like the use of the vomited qaf (9 Jordanian likely)

Word endings and lettering (12 Syrian highly likely)

The Egyptian dialect is due to the pronunciation of the Egyptian jim, and the manner of pronouncing the letters has a special tone in the Egyptian dialect (13 Egyptian highly unlikely)

Letter encryption and word endings (16 Syrian highly likely)

The dialect is Gulf. The vocabulary and pronunciation are not Syrian (17 Kuwaiti highly unlikely)

The dialect is Algerian because of the use of the word bezaf and the pronunciation of letters (21 Moroccan highly unlikely)

The dialect is close to the dialect of the people of Al-Jazirah, but it is similar to the Iraqi dialect as well, with exaggeration of the qaf and the zal (22 Jordanian uncertain)

Structure of the thousand and all open letters (24 Syrian highly likely)

Endings of words and the use of the word mu (27 Syrian highly likely)

The dialect is Iraqi due to the Iraqi vocabulary as Kalash, Zainin, and Takhfil (29 Iraqi highly unlikely)

Word endings and lettering (32 Syrian highly likely)

Word endings and lettering (36 Syrian highly likely)

The dialect is Iraqi due to the use of the words Kalash, Zainin, and the pronunciation of the capitalized letters (42 Iraqi highly unlikely)

Word endings and lettering (43 Syrian highly likely)

Word endings and lettering (50 Syrian highly likely)

The word 'Purpose' (l2aghra:D) and the way to pronounce the letter Qaf, in addition to the accentuation of the letter Alif, like the people of Aleppo (53 Lebanese highly likely)

The dialect is closer to Lebanese or the Syrian villages near Lebanon. The word mesh and the tilt of the alif and the pronunciation of the letter qaf (69 Lebanese likely)

Syria is vast .. the part close to Lebanon is influenced by the Lebanese dialect, and the part close to Iraq is influenced by the dialect of the people of Iraq, and the same is the case in the part near Jordan. They were close together

Dec 11, 2020 2:45 PM American 3 American nationality I lived in Egypt 1980-1983 I traveled between Morocco, Syria, Egypt and Kuwait 1988-1989 Short visits to Lebanon 1995-2017 every year or two 3 short visits to Morocco Residence in Lebanon since early 2018

We will add to the aforementioned features the word "Rafikat" (1 Syrian highly likely)

This is a clear Moroccan accent in terms of phoneme, lexical and accent (6 Moroccan highly unlikely)

Although it may be a rural Syrian dialect from the Essaouira point of view, I doubt that and think that it is a Palestinian dialect that is not good (i.e. moo manah etc.) (9 Jordanian highly unlikely)

An Egyptian accent is quite clear, phonetically, accent, and a landmark (13 Egyptian highly unlikely)

This dialect is close to the previous one, but it may be closer to a Levantine accent because of the voices of the fathah and the thousand mounah. (16 Syrian highly likely)

The dialect in general is Iraqi / Kuwaiti in terms of phonetics and lexical. A quarter of the people said to him, they hear the word "Deira" Worded in an Iraqi / Kuwaiti aspects of audio and Lexical quarter they said to him Kalash hear Deira (17 Kuwaiti highly unlikely)

It doesn't indicate a Palestinian accent, perhaps it is Jordanian (it has the same characteristic in the previous sample, and it may also be Jordanian-Palestinian) (22 Jordanian unlikely)

We return to Kalash Mo Zainin = Iraqi, unless Syria is close to Iraq (29 Iraqi unlikely)

Iraqi, it seems from the phoneme and the dictionary, Kalash Mo Zainin, unless it is a dialect from the far east of Syria near the borders of Iraq, not sure of the tilt (= IMALA) in the word Medina (42 Iraqi unlikely)

Very close to being Syria, but some features favor Lebanon: 2arya is for the village. Be careful - there is no way of negation with what instead of for mo (53 Lebanese unlikely)

Difficult to distinguish some of the Syrian dialects from the dialects of neighboring countries to convergence and at the same time, there is a large inside the Syrian border, the diversity is certainly not the tone of Damascus because of the sounds slot and a thousand and stress in general Syria features: the lexicon and sounds, for example, the word village points to Syria because the lexicon Syrians approaching eloquent sometimes tilt Spread in Syria and neighboring countries (69 Lebanese likely)

When the two phrases were repeated: Not good, and Kalash not Zainin, I began to doubt that the participants were reading from a text that might not reflect their accent. Perhaps the goal of the research is to try to determine the features that affect the most influential in determining a speaker's dialect. This may be related to the issues of political asylum, which is an important issue, of course, the Syrian dialects are varied. I really felt that the reading affected the tone in some cases, but I congratulate you on a beautiful and interesting search for success

Dec 11, 2020 1:13 PM Egyptian 2I was born in Egypt 1986 I lived in Saudi Arabia one year 1992-1993 I lived in America 2012-2015

The manner of speaking is real and not fabricated (12 Syrian likely)

The method of pronouncing words and emphasizing some letters is similar to the Maghreb (21 Moroccan highly unlikely)

The manner of speaking is contrived (22 Jordanian unlikely)

Some words such as "purposes (l2aghra:D), dir, ... the manner of pronouncing words." (32 Syrian highly likely)

The way the letters are pronounced Lebanese (soft) (69 Lebanese highly unlikely)

Dec 11, 2020 12:26 PM Saudi 8

Kalash (17 Kuwaiti highly unlikely)

Manah Hodol (50 Syrian likely)

Alolo Manah 2arya (53 Lebanese uncertain)

Hydol (69 Lebanese uncertain)

Most of the dialects revolve around the region of Syria, so it is very difficult to differentiate them.

Dec 11, 2020 11:15 AM Tunisian (linguist) France, yes, Tunisia, from the age of 5 to the age of 20, and the United States of America, from the age of 25 so far

This is Egyptian. Several pointers: lexicon, the [q] feature and the [g] feature. Also you can hear the attitude: very Egyptian. (haha) (2 Egyptian highly unlikely)

This is Algerian. The other one was then probably Moroccan. Moroccans are known to speak rather fast. Here, the speaker doesn't speak fast and sounds very close to my dialect (Tunisian). I also think the intonation gave it away. The way the speaker said weld elhram, it is very North African (6 Moroccan highly unlikely)

Kalash this Iraqi dialect (8 Kuwaiti highly unlikely)

Affrication. Not Syrian, however, I can't detect the exact location of this dialect. Peninsular Arabic? (9 Jordanian highly unlikely)

Egyptian! Very distinctive features! [g] and bitou3 and [q] awi etc (13 Egyptian highly unlikely)

Iraqi or Peninsular Arabic? Affrication and kolich (17 Kuwaiti highly unlikely)

This is Algerian or Morrocan because the speaker said "dial" and "bezaf". I am leaning towards Morrocan but I am not sure (21 Moroccan highly unlikely)

Affrication and [g] feature. But the speaker said mouch so I can't put my finger on their dialect. (22 Jordanian highly unlikely)

Peninsular Arabic? a little bit of affrication and intonation (prosody) is different (29 Iraqi highly unlikely)

I think this Saudi. There is more affrication than in what would Syrian sound like (42 Iraqi highly unlikely)

Again between Lebanese and Syrian. The speaker says Madine instead of Madina and I am not sure if that is a Lebanese feature or Syrian (50 Syrian uncertain)

This is Syrian. Although the speaker is speaking in a more formal way, they have an intonation that sounds like Syrian to my ear. (69 Lebanese highly likely)

I don't think that I know the difference between Syrian and Lebanese. Hence, I don't think I was very useful here. But I do know that both dialects are found to be very beautiful and are preferred to other dialects in the Arab world. I can recognize Egyptian easily because of very prominent pop culture. My exposure to Syrian and Lebanese is not as important as it was with Egyptian. Such a cool project. Good luck with everything.

Dec 11, 2020 10:43 AM Saudi 7 (linguist)

Take care of you, or any Egyptian vocabulary (2 Egyptian highly unlikely)

Vocabulary pronunciation of letters and intonation (6 Moroccan highly unlikely)

Letter logic (7 Syrian highly likely)

Iraqi because of the slow talk (8 Kuwaiti unlikely)

Speech tone and lettering areas (9 Jordanian unlikely)

HashinAoy Egyptian vocabulary (13 Egyptian highly unlikely)

Pronunciation of some letters and vocabulary (16 Syrian highly likely)

His Kuwaiti accent due to the slowness of the speaker, his tone, and the pronunciation of some letters (17 Kuwaiti highly unlikely)

Some vocabulary is Moroccan vocabulary and also the tone of the hadith (21 Moroccan highly unlikely)

Pronunciation of letters and vocabulary (24 Syrian highly likely)

Vocabulary and Tone areas for letters (29 Iraqi highly unlikely)

It seems from the pronunciation of the letters that it is the accent of Syria (36 Syrian likely)

Some words like Kalash and the speed and tone of the speaker (42 Iraqi unlikely)

Pronunciation of letters and vocabulary (50 Syrian highly likely)

Dec 11, 2020 2:41 AM American 2 I was born in the United States. I lived in southern Spain for 2 years.

