THE FOUNDATIONS OF PUBLIC LANGUAGE: WORDS AS SOCIAL ARTEFACTS

Nicholas Mark Tasker

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion and History of Science

July 2017

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2017 The University of Leeds and Nicholas Mark Tasker

The right of Nicholas Mark Tasker to be identified as Author of this work has been asserted by in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

My supervisors while writing this thesis were Robbie Williams, Gail Leckie and Paolo Santorio. Their extensive comments on drafts have been essential. I'm lucky to have been advised by three excellent philosophers who were generous with their time, extremely professional, and who helped me say what I wanted to say. I don't know how to thank them enough.

For four years I have been part of the *Nature of Representation* research project. This has been a privilege in so many ways. Thanks to Robbie Williams for engineering this particular social entity. The project – and through it, my research – was funded by the *European Research Council.* In my fourth year I was funded by a Jacobsen Studentship from the *Royal Institute of Philosophy*. I'm extremely grateful for these opportunities.

There are many wonderful philosophers in Leeds. Thanks to all of them, and to the many academic visitors who've given talks or come to conferences, and to the grad students, who are ace.

In 2016 I organised a conference on the notion of mental representation in generative linguistics. This was made possible by the *Nature of Representation* project and graduate attendance was supported by the *Analysis Trust.* The speakers were Frances Egan, Georges Rey, John Collins, Wolfram Hinzen, Ianthi Tsimpli and David Adger. A few months later some of us were reunited at another event on the same topic at the *Norwegian Summer Institute in Language and Mind* at *CSMN*, Oslo. These were formative experiences.

In June this year Gail Leckie organised a conference on the metaphysics of words, at which I gave a talk based on chapter three of this thesis. Also speaking were Linda Wetzel, Lee Walters, Mark Richard, Jess Pepp, Robert May and Stefano Predelli. Thanks to them for the discussions, as well as to the conference delegates.

Finally, for supporting me personally, thanks to my partner, Stéphanie, as well as to Anne, Brian, Carol, and all the family members who don't make it into this thesis for alphabetical reasons.

ABSTRACT

This thesis brings together topics in philosophy of language, social ontology, and generative linguistic theory. The first main contribution is to develop a theory of artefacts, and to apply it to linguistic entities. The general account of artefacts I offer here draws heavily on Amie Thomasson's work, but I go on to isolate a class of artefacts which I refer to as *essentially communicative artefacts*, ECAs, and I argue that words fall into this category. One benefit of this approach is that insights arising from social ontology can be used to remedy deficiencies in philosophical discussions of words: for example, I show why the failure of *form-theoretic* approaches to word individuation poses a significant obstacle to attempts to deploy Searleian assumptions about social ontology in a theory of words. The second main contribution is to provide an account of public language which is compatible with developments in generative linguistics. Too often, philosophical discussions of words ignore conceptions of language which are prevalent in linguistics, which means that fruitful connections between the disciplines are missed, and that worries expressed by linguists about philosophical conceptions of public language go unanswered. My account of words is intended not only to be compatible with generative linguistic theory, but also to be thoroughly embedded in the philosophy of science and mind which animates generative linguistic theorising. From this vantage point, I evaluate a range of sceptical arguments which have been levelled against public language views. I conclude that what the philosophy of generative linguistics recommends is not an eliminativist position with respect to public language, nor a naively scientist one, but a practical, principled, methodological preference. A third contribution is to provide original objections to extant theories of words, including those due to David Kaplan, Herman Cappelen, and Zoltan Szabo.

CONTENTS

ABBREVIATIONS

ECA   Essentially Communicative Artefact

FLB   Faculty of Language: Broad sense

FLN   Faculty of Language: Narrow sense

FT   The Form-theoretic approach to word individuation

NRP   Naturalistic Reality Principle (due to Collins)

RA   The Recognisability Argument (for the FT approach)

SLE   Standard Linguistic Entity

INTRODUCTION

Public Language

In philosophy, language is often thought of as a mind-external, public phenomenon. Words and sentences are assumed to be kinds of utterances and inscriptions. Their instances are concrete, physical entities found in books, on billboards and in acoustic blasts emitted during speech. After all, we are not telepathic. We cannot communicate, give orders or make promises just by beaming our thoughts from one mind directly into another. Linguistic communication among humans requires speakers to modify their physical environments in ways which are detectable by their interlocutors. Central to this picture is a view which takes a public language to be a system of concrete, mind-external signs which people know and use. An individual may have only a partial grasp of the language spoken in their linguistic community, but each person's grasp is of the same sign system, since one of the functions of the sign system is to facilitate coordinated communication among speakers.