The use of the word "Bizaf" and he sounds like my friends from Morocco. Also the use of "dialik." (6 Moroccan highly unlikely)

This is Moroccan. The vowel omissions and the use of the word "bizaf" (21 Moroccan highly unlikely)

This person uses the case markings much more clearly than I might expect in a dialect. (24 Syrian unlikely)

It was difficult for me to distinguish the dialects. I felt most comfortable when I heard MSA and Darija.

Dec 11, 2020 2:05 AM American 1 I was born in the United States and lived in Iraq for about a year and a half (from 2006-2007) and resided in Egypt from 2011-2012

This sounds Levantine and I believe Syrian. Phonological: 2 for q although not in qariye 'villiage' strong imaala but that can be in many other places like southern Egypt, the z for d3 morphology: haduul 'those guys' mu participle negation. rafa2aat = friends, aghradh 'things'. (1 Syrian likely)

Cairene Arabic. glottal stop for OA q. OA d3 as [g], butuu3 genitive exponent. stress patterns. duul for human proximal and distal. Lexical level e.g.khallibaal-ak min ...... (2 Egyptian highly unlikely)

This is North African most likely Moroccan Arabic. Phonological OA q = g Use of ka= imperfective proclitic, Use of raas as reflexive pronoun, use of fash for when, use of diyaalgenetive exponent. Bizzaf = a lot (6 Moroccan highly unlikely)

This sounds like Iraqi Arabic but it's weird to hear kulish at the end of the sentence I think it sounds better as hadhuulkullish mu zeeniin. also this person uses u instead of a for him. I wonder if Iraqi is not this person's dialect. (8 Kuwaiti unlikely)

haduul for those guys mish kwayissiin. no imaala ?? (9 Jordanian unlikely)

totally Cairene for the same reasons as the other example in here. (13 Egyptian highly unlikely)

Levantine and I think Syrian. Imaala use of aghraaD 'things' I wonder if this is Eastern Syria? Also use of SaaHib 'friend'. (16 Syrian likely)

This is Mesopotamian Arabic - not Syrian. Kulish 'very'. ra7 future marker … that or variants thereof are common though in a bunch of dialects. the big thing is keeping the nuun in yi3arafuun 'they know. the use of raba3 'the gang' for friends use of -a 'him' instead of 'u' or 'o' like in levantine. (17 Kuwaiti highly unlikely)

This is North African Arabic, again probably Moroccan. Phonological: Morphological: use of raas as reflexive pronoun, diyal for genitive exponent ka= imperative verbal proclitic. (21 Moroccan highly unlikely)

This sounds like a gulf Arabic speaker switching his dialect up. The 3 is super constricted and that sounds like either Iraqi or a gulf speaker to me. Theres also the retention of interdentals that appears in mesopotamian and gulf arabic. It's wierd to hear mish kwayyisin after that. He also has imaala in there but it just doesn't sound like that's natural for this speaker. (22 Jordanian highly unlikely)

This could be Syrian but maybe not standard. I dunno -a = him yilli for illi (24 Syrian uncertain)

This sounds Levantine to me and I think Syrian. Use of -o for him instead of -u. Use of z for OA d3. Imala in qariye. (27 Syrian likely)

This is a native Iraqi speaker. I think that this guy was changing his dialect up for other clips. kullish in right spot. (29 Iraqi highly unlikely)

sounds levantine and Syrian.2 for q imaala z for d3 aghraad 'things' diirbaal-ak can be Mesopotamian and levantine. (32 Syrian likely)

This sounds levantine I think Syrian. Imala, same sound correspondences. Use of -u for him.. use of aghrad for things. mnaa7 for good people (36 Syrian likely)

Why are some of these recordings faster? This sounds similar to Baghdadi Arabic or a variety of Mesopotamian Arabic . Phonological OA q = g, backed and low alif. Morphological: He kept the n in third person plural imperfect verbal inflection - not very common in a lot of Arabic dialects, but it is a feature in Iraqi varieties, mu for participle negation. Lexical: Use of Kullish "very" ,zeen = good in Iraqi. (42 Iraqi highly unlikely)

yes Syrian. imaala ma for negation of participles. 2ariyee (53 Lebanese highly likely)

This is levantine but I think Lebanese because of the fronted alif. deer byeel-ak (69 Lebanese unlikely)

Dec 10, 2020 6:50 PM Saudi 5 (linguist)

Egyptian dialect sound distinctive words (2 Egyptian highly unlikely)

Western Morocco tone of voice, distinctive words (6 Moroccan highly unlikely)

Tone of voice distinctive words (13 Egyptian highly unlikely)

The dialect is Shamy, almost Jordanian, but not Syrian (16 Syrian unlikely)

Kuwaiti dialect sound distinctive words (17 Kuwaiti highly unlikely)

Moroccan dialect distinctive words tone of voice (21 Moroccan highly unlikely)

Iraqi dialect, tone of voice, distinctive words (29 Iraqi highly unlikely)

Tone of voice distinctive words (32 Syrian highly likely)

Dec 10, 2020 6:41 PM Saudi 4

The tone is light and simple on the tongue (1 Syrian likely)

Smoothness of pronunciation and lightness of tone (12 Syrian likely)

The dialect is Damascene, but I don't know which country you belong to (24 Syrian uncertain)

Perhaps because it is light on the tongue (27 Syrian likely)

Perhaps because of its smoothness and lightness on the tongue (36 Syrian likely)

Smooth and subtle accent on the tongue (43 Syrian highly likely)

Perhaps for ease of articulation (50 Syrian highly likely)

Lebanese dialect is heavier than Syrian (53 Lebanese highly unlikely)

Dec 10, 2020 6:35 PM Saudi 3

Mo Manah Syrian words (1 Syrian likely)

the words (9 Jordanian likely)

Words used (16 Syrian likely)

Because it uses the word Mo Manah (24 Syrian likely)

Dec 10, 2020 6:19 PM Saudi 2 (linguist)

Vocabulary and pronunciation of letters in addition to tone (2 Egyptian highly unlikely)

Vocabulary (7 Syrian likely)

Some of the existing vocabulary and tones belong to Egypt (13 Egyptian highly unlikely)

There are some vocabulary that refers to the Syrian accent, such as Der Balk (16 Syrian likely)

Some of the sounds and tones do not belong to the Syrian dialect (22 Jordanian unlikely)

Kalash I think it is used in Iraq (42 Iraqi highly unlikely)

Dec 10, 2020 1:40 PM Syrian 4 I was born in Syria 1989

The word village The pronunciation of the letter qaaf is not similar to what Syrians generally use, but it may be from one of the villages in northern Syria (9 Jordanian uncertain)

Rah words and purposes (12 Syrian highly likely)

From some words like Rah (Syrian 36 highly likely)

Dec 10, 2020 1:25 PM UAE 1 (linguist) I was born in the United Arab Emirates and lived in the United Kingdom in the year 2015 for two years

The dialect is close to the Syrian accent, and the pronunciation of the letter R is closer to the way it is pronounced by some Syrian regions, where the letter R is thinned (36 Syrian likely)

The style of pronunciation of the dialogues (53 Lebanese likely)

There is a convergence between the Levantine dialects, especially the cities that are on the borders of the neighboring cities of another country. There may be some words used close to each other, or even the manner of pronouncing the words.

Dec 10, 2020 1:10 PM Syrian 3 (linguist)

The Egyptian dialect, especially the letter J (2 Egyptian highly unlikely)

Dialect and some vocabulary (8 Kuwaiti highly unlikely)

Pronounce some words such as: purposes (al aghrad) (16 Syrian highly likely)

The dialect and the exits of the letters (24 Syrian highly likely)

The dialect and some vocabulary, such as: Mu Zenin (42 Iraqi highly unlikely)

Residents of some areas in Syria speak dialects of neighboring countries, such as the people of the Syrian desert who speak the dialect of the people of Iraq and the people of Houran with a dialect of some of them very close to Jordan ,,,

Dec 10, 2020 12:34 PM Syrian 2 (linguist) Syria, yes, Turkey - Istanbul 2017-2019

From the word need, the dialect is Egyptian (2 Egyptian highly unlikely)

The word Balzaf, Algerian or Tunisian (6 Moroccan highly unlikely)

The dialect is Levantine, from the word mu (7 Syrian highly likely)

From the word Kalash, the dialect is Gulf (8 Kuwaiti highly unlikely)

Egyptian dialect, from the word of your companions (13 Egyptian highly unlikely)

The word Kalash, the dialect is Kuwaiti (17 Kuwaiti highly unlikely)

Your dialect is a Maghreb word (21 Moroccan highly unlikely)

Hazul word, Houranic word (22 Jordanian highly likely)

The dialect is Gulf, from the word Zina (29 Iraqi highly unlikely)

From the word Kalash, this is Direya (42 Iraqi likely)

From the word Munaih, instead of Manah, the Lebanese dialect (53 Lebanese highly unlikely)

Dec 10, 2020 11:09 AM Sudanese, UAE born

Egyptian dialect due to the use of the word "dole, monster, and awi" " dual wawahashin w 'awy" (2 Egyptian highly unlikely)

The word "Balzaf" is specific to the Moroccan dialect (6 Moroccan highly unlikely)

The word "Alu", "Derbalak" and "Hadul" is used in the Levant (7 Syrian likely)

Using "not good" (9 Jordanian unlikely)

His accent is Egyptian, using two monsters, and the gym is not thirsty lahujhmisriataistikhdamwahashinwaljimghyrmaetisha (13 Egyptian highly unlikely)

The word "Deira" is used in Kuwait and the Gulf, and they also use Kalash "but differently from the Iraqis." (17 Kuwaiti highly unlikely)

Balsaf, dialyk, and they will most likely know you with a Moroccan accent (21 Moroccan highly unlikely)

Most likely, a Palestinian or Jordan dialect, because of this (hudhul) and its way of pronouncing qaf in the word "village" (22 Jordanian highly unlikely)

Stretch in the word "your accent" (36 Syrian highly likely)

The word "Kalash" is specific to the Iraqi dialect (42 Iraqi highly unlikely)

Dec 10, 2020 10:45 AM Syrian 19 I was born in Syria, I left Syria and came to England three years ago

The way the Syrian Bedouins speak (8 Kuwaiti highly likely)

The way of the Bedouins (9 Jordanian highly likely)

The accent of the Syrian Bedouin people (17 Kuwaiti likely)

Dec 10, 2020 10:44 AM Egyptian 1

Hiring the word minah (1 Syrian highly likely)

This is the accent of the people of Egypt using the word "let your mind", which is an Egyptian colloquial term (2 Egyptian highly unlikely)

The use of the pronunciation of Bazaf and phonetic composition of sentences (6 Moroccan highly unlikely)

I have spoken to Syrians before, and it is almost the same as their use of the word objects 'aghrad and their breaking of the last sounds in words, but it is also close to the dialect of the people of Lebanon. (7 Syrian uncertain)

Use it to pronounce Kalash (8 Kuwaiti highly unlikely)

Use the word village instead of village (qaryatbadalaan min qarya) (9 Jordanian unlikely)

His use of the word purposes (aghrad) to break it for the last sounds in words (12 Syrian likely)

The use of the word "Tawaa," countryside and need (13 Egyptian highly unlikely)

The use of the word Kalash (17 Kuwaiti highly unlikely)

The use of the word Bazaf, one of the words of the people of the Maghreb is also the speed of speech and because I already know the Syrian dialect (21 Moroccan highly unlikely)

The use of the word for the purposes of Manah (24 Syrian highly likely)

The last sound in the words broke the use of a word and the pronunciation of objects (aghrad) (27 Syrian highly likely)

The speaker's use of the sound of u at the end of a word is not uncommon in the Syrian dialect. (29 Iraqi unlikely)

Because the speaker enjoys speaking as if he is reading it and not at liberty with it (32 Syrian unlikely) (IMITATION??)

Use it to pronounce objects (l2aghra:D), village and break the last sounds into words (36 Syrian highly likely)

The use of the word Kalash (42 Iraqi highly unlikely)

The use of pronouncing purposes and breaking the last sound in the word aistikhdamlifaz 'aghradwakasralsawtal'akhir fi alkalima (43 Syrian highly likely)

The words are Syrian expressions, and the phonemic structure of the word refers to the people of Levant (50 Syrian highly likely)

Use the word 2arya instead of village (53 Lebanese highly unlikely)

The use of the pronunciation of Hedol (69 Lebanese unlikely)

Dec 10, 2020 10:38 AM Lebanese 1 (linguist)

Egyptian. The ج as g gives it away (2 Egyptian highly unlikely)

Either too Syrian or not at all (8 Kuwaiti uncertain) IMITATION??

حانمولدوه: الكلماتالتالية (The following words: Hadul Mu Manah) But it also sounds a tiny bit fake as the same letter ق was once pronounced as ق and once as أ. This could be because of the many different Syrian dialects, but I'm not sure. (12 Syrian likely) IMITATION??

Egyptian, most probably. The pronunciation of the ج (jeem) as g gives it away (13 Egyptian highly unlikely)

The ش (shin) sound at the end hints at Khaliji dialect more than Syrian (17 Kuwaiti unlikely)

The fast pace indicates a more Jazaeri/Tunisi dialect (21 Moroccan unlikely)

غالودهولكلشمورحيغلونعليك : الكلماتالتالية (alkalimataltaaliat: ghaludhwlklsh mw rhinyaghlunealayk) The pronunciation of the ق as g makes it sound as if it's more Khaliji than Syrian (29 Iraqi highly unlikely)

Sounds like a Syrian who has been in Lebanon for a while. Has a hint of Syrian but the more drawn out pronunciation hints at the Lebanese dialect (36 Syrian uncertain)

Lebanese. Too drawn out. The Lebanese dialect is actually sometimes made fun of by Palestinians because it's too "ممحون" / soft (53 Lebanese highly unlikely)

Lebanese. مشمناهmshmanah gave it away. Even if the ق (qaf) is pronounced as 69(ق Lebanese highly unlikely)

This was so interesting and fun to do. Even though I am Lebanese I still get incredibly confused. I'll send it out to my sister who's good at differentiating the dialects. I hope you used people who are natives in order to get genuine pronunciations

Mar 25, 2021 8:58 PM Jordanian 6

the sound of the words and letters used is similar. (1 Syrian likely)

this is Egyptian. I can identify it from the sound of the letters (2 Egyptian highly unlikely)

the words used are not Syrian, Bizzat is not Syrian (6 Moroccan highly unlikely)

the word killish is not Syrian (8 Kuwaiti highly unlikely)

the words used are not syrian. (9 Jordanian highly unlikely)

this is not Syrian. they do not say the Agrad they would say grad. (12 Syrian unlikely)

from the words we7shin and 7aga (13 Egyptian highly likely)

from the word grad (16 Syrian likely)

killish is not Syrian (17 Kuwaiti highly unlikely)

the choice of words is more like Jordanian accent. (22 Jordanian highly unlikely)

kolish is not Syrian (29 Iraqi highly unlikely)

this is Syrian from the way they pronounce the words (36 Syrian highly likely)

the word mozenin and kollesh is not Syrian (42 Iraqi highly unlikely)

it is Lebanese and not Syrian. the sound of the letters are different (69 Lebanese highly unlikely)

Syrians in general have specific way of pronouncing words, that is known to Middle Eastern people. Each region has a different accent, Gulf has one accent including but not limited to Iraqi, Middle Eastern has a different accent and Egyptians have a completely different accent. Furthermore, Moroccons and Algeria, Tunisia who were occupied by the French, have an accent that is similar to French language in the way they divide the phonetics

Lebanese 4 Jul 2, 2021 11:40 AM

the way she said "hadol" not "hole" or "haydole" (8 Kuwaiti likely)

the speaker says "garyeh" instead of "qaryeh" and the dialect seems more egyptian or a Lebanese person from the Chouf (Deir lQamar) area (9 Jordanian unlikely)

the way the speaker says "moumnah" and "lahejtak" (36 Syrian likely)

the speaker talked to fast so it was hard to focus on his dialect however, he did say "2al ashya2" but still some Lebanese dialects might say that too. (42 Iraqi uncertain)

Syrian 37 Oct 12, 2021 8:41 AM Oct 12, 2021 8:41 AM The country of origin is Palestine, born in Syria, the countries in which Syria has lived for 28 years, Iraq - Kurdistan (4 months) Turkey (3 months) One country contains more than one dialect, for example, southern Syria (the Hauran region) speaks a dialect closer to Jordanian than to Levant, and the northeastern regions have a dialect closer - and may be identical - to Iraqi due to family and clan overlap. also. Therefore, it is necessary to specify more precisely, whether the research is conducted on the Levantine dialect, or the reduced Levantine dialect that takes the character of the Syrian dialect.