Public languages are typically taken to depend on speakers' communicative intentions and on conventions which are maintained within a linguistic community. They are therefore regarded as social entities, aspects of the manifest image. Indeed, public languages give every sign of being thoroughly cultural entities. After all, it is perfectly true that the English language was shaped by, among other things, the Norman invasion of Britain, the work of Chaucer and Shakespeare, the invention of the printing press, dictionaries, etc. One thinks also of language as a social phenomenon in considering the impact of the French government on the marginalisation of regional languages in France in the 19th century, and the influence of bodies such as the *Académie Française*. Language, in this sense, is something which is specific to a community, not to the species. Finally, language is generally understood to be the focus of a welter of research in philosophy and the social sciences.

I-language

A contrasting picture emerges from generative linguistics. There, the explanatory focus is targeted inward on the cognitive processes which underlie and explain aspects of our linguistic behaviour. It is assumed that linguistic facts are ultimately facts about individual language users, not about some mind-independent linguistic reality. To explain why humans acquire languages with specific properties (while monkeys and rocks do not) even without formal instruction, it is assumed that the human mind is innately configured in certain specific ways. The task of linguistics is then to understand what properties human mind/brains have to have in order to explain our remarkable ability to systematically pair sounds with meanings over an unbounded range.

On this approach, the goal is not to provide a theory of communication, or a theory of why people say what they say. For general reasons pertaining to scientific methodology, an attempt is made to abstract away from the messy detail of our linguistic behaviour in order to make progress. Notably, generative linguistics has long been understood as idealising away from facts about linguistic performance in order to get a theoretical grip on linguistic competence, the relatively stable state which underlies the human capacity to pair sound and meaning in systematic ways. According to modern jargon, this capacity of an individual is labelled *I-language*:

> The word "language" has highly divergent meanings in different contexts and disciplines. In informal usage, a language is understood as a culturally specific communication system (English, Navajo, etc.). In the varieties of modern linguistics that concern us here, the term "language" is used quite differently to refer to an internal component of the mind/brain (sometimes called "internal language" or "I-language"). We assume that this is the primary object of interest for the study of the evolution and function of the language faculty. (Hauser, Chomsky, Fitch, 2002: 1570)

An individual's I-language is thus understood as a subconscious capacity to pair sounds and meanings:

> So understood, competence designates the capacity to pair discrete, structured messages or meanings with sounds (or some other vehicle, such as hand gestures in sign language) over an unbounded range. The science is concerned with the structure and development of this capacity; its integration with other cognitive capacities; its realization in the human brain; and how it has evolved in the human species alone. (Collins, 2010:46)

The hypothesis that there are I-languages should be distinguished from the hypothesis that there is such a thing as *Universal Grammar*. The study of Universal Grammar is not the study of I-languages as such, but of the initial state of the language faculty which is part of an individual's genetic endowment, and is assumed to be more or less the same across individuals throughout the species. An individual's linguistic competence develops over time in response to their exposure to spoken language, before eventually reaching a relatively stable state. An I-language can be construed as a state of linguistic competence which characterises an individual at a specific stage of development. An individual's I-language is a natural variation on a species property. I-languages are thus natural objects, not socially constructed ones.

Evidence for this view of linguistic competence comes from a variety of sources including, the well-known *Poverty of Stimulus* argument, the existence of linguistic universals, the process of creolisation, over-regularisation errors, language pathologies, the ubiquity of language in humans and its absence in other species,[1] as well as evidence that linguistic processing is associated with specific parts of the brain. Pursuing the inquiry into language as an aspect of human psychology and, ultimately, biology has yielded theories which are precise, explanatory and which have achieved a degree of integration with other successful sciences.

Theorists in the generativist tradition insist that inquiry into language is best served by pursuing this internalist perspective. Some put the point more forcefully, holding that there is *no such thing* as public language. This eliminativist attitude is contentious. Even many Chomskians recognise that eliminativism

---

[1] 'Language' is used here in a sense which is not equivalent to 'communication'.

about mind-external linguistic phenomena is not strictly mandated by the success of generative linguistics. Instead, a more moderate stance is typically adopted which simply denies that externalist conceptions of language enter into serious naturalistic inquiry.