Highly unlikely--One of the dialects used in Palestine or Jordan for its use of the word "hadula" and the word "lamma" and how to pronounce the word "rafqatu"

Highly unlikely--" al'aryaflafzkamlathajana "wahshin 'uwy"

Highly unlikely

Highly likely

Highly unlikely--I think it's Saudi

Highly unlikely

Highly likely

Highly unlikely

Highly likely

Highly unlikely

Highly unlikely--Algerian dialect due to the use of the word balzaf to describe a lot of things and how to pronounce the exits of wars and other vocabulary

Highly likely--This dialect is used in southern Syria and northern Jordan, especially the Hawran region, in the manner of pronouncing the word village, the word hudoul, and others

Highly likely--Hadol: a word used in some Syrian dialects as a sign tool. They told him: Aalulu: a colloquial word in some Syrian dialects. Therefore, this is a Syrian dialect

Highly likely

Highly unlikely--An Iraqi dialect for its use of the word "kalash" to describe a lot, as well as some other words that show the exits of letters and stress in other places, such as the word "miglon" and the intensity on the gen.

Highly likely

Highly likely--alqaryat - hadual - algharadi- alkh

Highly unlikely

Highly likely

Highly likely

Highly unlikely--Baya'in: with the ba'a fraction used by the Lebanese/the village - where the qaf is pronounced as a hamza, and this is a Lebanese word, but some Syrian dialects pronounce the qaf as a hamza, but they do not use the description of the village, but rather a village. Therefore, this is a Lebanese dialect, not Syrian.

Highly unlikely--Lebanon

Syrian 33 Oct 11, 2021 4:47 PM

Highly likely

Highly unlikely--alhajatu: al'aghradi, jiha: jimtulfaz g lahjatmisria

Highly unlikely--dyalik

Highly likely--lima- rah- dirbialka- hadual When - I will - take care of you - the words stretch a little

Highly unlikely--Mo Zineen Kalash - Iraqi dialect

Uncertain--It may be the dialect of Daraa due to the use of the word "Hoodul" - it is not good without exaggeration

Highly likely--Deir Balak - Hadol

Highly unlikely--ashabuhu- bitueu- wahshini- al'aryaf- jaha Al-Jim between Al-Jim and Al-Gin (Egyptian dialect)

Highly likely--dirbialki, hadul, wasalalkalam

Highly unlikely--Kalash

Highly unlikely

Highly unlikely--Hazoul-Raehoglu (Jordanian)

Highly likely--Is someone Syrian imitating the Levantine dialect 😆 Stretch the words a little - using the term Deir Balak

Highly likely--Stretch the words a little, words like "Hadoul" and "Dir Balak"

Highly unlikely--mozenin - everything

Highly likely--Deir Balak - Hadol

Highly likely--rah yaghlu- lima yasmaeu- hadwal - stretch the words a little

Unlikely--The use of the word Kalash, which is an Iraqi dialect

Highly likely--lamaayasmaeu, hadual, algharad,

Highly likely--The use of the word Hdol

Highly unlikely--Lebanese dialect - converting A to Z is light like the word balak-belek

Highly unlikely--Sabu, Hydol, Belek

Syrian 35 Oct 11, 2021 10:06 PM I was born in Syria as a Palestinian refugee in 1985. I have been living in Europe since 2014.

Highly likely

Highly unlikely--The Egyptian dialect is one of the most famous dialects in the Arab world. It is famous for changing the letter jim to a letter that comes from the throat, not the tongue.

Highly unlikely

Highly likely

Highly unlikely

Highly unlikely

Highly likely

Highly unlikely

Highly likely

Highly unlikely

Highly unlikely--Tunisian dialect, fast dialect and the word Bazaf Tunisian identity

Highly unlikely

Highly likely

Highly likely

Highly unlikely

Highly likely

Highly likely

Highly unlikely

Highly likely--Switch the qaf to the hamza. A clear feature in the dialect of the Levant. But the Syrian is known from the long sounds of the vowels, (the) the definition is broken.

Highly likely

Highly unlikely

Highly unlikely

Syrian 31 Oct 11, 2021 8:57 AM Syria UK (8 years)

Highly likely

Highly unlikely

Highly unlikely

Highly likely

Uncertain--could be from eastern syria or iraq

Likely--might be jordanian or from Horan

Highly likely

Highly unlikely

Highly likely

Uncertain--might be from eastern syria (close to Iraq)

Highly unlikely

Unlikely--most likely jordanian

Highly likely

Highly likely

Highly unlikely

Highly likely

Highly likely

Uncertain--either iraqi or from eastern syria

Highly likely

Highly likely

Highly unlikely--lebanese

Highly unlikely

References

Abdi, H. (2009). Signal detection theory. In McGaw, B., Peterson, P.L., Baker, E. (Eds.): *The International Encyclopedia of Education* (3rd Ed). New York: Elsevier.

Abu-Haidar, F. (2012). Baghdad Arabic. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Aguade, J. (2018). The Maghrebi dialects of Arabic. In Holes, C. (Ed.) *Arabic Historical Dialectology* (pp. 259-292). Oxford: Oxford University Press.

Adank, P., Evans, B. G., Stuart-Smith, J. & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance,* 35(2), pp. 520-529. https://psycnet.apa.org/doi/10.1037/a0013552

Aitken, C. G. G. et al. (2011). Expressing evaluative opinion: a position statement. *Science and Justice*, 51(1–2). https://doi.org/10.1016/j.scijus.2011.01.002

Almbark, R., & Hellmuth, S. (2015). Acoustic analysis of the Syrian Arabic vowel system. *Proceedings of the 18th ICPhS.* Glasgow, Scotland.

Al-Wer, E., Horesh, U. Herin, B. & Fanis, M. (2015). How Arabic Regional Features Become Sectarian Features. Jordan as a Case Study. *Journal of Arabic Linguistics*, 62, pp. 68-87.

Arnaut, K. & Spotti, M. (2015). Superdiversity Discourse. In Tracy, K., Ilie, C., & Sandel, T. (Eds.), *The International Encyclopedia of Language and Social Interaction* (pp. 1-7). United Kingdom: Wiley-Blackwell. https://doi.org/10.1002/9781118611463.wbielsi138

Atkinson, N. (2015). *Variable factors affecting voice identification in forensic contexts*. PhD thesis, University of York.

Augusto, L.M (2010) Unconscious knowledge: A survey. *Advances in Cognitive Psychology,* 6, pp. 116-141.

Baker, W., Eddington, D. & Nay, L. (2009). Dialect identification: the effects of region of origin and amount of experience. *American Speech*, 84(1), pp. 48-71. https://doi.org/10.1215/00031283-2009-004

Baltisberger, E. & Hubbuch, P. (2010). LADO with specialized linguists – the development of LINGUA's working method. In Zwaan, K., Muysken, P. & Verrips, M. (Eds.), *Language and Origin: The Role of Language in European Asylum Procedures: Linguistic and Legal Perspectives* (pp. 9-19). Nijmegen: Wolf Legal Publishers.

Baranowski, M. (2017). Class matters: the sociolinguistics of GOOSE and GOAT in Manchester English. *Language Variation and Change*, 29, pp. 301–339. https://doi.org/10.1017/S0954394517000217

Baranowski, M. & Turton, D. (2015). Manchester English. In Hickey, R. (Ed.), *Researching Northern Englishes* (pp. 293-316). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Bargh, J. & Chartrand, T. (1999). The Unbearable Automaticity of Being. *American Psychologist,* 54(7), pp. 462-479. https://doi.org/10.1037/0003-066X.54.7.462

Barras, W. (2006) *The SQUARE-NURSE merger in Greater Manchester: the impact of social and spatial identity on phonological variation*. MSc dissertation, University of Edinburgh.

Barras, W., Honeybone, P. & Trousdale, G. (2007). Is word-frequency a factor in phonological demergers? The sociophonology of SQUARE and NURSE in Lancashire English. *Paper presented at ICLCE2.* Toulouse, France.

Bassiouney, R. (2018). An alternative approach: Understanding diglossia/code switching through indexicality: The case of Egypt. In Benmamoun, E. & Bassiouney, R. (Eds.), *The Routledge Handbook of Arabic Linguistics* (pp. 345-358). London: Routledge.

Behnstedt, P. (2012). Syria. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Behnstedt, P. &Woidich, M. (2013). Arabic Dialects. In Owens, J. (Ed.), *The Oxford Handbook of Arabic Linguistics* (online edition). Oxford Handbooks Online.

Beal, J.C. (2010). *An Introduction to Regional Englishes: Dialect Variation in England*. Edinburgh: Edinburgh University Press.

Bener, A. & Mohammad, R.R. (2017). Global distribution of consanguinity and their impact on complex diseases: Genetic disorders from an endogamous population. *The Egyptian Journal of Medical Human Genetics,* 18, pp. 315–320. https://doi.org/10.1016/j.ejmhg.2017.01.002

Bentahila, A., Davies, E. & Owens, J. (2013). Codeswitching and Related Issues Involving Arabic. In Owens, J. (Ed.), *The Oxford Handbook of Arabic Linguistics* (online edition). Oxford Handbooks Online.

Blommaert, J. (2009). Language, Asylum, and the National Order. *Current Anthropology*, 50(4), pp. 415-441.

Blommaert, J. (2001). Investigating narrative inequality: African asylum seekers' stories in Belgium. *Discourse and Society*, 12(4), pp. 413–49. https://doi.org/10.1177/0957926501012004002

Blommaert, J. & Verschueren, J. (1991). The Pragmatics of Minority Politics in Belgium. *Language in Society*, 20(4), pp. 503-531. https://doi.org/10.1017/S0047404500016705

Blommaert, J. & Verschueren, J. (1992). The Role of Language in European Nationalist Ideologies. *Pragmatics*, 2, pp. 55-75.

Bobda, S., Wolf, H. and Peter, L. (1999). Identifying regional and national origin of English-speaking Africans seeking asylum in Germany. *Forensic Linguistics*, 6(2), pp. 300-331. https://doi.org/10.1558/ijsll.v6i2.300

Braber, N., Smith, H., Wright, D., Hardy, A., & Robson, J. (2022). Assessing the Specificity and Accuracy of Accent Judgments by Lay Listeners. *Language and Speech*. https://doi.org/10.1177/00238309221101560

Broeders, A. P. A. (2010). Decision making in LADO – A view from the forensic arena. In Zwaan, K., Muysken, P. & Verrips, M. (Eds.), *Language and Origin: The Role of Language in European Asylum Procedures: Linguistic and Legal Perspectives* (pp. 51–60). Nijmegen: Wolf Legal Publishers.

Brustad, K.E. (2000). *The syntax of spoken Arabic : a comparative study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*. Washington D.C.: Georgetown University Press.

Cambier, T. & van den Boogert, N. (2008). Language analysis in the Netherlands. Presentation to *Eurasil workshop on language analysis*. Brussels, Belgium.

Cambier-Langeveld, T. (2010a). The role of linguists and native speakers in language analysis for the determination of speaker origin. *International Journal of Speech, Language and the Law*, 17(1), pp. 67–93. https://doi.org/10.1558/ijsll.v17i1.67

Cambier-Langeveld, T. (2010b). The validity of language analysis in the Netherlands. In Zwaan, K, Muysken, P., & Verrips, M. (Eds.), *Language and origin: The role of language in European asylum procedures: A linguistic and legal survey* (pp. 21–33). Nijmegen: Wolf Legal Publishers.

Cambier-Langeveld, T. (2012). Clarification of the issues in language analysis: a rejoinder to Fraser and Verrips. *International Journal of Speech, Language and the Law,* 19(1), pp. 95–108. https://doi.org/10.1558/ijsll.v19i1.95276

Cambier-Langeveld, T. (2014). State-of-the-art in language analysis: a response to the chapter on LADO in the Oxford Handbook of Language and Law. *International Journal of Speech, Language and the Law,* 21(2), pp. 371–381. https://doi.org/10.1558/ijsll. v21i2.371

Cambier-Langeveld, T. (2016). Language analysis in the asylum procedure: a specification of the task in practice. *International Journal of Speech, Language and the Law,* 23(1), pp. 25–41. https://doi.org/10.1558/ijsll.v23i1.17539

Cambier-Langeveld, T. (2018a). Language analysis in the asylum procedure: consider the context. In Nick, I. M. (Ed.), *Forensic Linguistics: Asylum-Seekers, Refugees and Immigrants* (pp. 1–22). Malaga: Vernon Press.

Cambier-Langeveld, T. (2018b). *Language Indication*. Presentation to BAMF. Federal Republic of Germany.

Campbell, J. (2013). Language analysis in the United Kingdom's refugee status determination system: seeing through policy claims about 'expert knowledge'. *Ethnic and Racial Studies*, 36(4), pp. 670-690. https://doi.org/10.1080/01419870.2011.634506

Channon, A., Foulkes, P. & Walker, T. (2018). 'But what is the reason why you know such things?' Question and response patterns in the LADO interview. *Journal of Pragmatics*, 129, pp. 154-172. https://doi.org/10.1016/j.pragma.2018.03.015

Clopper, C. G. & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change, 18*(2), 193–221. https://doi.org/10.1017/S0954394506060091

Cohen, E. (2012). The Evolution of Tag-Based Cooperation in Humans: The Case for Accent. *Current Anthropology*, 53(5), pp. 588-616. https://doi.org/10.1086/667654

Corcoran, C. (2004). A critical examination of the use of language analysis interviews in asylum proceedings: a case study of a West African seeking asylum in the Netherlands. *International Journal of Speech, Language and the Law*, *11*(2), pp. 200–221.

De Taalstudio (2022). *Termination of Service*. Retrieved 12 January 2023 from https://www.taalstudio.nl/taalanalyse/index_uk.html

Deutsche Welle (2017). *Automatic speech analysis software used to verify refugees' dialects*. Retrieved 26 October 2018 from https://www.dw.com/en/automatic-speech-analysis-software-used-to-verify-refugees-dialects/a-37980819

Dillon, B. & Wagers, M.W. (2021). Approaching Gradience in Acceptability with the Tools of Signal Detection Theory. In Goodall, G. (Ed.), *The Cambridge Handbook of Experimental Syntax*, (pp. 62 – 96). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108569620.004

Dukhan, H. (2014). Tribes and Tribalism in the Syrian Uprising. *Syria Studies*, 6(2), pp. 1-28.

Eades, D. (2005). Applied linguistics and language analysis in asylum seeker cases. *Applied Linguistics,* 26(4), pp. 503–526. https://doi.org/10.1093/applin/ami021

Eades, D., & Arends, J. (2004). Using language analysis in the determination of national origin of asylum seekers: An introduction. *International Journal of Speech, Language and the Law,* 11(2), pp. 179–199. https://doi.org/10.1558/ijsll.v11i2.179

Eades, D., Fraser, H., Siegel, J., McNamara, T., & Baker, B. (2003). Linguistic identification in the determination of nationality: A preliminary report. *Language Policy,* 2(2), pp. 179–199. https://doi.org/10.1023/A:1024640612273

Eid, M. (2012). Locatives. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Elvers, G. (n.d. a). *Signal Detection Theory: d' & β*. Retrieved 20 August 2022 from https://elvers.us/perception/sdtGraphic/

Elvers, G. (n.d. b). *Signal Detection Theory.* Retrieved 16 June 2020 from https://elvers.us/perception/sdtCalculator/

Eriksson, A. (2008). Guidelines? What guidelines? Paper presented at *IAFPA 17th Annual Conference.* Lausanne, Switzerland.

Eriksson, A. (2010a). LADO research from a forensic point of view – Some key issues to be considered. Paper presented at *Gothenburg Workshop on Language Analysis*. Gothenburg, Sweden.

Eriksson, A. (2010b).The Disguised Voice: Imitating Accents or Speech Styles and Impersonating Individuals. In Llamas, C. Watt, D. & Llamas, C. (Eds.), *Language and Identities* (pp. 86–98). Edinburgh: Edinburgh University Press.

Eurostat (2022). *Number of asylum applicants (non-EU citizens), EU, 2008–2020*. Retrieved 17 June 2022 from https://web.archive.org/web/20220408080611/https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Figure_1_Number_of_asylum_applicants_(non-EU_citizens),_EU,_2008%E2%80%932020.png

Findahl, J. (2018). Collecting contemporary knowledge on Gorani spoken east of Mosul. In Nick, I.M. (Ed.), *Forensic Linguistics: Asylum-seekers, Refugees and Immigrants* (pp. 57-74). Malaga: Vernon Press.