Language as biology and culture

In short, while it seems obvious that language is partly a cultural phenomenon, we have an empirically successful science which views language as a narrow aspect of human biology. Some linguists and philosophers see the foregoing contrasts as amounting to an unresolvable conflict. One of the things I want to do in this thesis is to show that the two approaches are compatible and complementary, though they reflect different theoretical interests and priorities.

One way of seeing this is to view language as analogous to such phenomena as love, jealousy, fear, disease, or disability. These are phenomena which have both biological and cultural facets. Similarly, take the example of sexual reproduction. Biologists have a technical conception of the biological processes involved in sexual reproduction including its role in natural selection. But sexual reproduction, in humans at least, is not only a biological phenomenon. It can be studied from the point of view of the social sciences. For example, sexual reproduction in the broad sense is a phenomenon of interest to experts from a variety of disciplines, right up to the sociologist or historian of the institution of marriage.

If we are scientific realists of one stripe or another – as I am – then we ought to believe in I-languages. This discovery illuminates crucial aspects of natural language, what it is, how individuals acquire it, and enables us to ask questions about how it might have evolved. But there are aspects of language – the ways people use it, the role it plays in communication, how it unites people under a common identity or alienates them, how it can be used to persuade or to denigrate and slur, how it can be deployed in literature – which are not addressed by the biological approach pursued in the cognitive science of linguistics.

Generative linguistics quite rightly seeks to abstract away from such parochial, human centric, norm-governed phenomena, just as the biology of sexual reproduction abstracts away from the details of marriage law in seventeenth century Scotland. But philosophy can afford to treat language as a multi-faceted phenomenon which cross-cuts distinctions such as 'biological vs cultural,' 'natural vs social' or 'mental vs non-mental'. It turns out, I think, that I am broadly in agreement with people like Chomsky about the phenomenon of public language, its relations to human intentions, etc. We differ only in our levels of enthusiasm for and optimism about a programme in philosophy of language which continues to pursue the phenomenon of public language in the social ontological vein.

Words

My choice to focus on words does not reflect any prejudice against larger units such as sentences. Sometimes I extend the discussion to units such as sentences, phonemes, etc., but I mostly confine my attention to words in order to avoid needless complexity. A further reason to focus on words is that they are of independent interest to specific areas of inquiry in philosophy, such as referential semantics (where a referential semantic theory is understood to involve relations between words and the world) or work on slurs. Moreover, a small literature on words has emerged in philosophy of language, especially since Kaplan (1990). Since this literature has rarely been integrated with topics in social ontology on the one hand and generative linguistics on the other, this thesis is intended to do just that.

Thinking about words is in equal measures fascinating and infuriating in light of the fact that the notion of a word belongs at once to everyday, common sense ways of understanding the world, and to more theoretical discourse. Worse, there is no single notion of word. Both common sense and theory distinguishes different senses of word. Worse still, common sense and theoretical conceptions have intermingled to a certain degree. Sometimes this is because some linguist or

philosopher over the years has taken on some common sense notion and put it to work in a theoretical context; sometimes it is because ordinary people are exposed to theoretical conceptions of language (through education, television, etc.)

What a mess! With such a plethora of ways of thinking about language and of candidates for what we might be thinking about when we do, it is impossible not to feel confusion when trying to answer the question, 'what is a word?'. Philosophers, linguists and the folk use a bewildering variety of word-like concepts. It's a paradigmatically philosophical task to try to clarify such concepts, and to work out what the relations between them are. That is the task partially undertaken in this thesis. Separating out theoretical and common sense conceptions of words and inquiring into the relations between them is a kind of Sellarsian task of showing "how things in the broadest sense of the term hang together in the broadest sense of the term" (Sellars, 1963:1). In this thesis I try to pick apart the tangled relations between the theoretical entities posited in linguistics and the linguistic objects which furnish our everyday lives.