Fitzmaurice, S. (2019). Transnational Languages, Multilinguals and the Challenges for LADO. In Patrick, P., Schmid, M. & Zwaan, K. (Eds.) *Language analysis for the determination of origin: current perspectives and new directions* (pp. 167-192). Cham: Springer.

Foley, R.A. & Mirazón Lahr, M. (2011). The evolution of the diversity of cultures. *Philosophical Transactions of the Royal Society: Biological Sciences*, Vol. 366, No. 1567, pp. 1080-1089 https://doi.org/10.1098/rstb.2010.0370

Fortress Europe Circular Letter (1998). *Controversial language tests for the determination of asylum seekers' country of origin*. Retrieved 24 November 2018 from http://www.fecl.org/circular/5304.htm

Foulkes, P., & French, P. (2012). Forensic speaker comparison: A linguistic-acoustic perspective. In Tiersma, P. M. & Solan, L. M. (Eds.), *The Oxford handbook of language and law* (pp. 418–421). Oxford: Oxford University Press.

Foulkes, P., French, P. & Wilson, K. (2019) *LADO as forensic speaker profiling*. In Patrick, P., Schmid, M. & Zwaan, K. (Eds.), *Language analysis for the determination of origin: current perspectives and new directions* (pp. 91-116). Cham: Springer.

Foulkes, P. & Hughes, V. (in press) Dialectological and sociolinguistic foundations of forensic speaker comparison. To appear in Nolan, F., McDougall, K. & Hudson, T. (Eds.), *The Oxford Handbook of Forensic Phonetics*. Oxford: Oxford University Press.

Foulkes, P., & Wilson K. (2011). Language analysis for the determination of origin: An empirical study. *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 691-694. Hong Kong.

Fraser, H. (2009). The role of "educated native speakers" in providing language analysis for the determination of the origin of asylum seekers. *International Journal of Speech Language and the Law,* 16(1), pp. 113–138. https://doi.org/10.1558/ijsll.v16i1.113

Fraser, H. (2011). The role of linguists and native speakers in language analysis for the determination of speaker origin: A response to Tina Cambier-Langeveld. *International Journal of Speech, Language and the Law,* 18(1), pp. 121–130. https://doi.org/10.1558/ijsll.v18i1.121

Fraser, H. (2019). The role of native speakers in LADO: Are we missing a more important question? In Patrick, P., Schmid, M. & Zwaan K. (Eds.), *Language analysis for the determination of origin: current perspectives and new directions* (pp.71–89). Cham: Springer. https://doi.org/10.1007/978-3-319-79003-9_5

Goodman, J., Crema, E., Nolan, F., Cohen, E., & Foley, R. (2021). Accents as honest signals of in-group membership. *Cambridge Open Engage*. https://doi.org/10.33774/coe-2021-c9zhv

Grabe, E., Post, B. & Nolan, F. (2001). *The IViE Corpus*. Department of Linguistics, University of Cambridge. http://www.phon.ox.ac.uk/files/apps/IViE/index.php

Haidt, J. (2012). *The Righteous Mind: why good people are divided by politics and religion*. New York: Pantheon Books.

Harrison, K.D. & Raimy, E. (2007). Language as an Emergent System. *Soundings: An Interdisciplinary Journal*, 90(1/2), pp. 77-90. https://doi.org/10.2307/41179192

Hedegard, H. (2015). *Language Analysis for the Determination of Origin (LADO): The Native Speaker Dimension.* MSc dissertation, University of York.

Heine, B. & Nurse, D. (2000). Introduction. In Heine, B. and Nurse, D. (Eds.), *African Languages: An Introduction* (pp. 1-10). Cambridge: Cambridge University Press.

Hellmuth, S. (2013) Phonology. In Owens, J. (Ed.), *The Oxford Handbook of Arabic Linguistics* (pp. 45-70). Oxford: Oxford University Press.

Hellmuth, S. (2019). Prosodic Variation. In Al-Wer, E. &Horesh, U. (Eds.), *The Routledge Handbook of Arabic Sociolinguistics.* London: Routledge.

Hellmuth, S. (2020). Contact and variation in Arabic intonation. In Lucas, C. & Manfredi, S. (Eds.), *Arabic and contact-induced change* (pp. 583–601). Berlin: Language Science Press.

Hellmuth, S. & Almbark, R. (2017). Intonational Variation in Arabic Corpus. Essex: UK Data Service. https://doi.org/10.5255/UKDA-SN-852878

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences,* 33(2-3), pp. 61-83. https://doi.org/10.1017/S0140525X0999152X#

Holes, C. (2012a). Kuwaiti Arabic. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Holes, C. (2012b). Omani Arabic.In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic* Language and Linguistics (online edition). Brill Online.

Holes, C. (2018). *Arabic Historical Dialectology*. Oxford: Oxford University Press.

Hollien, H. (2002). *Forensic Voice Identification*. New York: Academic Press.

Hoskin, J.A. (2018). Native speaker non-linguists in LADO: an insider perspective. In Nick, I.M. (Ed.), *Forensic Linguistics: Asylum-seekers, Refugees and Immigrants* (pp. 23-40). Malaga: Vernon Press.

Hoskin, J.A., Cambier-Langeveld, T. & Foulkes, P. (2020). Improving objectivity, balance and forensic fitness in LAAP: a response to Matras. *International Journal of Speech, Language and the Law,* 26(2), pp. 257–277. https://doi.org/10.1558/ijsll.39208

Home Office (2018). *Language analysis, version 21.0.* Retrieved 17 August 2022 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685203/Language-analysis-AI-v21.0EXT.pdf

Hubbuch, P. (2019). "We Only Want to Talk...": Lingua Interviews for Linguistic Analyses for the Determination of Origin (LADO). In Patrick, P., Schmid, M. & Zwaan, K. (Eds.), *Language analysis for the determination of origin: current perspectives and new directions* (pp. 41-59). Cham: Springer.

Hughes, V. Trudgill, P. & Watt, D. (2012). *An Introduction to Social and Regional Varieties of English in the British Isles* (5ᵗʰ Edition). London: Routledge.

IDEA: International Dialects of English Archive (n.d.).*England*. https://www.dialectsarchive.com/england.

International Association for Forensic Phonetics and Acoustics (2009). *Language and Determination of National Identity Cases Resolution*. Retrieved 14 July 2022 from http://www.iafpa.net/about/resolutions/.

Jacquemet, M. (2009). Transcribing refugees: the entextualization of asylum seekers' hearings in a transidiomatic environment. *Text & Talk*, 29(5), pp. 525-546.

Jamestown Foundation (2012). Syrian Tribal Networks and their Implications for the Syrian Uprising, *Terrorism Monitor*, 10(11). Retrieved 20 November 2020 from https://jamestown.org/program/syrian-tribal-networks-and-their-implications-for-the-syrian-uprising/

Jamestown Foundation (2013). The Tribal Factor in Syria's Rebellion: A Survey of Armed Tribal Groups in Syria. *Terrorism Monitor*, 11(13). Retrieved 20 November 2020 from https://www.refworld.org/docid/51cd4b024.html

Jastrow, O. (2012a). Anatolian Arabic. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Jastrow, O. (2012b). Iraq. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

JASP Team (2022). *JASP* (Version 0.16.3) [Computer software]. https://jasp-stats.org/

Johnson, M.N.P. & McLean, E. (2020). Discourse Analysis. In Kobayashi, A. (Ed.) *The International Encyclopedia of Human Geography* (2nd Edition) (pp. 377-383). New York: Elsevier.

Keating, P. (2005). *D-prime (signal detection) analysis*. Retrieved 9 October 2019 from http://phonetics.linguistics.ucla.edu/facilities/statistics/dprime.htm

Khattab, G. & Foulkes, P. (in press) Arabic sociophonetics. To appear in Strelluf, C. (Ed.), *The Routledge Handbook of Sociophonetics*. London: Routledge.

Klein, W. (1988). The Unity of a Vernacular: Some Remarks on "Berliner Stadtsprache". In Dittmar, N., & Schlobinski, P. (Eds.), *The Sociolinguistics of Urban Vernaculars: Case Studies and Their Evaluation*, (pp. 147-153). Berlin: Walter de Gruyter.

Kudera, J., Stenger, I., Möbius, B., Avgustinova, T., & Klakow, D. (2022). Phonetic Cues in Auditory Identification of Bulgarian, Czech, Polish, and Russian Language of Origin. *Language and Speech*. https://doi.org/10.1177/00238309221119098

Lähteenmäki, M. (2004). Between Relativism and Absolutism: Towards an Emergentist Definition of Meaning Potential. In Bostad, F., Brandist, C. Evesen, L. & Faber, H. (Eds.), *Bakhtinian Perspectives on Language and Culture: Meaning in Language, Art and New Media* (pp. 91–113). London: Palgrave.

Language and National Origin Group (2004). Guidelines for the use of language analysis in relation to questions of national origin in refugee cases. *International Journal of Speech, Language and the Law,* 11(2), pp. 261-266. https://doi.org/10.1558/ijsll.v11i2.261

Lentin, J. (2012). Damascus Arabic. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

LINGUA (2009). *SPRAKAB, Scandinavian Language Analysis, Sweden*. Retrieved 22 July 2022 from https://www.sem.admin.ch/sem/en/home/publiservice/service/sprachanalysen/workshop_2008/programm/sprakab.html

LINGUA (2021). *What is LINGUA?* Retrieved 14 July 2022 from https://www.sem.admin.ch/sem/en/home/publiservice/service/sprachanalysen/lingua/wer_ist_lingua.html

Lucas, C. (2017). *Determining the origin of asylum claimants through language analysis: The case of Arabic*. Retrieved 28 July 2022 from https://www.youtube.com/watch?v=Lbgpoahbh30

Lucas, C. & Ismail, H. (2018). How and why to test lexicon in LADO for Arabic. Paper presented at the *27th Annual Conference of the International Association for Forensic Phonetics and Acoustics*. Huddersfield, England.

McGregor, G. (2001). Review of Folk Linguistics, Nancy A. Niedzielski and Dennis R. Preston, Berlin, Germany: Mouton De Gruyter, 1999. *Journal of Language and Social Psychology*, 20(4), pp. 480-83.

McNamara, T., Verrips, M. & van den Hazelkamp, C. (2010). Language testing, validity and LADO. In Zwaan, K, Muysken, P., & Verrips, M. (Eds.), *Language and origin: The role of language in European asylum procedures: A linguistic and legal survey* (pp. 61–72). Nijmegen: Wolf Legal Publishers.

Markham, D. (1999). Listeners and disguised voices: The imitation and perception of dialectal accent. *International Journal of Speech, Language and the Law,* 6(2), pp. 290–299. https://doi.org/10.1558/sll.1999.6.2.289

Maryns, K. (2010). Identifying the asylum speaker: reflections on the pitfalls of language analysis in the determination of national origin. *International Journal of Speech, Language and the Law,* 11(2), pp. 240-260.

Maryns, K. & Blommaert, J. (2001). Stylistic and thematic shifting as a narrative resource: Assessing asylum seekers' repertoires*. Multilingua,* 20(1), pp. 61-84. https://doi.org/10.1515/MULTI.2001.003

Matras, Y. (2018). Duly verified? Language analysis in UK asylum applications of Syrian refugees. *International Journal of Speech, Language and the Law*, 25(1), pp. 53–78. https://doi.org/10.1558/ijsll.35710

Matras, Y. (2021). Process, tools and agendas in LADO: a rejoinder. *International Journal of Speech, Language and the Law*, 28(2), pp. 233–250. https://doi.org/10.1558/ijsll.20224

Migration Observatory (2022). *Net migration to the UK*. Retrieved 27 August 2022 from https://migrationobservatory.ox.ac.uk/resources/briefings/long-term-international-migration-flows-to-and-from-the-uk/

Milroy, L. (1980). *Language and social networks*. Oxford: B. Blackwell.

Milroy, J. & Milroy, L. (1985). *Authority in language: Investigating language prescription and standardisatio*n. London: Routledge & Kegan Paul.

Moosmüller, S. (2010). The roles of stereotypes, phonetic knowledge, and phonological knowledge in the evaluation of dialect authenticity. *Proceedings of the Workshop "Sociophonetics, at the crossroads of speech variation, processing and communication".* Pisa, Italy.

Moosmüller, S. (2011). Evaluation of the Speech Behaviour of Reference Speakers. *International Journal of Speech, Language and the Law,* 18(2), pp. 179-205. https://doi.org/10.1558/ijsll.v18i2.179

Morgades, S. (2010). The Asylum Procedure in Spain: The Role of Language in Determining the Origin of Asylum Seekers. In Zwaan, K., Muysken, P. & Verrips, M. (Eds.), *Language and Origin: The Role of Language in European Asylum Procedures: Linguistic and Legal Perspectives*(pp. 159-174). Nijmegen: Wolf Legal Publishers.

Morrison, G. S. (2009). Forensic voice comparison and the paradigm shift. *Science and Justice* 49(4), pp. 298–308. https://doi.org/10.1016/j.scijus.2009.09.002

Mufwene, S.S. (2009). Kituba, Kileta, or Kikongo? What's in a name? In *Le nom des langues III: nom des languesen Afrique sub saharienne: pratiques, dénominations, categorizations* (pp. 211-222). Louvain: Peeters Publishers.

Muhammad, U.G. (2021). *A Comparative Phonological Analysis of Varieties of English Spoken by Native Speakers of Nigerian Languages (Hausa, Igbo, Kanuri and Yoruba) for the Determination of Speakers' Origins*. PhD thesis, University of York.

Munro, M. J., Derwing, T. and Burgess, C. (2003). The detection of foreign accent in backwards speech. *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 535–538. Barcelona, Spain.

Naïm, S. (2012). Beirut Arabic. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Neuhauser, S. and Simpson, A. P. (2007) Imitated or authentic? Listener's judgements of foreign accents. *ICPhS XVI Proceedings*, pp.1805-1808. Saarbrücken, Germany.

Nettle, D. and Dunbar, R. I. M. (1997) Social Markers and the Evolution of Reciprocal Exchange. *Current Anthropology*, 38(1), pp. 93-99.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1), pp. 62-85. https://doi.org/10.1177/0261927X99018001005

Nolan, F. (2012). Degrees of freedom in speech production: An argument for native speakers inLADO. *International Journal of Speech, Language & the Law,* 19(2), pp. 263–289. https://doi.org/10.1558/ijsll.v19i2.263.

Nurse, D. (2019). The Case of the Bajunis. In Patrick, P., Schmid, M. & Zwaan, K. (Eds.) *Language analysis for the determination of origin: current perspectives and new directions* (pp. 167-192). Cham: Springer.

Oliver, R., Bjoertomt, O., Greenwood, R. & Rothwell, J. (2008). 'Noisy patients'—can signal detection theory help? *Nature Clinical Practice Neurology*, 4, pp. 306–316. https://doi.org/10.1038/ncpneuro0794

Othman, H. & Saadat, M. (2009). Prevalence of Consanguineous Marriages in Syria. *Journal of Biosocial Science,* 41, pp. 685–692. https://10.1017/S0021932009003411

Owens, J. (2001). Arabic Sociolinguistics. *Arabica,* 48(4)*: Linguistique Arabe: Sociolinguistique et Histoire de la Langue*, pp. 419-469.

Owens, J. (2006). *A Linguistic History of Arabic.* Oxford: Oxford University Press.

Palva, H. (2012). Dialects: Classification. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (pp. 604-613). Leiden: Brill.

Patrick, P. (2010). Language Variation and LADO. In Zwaan, K., Muysken, P. &Verrips, M. (Eds.), *Language and Origin: The Role of Language in European Asylum Procedures: Linguistic and Legal Perspectives* (pp. 73-87). Nijmegen: Wolf Legal Publishers.

Patrick, P. (2012). Language analysis for determination of origin: objective evidence for refugee status determination. In Tiersma, P.M. & Solan, L. (Eds.), *The Oxford Handbook of Language and*

Law (pp. 533–546). Oxford: Oxford University Press.
https://doi.org/10.1093/oxfordhb/9780199572120.013.0039

Patrick, P. (2016). What is the role of expertise in language analysis for determination of origin (LADO)? *International Journal of Speech Language and the Law, 23*(1), 133–139. https://doi.org/10.1558/ijsll.v23i1.27537.

Pinker, S. (2012). *The False Allure of Group Selection*. Retrieved 10 July 2019 from https://www.edge.org/conversation/steven_pinker-the-false-allure-of-group-selection

Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge & Kegan Paul.