Speaking, writing, signing

I have chosen to include writing in the theory of words, a choice which is standard in the philosophy of language literature but not free of controversy. For example, echoing Aristotle, Stebbing writes:[2]

> Sound-tokens are historically prior to shape-tokens; it is in this sense that spoken words are more fundamental than written words. Perhaps Aristotle was right in saying that the latter are symbols of the former. I think, however, that it is better to say that shape-tokens represent sound-tokens. (Stebbing, 1935:7)

Stebbing's idea is that for a given language there is only one fundamental medium in which words are *instantiated* (speech or sign, etc.); inscriptions are *mere*

---

[2] The passage from Aristotle which Stebbing cites is as follows: "Spoken words…are symbols of affections of the soul and written words are symbols of spoken words." (Aristotle, De Interpretatione 1, 16a3-8 as translated in Modrak, 2001).

representations of words.[3] One might wonder why the *historical* priority of speech over writing entails anything of import about their metaphysics. Nonetheless, Stebbing is in auspicious company. Saussure also maintains that written words are mere representations of spoken ones, and that linguistic theory should be concerned only with the latter:

> Language and writing are two distinct systems of signs; the sole purpose for the existence of the second is that of representing the first. The object of linguistics is not defined in terms of both the written and spoken forms of words; the latter alone constitute its focus. But the written word is so closely associated with the spoken word of which it is the image that it manages to usurp the main role. People attach more importance to the representation of a vocal sign than to the sign itself. It is as if someone thought that a better way of getting to know someone were to look at his photograph rather than view him directly. (Saussure, 1916/1983:45)

It should be noted that linguists typically do set writing to one side, insisting that speech is more basic and more universal than writing, with the latter being obviously a cultural and intellectual achievement passed on through explicit instruction. Almost everybody who can read and write can speak or sign, but huge numbers of people who can speak or sign cannot write, especially when we consider all the humans that have ever lived. Most languages past and present don't even have a writing system, but there are no natural languages which have only a writing system and are never spoken or signed. Writing appears therefore to be something of a secondary medium of communication. Moreover, the capacity to speak is acquired by infants without formal instruction and the development of this capacity proceeds in stages which are uniform across all cultures. There is also significant evidence for the existence of a critical period for language acquisition such that if a child is not exposed to language use before roughly twelve years old they will never acquire a facility with speech comparable to that of a competent native speaker. In contrast, the ability to read and write is

---

[3] Stebbing's position should not be confused with Szabo's (2000) view, discussed in chapter four (§2) that not only inscriptions but also utterances are mere representations of words, with the latter considered as abstract particulars. For Stebbing, utterances genuinely instantiate words.

learned through more or less formal education and can be acquired at any stage in life.

Nevertheless, philosophers typically take it that words can be tokened in speech or in writing. My stance on all this is as follows. I'll be trying to develop a theory of words which allows that words can be instantiated in both utterances and inscriptions. If that were to prove impossible it would be comforting to know that we could on principled grounds exclude inscriptions from the picture and give a theory of words purely as kinds of spoken tokens. But we should set our sights higher than that. A theory which treats utterances and inscriptions on all fours promises to contribute to an understanding of a wider range of linguistic behaviour, and is less revisionary of common sense attitudes to language. Moreover, it turns out there's a perfectly good sense in which inscriptions instantiate words, with many strong comparisons between utterances and inscriptions. There are some differences, and where these matter I'll flag them as we go.

I will exclude sign languages from the present inquiry, though the subject is fascinating. I believe my account applies equally well to sign languages, though many of the details are somewhat different. Since there's no theoretical payoff in keeping the extra balls in the air, I'll maintain simplicity and leave sign languages out.

Since it is tediously verbose to keep stating that a certain claim applies to both utterances and inscriptions, I will in places speak only of utterances, except where making the wider scope of the claim clear is particularly important. I will also, in places, use the neutral term 'token'.)

Types and tokens

Philosophers frequently describe the word-utterance relation in terms of *types* and *tokens*: for example, an inscription of 'alone on a wide wide sea' has six word tokens of five different word types. Here, I interpret this to mean that word

types are *kinds of utterances and inscriptions*, and that word tokens *instantiate* word types in virtue of having properties associated with the type. This is not the only way of understanding type-token talk. Quine (1987:216-219) holds that a word type is the class of its tokens. Kaplan (1990) thinks of words as concrete individuals made up of utterances and inscriptions (and some mental stuff), while Szabo (1999) considers words to be abstract particulars represented but not instantiated by word tokens. These metaphysical choices are largely orthogonal to present concerns, though I will I will discuss Kaplan's and Szabo's ideas in chapter four.

To be clear, I assume that word tokens *instantiate* word types. In general, I make no distinction here between types and tokens on the one hand and kinds and instances on the other, and for convenience I will employ both sets of vocabulary and treat them as equivalent. I'm not denying that there could be useful metaphysical distinctions between types and kinds, but I don't think they will show up in this thesis.

Two accounts of words

Kaplan (1990) is an important milestone in framing the debate about words, spawning a small literature. He asked two main questions: What are words? What makes two utterances utterances of one word? Noting that having a certain form is neither necessary nor sufficient for being a token of a given word, Kaplan argued for a kind of *intentionalism*. According to his view, whether or not something is a token of a given word depends on whether it was produced with the right intention in mind. Cappelen (1999), on the other hand, rejects intentionalism, insists on a view of words as kinds of sounds and shapes, and embeds his view within Searle's social ontological theory involving collective acceptance of rules of the form *X counts as Y (in context C)*. According to both of these views, words have a kind of mind-dependence. They depend constitutively on intentions (Kaplan) or conventions (Cappelen).[4]

---

[4] Cappelen appears to think of conventions as involving *collective agreement*.

As I see it, these accounts of words have two serious deficiencies. First, they have not been accompanied by detailed attempts to incorporate insights from social ontology. The models of social construction employed are underspecified or have flaws which make them unable to do the work they are supposed to do. Second, they have been worked out largely in isolation from research in generative linguistics. This is a shame because (i) Chomskian complaints about philosophers' talk of public language are not responded to, and (ii) insights from generative linguistics which can be fruitfully put to work in a theory of public language go overlooked.

The standard model of social ontology

Cappelen's and Kaplan's approaches are broadly within what is sometimes called the *standard model of social ontology*, according to which many social objects are taken to be partially constituted by human mental representations. I argue that this is a fruitful framework for thinking about words, though there are pitfalls and controversies which we need to avoid or take a stance on. So one of the main things I undertake to do in this thesis is to set out a decent framework drawing on the standard model.

I'll leave out approaches in social ontology which are outside the standard model, e.g. Millikan (1984), Elder (2007). This is an attempt to give a theory of words from within the standard model of social ontology. If it should turn out that the various tools of the standard model (intentions, conventions, etc.) are not up to the job of giving a theory of words, then that would be a result of some importance in general social ontology. In fact, the standard model has the resources to provide a pretty good theory of words, at least when allied with a sensitivity to some of the insights of modern linguistics. It also enables us to interpret the Chomskian complaints about public language in a way which is reasonable and worth engaging with.

I develop an account of linguistic entities as social artefacts which depend constitutively on speaker intentions. Drawing on Amie Thomasson's work on

artefacts I argue that part of what makes it the case that an utterance is of a given word is that it is intended to be recognisable as an utterance of that word. The account I develop is independently plausible as an account of certain kinds of artefacts, and it harmonises beautifully with certain interpretations of linguistic theory. My view therefore remedies two significant deficiencies of Cappelen's and Kaplan's views.

What is common to approaches in the standard model of social ontology is analogous to what is sometimes called the head-first approach to the problem of intentionality. How is it that an acoustic blast can be *about* cats? In philosophical discussions of intentionality, it's fairly common to say that a spoken utterance has only *derived intentionality*, as opposed to the *original intentionality* which characterises human mental states. The idea is that human beliefs and intentions are somehow responsible for the intentional profile of the utterance. To point out that word tokens get their meanings via human intentions is not, of course, to resolve the problem of intentionality. Assuming we can make sense of how a meaning can be conferred on an inanimate object by intention, we only manage to explain the semantic properties of the external object in virtue of mental states which do themselves have semantic features. As Putnam notes:

> [T]o have the intention that anything, even private language (even the words 'Winston Churchill' spoken in my mind and not out loud), should represent Churchill, I must have been able to think about Churchill in the first place. (Putnam, 1981:4)

This *head-first* approach to intentionality is fairly common. As noted, it just pushes the issue of intentionality back a step. Philosophical attempts to naturalise intentionality have focussed their efforts on mental representation, including notably Dretske (1981) and Millikan (1984). Though naturalistic efforts to reduce mental representation have not been entirely successful, I will continue to assume that we can employ contentful mental states in explanations of the intentionality of external objects. Though we cannot complete what Rey (1996) calls the *downward project* of providing a reductive account of mental representation, we can undertake the *horizontal project* which involves pointing out the explanatory

roles played by posited mental representations in cognitive science and everyday life. The head-first approach puts the intentionality of mental states like beliefs and intentions to work explaining the linguistic properties of mind-external words.