Polanyi, M. (1968). Life's Irreducible Structure. *Science*, Vol. 160, No. 3834, pp. 1308- 1312. https://doi.org/10.1126/science.160.3834.1308

Polanyi, M. (1970). Transcendence and Self-Transcendence. *Soundings: An Interdisciplinary Journal*, 53(1), pp. 88-94.

Preston, D. (2002). Language with an Attitude. In Chambers, J. K., Trudgill, P., Schilling-Estes, N. (Eds.), *The Handbook of Language Variation and Change* (pp. 40-66). Oxford: Blackwell Publishers.

Preston, D. (2019). Trouble in LADO-Land: How the Brain Deceives the Ear. In Patrick, P., Schmid, M. & Zwaan, K. (Eds.) *Language analysis for the determination of origin: current perspectives and new directions* (pp. 131-153). Cham: Springer. https://doi.org/10.1007/978-3-319-79003-9_5#

Pretto, C. (2010). The Use of Language Analysis in the Italian Asylum Procedure. In Zwaan, K., Muysken, P. &Verrips, M. (Eds.), *Language and Origin: The Role of Language in European Asylum Procedures: Linguistic and Legal Perspectives* (pp. 187-197). Nijmegen: Wolf Legal Publishers.

Procházka, S. (2012). Prepositions. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Procházka, S. (2018). The Northern Fertile Crescent. In Holes, C. (Ed.) *Arabic Historical Dialectology* (pp. 259-292). Oxford: Oxford University Press.

Prokofyeva, T. (2018) Linguistic origin identification in focus: theory and practice in LOID. In Nick, I. M. (Ed.), *Forensic Linguistics: Asylum-Seekers, Refugees and Immigrants* (pp. 41–55). Malaga: Vernon Press.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Reath, A. (2004). Language analysis in the context of the asylum process: Procedures, validity, and consequences. *Language Assessment Quarterly,* 1(4), pp. 209–233. https://doi.org/10.1207/s15434311laq0104_2.

Reich, D. (2018). *Who We Are and How We Got Here*. Oxford: Oxford University Press.

Richerson, P.J. & Boyd, R. (2010). Why Possibly Language Evolved.*Biolinguistics* 4 (2-3), pp. 289-306. https://doi.org/10.5964/bioling.8793

Rose, P., & Morrison, G. S. (2009). A response to the UK Position Statement on forensic speaker comparison. *International Journal of Speech, Language and the Law*, 16 (1), pp. 139–163. https://doi.org/10.1558/ijsll.v16i1.139

Schmid, M. (2019). Language Attrition as a Problem for LADO. In Patrick, P., Schmid, M. & Zwaan, K. (Eds.) *Language analysis for the determination of origin: current perspectives and new directions* (pp. 155-165). Cham: Springer.

Schumacher, E. F. (1977). *A guide for the perplexed*. Abacus: London.

Shen, C. & Watt, D. (2015). Accent categorisation by lay listeners: which type of "native ear" works better? *York Papers in Linguistics* 2(14), pp. 106-129.

Shorrocks, G. (1998) *A grammar of the dialect of the Bolton area*. Frankfurt am Main: Peter Lang.

Singler, J. (2004). The "linguistic" asylum interview and the linguist's evaluation of it, with special reference to applicants for Liberian political asylum in Switzerland. *International Journal of Speech, Language and the Law*, 11(2), pp. 222–239. https://doi.org/10.1558/ijsll.v11i2.222

Sprakab (2022). *Our Services*. Retrieved 21 July 2022 from https://www.sprakab.se/language-analysis/our-services/

Stansislaw, H. & Todorov, N. (1999). *Calculation of signal detection theory measures*. *Behavior Research Methods, Instruments, & Computers*. 31(1), pp. 137-149. https://doi.org/10.3758/BF03207704

Stanley, J. & and Williamson, T. (2001). Knowing How. *The Journal of Philosophy*, 98(8), pp. 411-444.

Stevenage, S., Clarke, G. & McNeill, A. (2012). The "other-accent" effect in voice recognition. *Journal of Cognitive Psychology*, 24(6), pp. 647-653, https://doi.org/10.1080/20445911.2012.675321

Sumner, M. & Samuel, A.G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60, pp. 487–501. https://doi.org/10.1016/j.jml.2009.01.001

UN (2015). *World Population Prospects: Data Booklet*. Retrieved 24 October 2018 from https://population.un.org/wpp/Publications/Files/WPP2015_DataBooklet.pdf

UNHCR [United Nations High Commissioner for Refugees] (2019). *Handbook and Guidelines on Procedures and Criteria for Determining Refugee Status under the 1951 Convention and the 1967 Protocol Relating to the Status of Refugees*. Retrieved 21 July 2022 from https://www.refworld.org/docid/4f33c8d92.html

UNHCR Data Portal (2022). Ukraine Refugee Situation. Retrieved 10 August 2022 from https://data.unhcr.org/en/situations/ukraine

Verified (2012). *LOID – Method and Limitations.* Retrieved 17 June 2019 from

https://cms.qz.com/wp-content/uploads/2015/09/metod-och-begrc3a4nsningar-loid-ho.pdf

Verrips, M. (2010). Language analysis and contra-expertise in the Dutch asylum procedure.
International Journal of Speech, Language and the Law, 17(2), pp. 279–294.

https://doi.org/10.1558/ijsll.v17i2.279

Verrips, M. (2011). LADO and the pressure to draw strong conclusions: A response to Tina
Cambier-Langeveld. *International Journal of Speech, Language and the Law,* 18(1), pp. 131–143.

https://doi.org/10.1558/ijsll.v18i1.131

Versteegh, K. (2001). *The Arabic Language*. New York: Columbia University Press

Vuorre, M. (2017). *Bayesian Estimation of Signal Detection Models.* Retrieved 9 October 2019
from https://sometimesir.com/posts/2017-10-09-bayesian-estimation-of-signal-detection-
theory-models/

Wakelin, M. F. (1977) *English Dialects: An Introduction*, (Revised Edition). London: The Athlone
Press.

Watson, K. & Clark, H. (2013). How salient is the NURSE/SQUARE merger*? English Language
and Linguistics*, 17(2), pp. 297–323. https://doi.org/10.1017/S136067431300004X

Weinreich, U. Labov, W. & Herzog, M. (1968). Empirical Foundations for a Theory of Language
Change. In Lehmann, W. & Malkiel, Y. (Eds.), *Perspectives on Historical Linguistics* (pp. 95-195).
Austin: University of Texas Press.

Wells, J. C. (1982). *Accents of English, Vol. 2: The British Isles.* Cambridge: Cambridge University
Press.

Wikimedia Commons (n.d.) *Arabic Dialects* (map). Retrieved 11 September 2022 from

https://commons.wikimedia.org/w/index.php?curid=17758938

Wilmsen, D. (2012). Egypt. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Woidich, M. (2012). Cairo Arabic. In Edzard, L. & de Jong, R. (Eds.), *Encyclopedia of Arabic Language and Linguistics* (online edition). Brill Online.

Wilson, K. (2009). *Language analysis for the determination of origin: Native speakers vs. Trained linguists.* MSc dissertation, University of York.

Wilson, K. (2016). *Language Analysis for the Determination of Origin (LADO): An Investigative Study*. MPhil dissertation, University of York.

Wilson, K., & Foulkes, P. (2014). Borders, variation, and identity: Language analysis for the determination of origin (LADO). In Watt, D. & Llamas, C. (Eds.), *Language, Borders and Identity* (pp. 218-229). Edinburgh: Edinburgh University Press.

Yarmey, A.D. (2012) Factors Affecting Lay Persons' Identification Of Speakers. In Tiersma, P. M. & Solan, L. M. (Eds.), *The Oxford handbook oflanguage and law* (pp. 533–546). Oxford: Oxford University Press.

Yorkshire Dialect Society (n.d.). *The Yorkshire Dialect Border*. Retrieved 10 September 2019 from https://www.yorkshiredialect.com/Border%20text.htm

Zebian, S. & Denny, J.P. (2001). Integrative Cognitive Style in Middle Eastern and Western Groups: Multidimensional Classification and Major and Minor Property Sorting. *Journal of Cross-Cultural Psychology*, 32(1), pp. 58-75. https://doi.org/10.1177/0022022101032001008

Zwaan, K. (2010). Dutch Court decisions and language analysis for the determination of origin. In Zwaan, K., Muysken, P. & Verrips, M. (Eds.), *Language and Origin: The Role of Language in European Asylum Procedures: Linguistic and Legal Perspectives* (pp. 215-224). Nijmegen: Wolf Legal Publishers.