Objectives

My aims in this thesis are as follows:

(i) I want to uphold a scientific realist attitude regarding linguistic theory. I do not seek to censure the practice of linguistics or to impose a revisionary interpretation on it.

(ii) I do want to show that a theory of I-language is not a complete theory of language. (Chomskians won't find this surprising or incorrect). Moreover, I want to give a plausible story about how words can be regarded as social artefacts, and argue for a pluralistic stance which makes room for Chomskian "I-languages" and public languages answering to different explanatory roles. My goal is to articulate a notion of public language words as essentially communicative artefacts which preserves common sense notions of language and resists sceptical attitudes towards disciplines – such as philosophy, forensic linguistics, sociology etc. – which inquire into language as a social phenomenon.

(iii) I want to show the deficiencies of certain rival theories of words.

(iv) A key objective is to charitably articulate Chomskian complaints about philosophers' attitudes to public language and show how they can be answered.

The ultimate goal is to achieve peace, love and theoretical integration between philosophers and linguists by giving some content to the notion that language is both biology and culture.

Chapter overview

Chapter 1. Many of the contributions to the topic of words in philosophy of language assume that words are in some sense social objects, perhaps depending on the intentions of speakers or on conventions which hold in a community. It is therefore useful to draw heavily on debates in social ontology in order to establish a framework within which to pursue the inquiry into words. In chapter one I focus on debates about artefacts in recent social ontology. In particular, I develop an account of *essentially intentional artefacts*, instances of which depend constitutively on their creators' intentions. I provide an account of artefact categorisation and attend to a significant class of artefacts which have an essentially communicative or symbolic character: I refer to these as *essentially communicative artefacts*, ECAs. The point is to provide a substantive perspective on artefacts which can be brought to bear on the theory of words in the remainder of the thesis.

Chapter 2. In chapter two I apply the account to words, arguing that the latter can fruitfully be regarded as ECAs. I argue that the ECA view of words can be embedded within the philosophy of science and mind associated with generative linguistics in a way which recruits extra support for the view. In this chapter I also address two major objections which hold, for one reason or another, that the intentions my story needs to get off the ground are not available. In addition, I indicate the kind of approach to questions of word individuation which is suggested by the ECA view of words, and respond to a variety of objections.

Chapter 3. The view that words are kinds of sounds and shapes (the form-theoretic view, FT) remains relatively popular, despite attracting criticism. Cappelen (1999) defends a version of FT which is implemented within the social ontological framework provided by Searle. After outlining the Searle-Cappelen model of words, I spend some time articulating the argument which motivates FT, something often ignored in the literature. I then state the case against FT, which partly involves rehearsing empirical arguments which are well known. In addition, I argue that FT's appeal lies in a tempting but misleading conception of

the processes involved in word recognition. I expose the misconception, thereby undermining FT. I suggest that the demise of FT throws up a problem which hinders any attempt to apply the Searleian social ontological theory to words. Next, a theory of words ought to say something about the formal features of utterances and inscriptions: Kaplan (1990) was wrong to think that form was irrelevant to word type. I give an account of this matter drawing on my theory of words as artefacts. Finally, I make it clear that my account of words is distinct from Cappelen's in important ways, meaning that it is immune to the problems I raise for Cappelen's view.

Chapter 4. Having put most of my positive account in place in the first three chapters, as well as criticising one of the main extant alternatives to my view, I turn in the fourth chapter to providing criticism of two further related views. Kaplan is well known for endorsing an intentionalist theory of words, though his version of the approach is significantly different to mine. I raise a number of problems for his account. In the second part of the chapter I turn to a critique of Szabo's suggestion that words are created abstract particulars, represented but not instantiated by utterances and inscriptions.

Chapter 5. In the final chapter I turn to the various sceptical attitudes taken by theorists in the generative tradition towards public language notions such as that developed in this thesis. First of all I outline the most radical critiques of public language, views according to which words considered as mind-external entities can be eliminated from our ontology. Having criticised these radical positions, I consider the claim that artefacts (and by extension, words) are somehow inapt for entering into scientific inquiry. After showing how these objections can be met I go on to outline what remains of the Chomskian critique of public language. What remains is an agreement between the Chomskians and some philosophers that the arrows of linguistic explanation point from public language inwards to human cognition. What the Chomskians add to this is a methodological scepticism about the prospects for any serious scientific inquiry targeting public language. That critique has to be taken seriously, though it is by

no means the death knell for philosophical inquiry into public language which some have claimed.

CHAPTER 1: ARTEFACTS

Examples of artefacts include chopsticks, knives, Frisbees, cars, computers, tuxedos, houses, apple crumbles, paintings, the Eiffel Tower, one pound coins, crucifixes, paperweights, ballistic missiles, paperclips, tattoos, pacemakers, wedding rings, newspapers, abacuses, handwritten signatures, etc.

That such and such an object is a chopstick or a one pound coin involves, in some sense, facts about people and the attitudes they bear to objects in their environment. Thomasson (2007:52) expresses the view that "artifacts and other social and cultural objects are 'creations of the mind', depending in certain ways on human beliefs or activities." For Thomasson, this dependence is not just causal but constitutive. That an artefact is an object which has been intentionally made or adapted is taken to be a conceptual truth.

The view that artefacts are partially constituted by human intentions is widespread in the field of social ontology. Adherents include Baker (2004), Dipert (1993), Hilpinen (1993), and Khalidi (2015). Within this intentionalist approach to artefacts, Thomasson (2003a, 2003b, 2007, 2009, 2014) has provided one of the most developed accounts currently available. More broadly, these views are situated within what Guala (2007:956) calls "the Standard Model of Social Ontology" which seeks to ground aspects of social reality in human beliefs, attitudes and intentions. One finds variations on themes associated with the standard model entering into the literature on words, e.g. Kaplan (1990), Bromberger and Halle (2000), Barber (2006, 2013), Stainton (2014) and McCulloch (1991).

The purpose of this chapter is to give a theory of artefacts, and in particular a class of artefacts which have an essentially communicative function, in a sense to be explained. To this end I draw heavily on the work of Amie Thomasson. With this substantive account of an important class of artefacts in hand, I'll go on in

chapter two to apply the theory to the case of words. I'll argue that from the perspective developed here, it is natural to consider words as intentional artefactual kinds, and that doing so provides a rich account of the nature of words.

The most immediate benefit is that a theory of words can be embedded in a wider project in social ontology, allowing insights developed there to be applied to the account of words. A secondary benefit is that words provide an interesting test case for the general theory of artefacts. To the extent that the theory can be successfully applied to words, it increases the breadth of its explanatory power. To the extent that it cannot be so applied, we may be able to see ways of refining the theory of artefacts to account for a wider variety of phenomena.

1. INTENTIONALISM ABOUT ARTEFACTS

This section proposes a definition of artefacts, distinguishing them from non-artefacts, and also delineates certain kinds of contrasts within the broad class of artefacts.

1.1 The authorial connection

According to Hilpinen, something is an artefact "only if it is intentionally produced under some description of the object" (Hilpinen, 1992:60). Thomasson adopts a similar characterisation, writing that artefacts are "things that are intentionally made and which have at least some intended features" (Thomasson, 2014:48). On this conception, artefacts are the intentional products of human activity.[5] There is an authorial connection between an artisan and the artefacts they produce: artisans exercise agency in becoming the authors or creators of a new artefact.

---

[5] Crows modify tools, beavers build dams, bowerbirds build bowers. Are these artefacts? Do animals have intentions? I don't know the answers to these questions. It seems fair to bracket such issues and say the account just applies to human artefacts. However, suppose the answer to the second question is negative. Then I think there will be some pressure to allow that *some* artefacts have their natures independently of their creators' intentions, even some human artefacts. Still, I don't think this would undermine my claims about the class of artefacts which are most at issue here (the essentially communicative artefacts). See §5.3 of this chapter for (a little) more discussion.

We can refer to this conception of artefacts as *intentionalism*. The intentionalist defines artefacts in such a way as to rule out, in the first place, naturally occurring objects such as rocks, shells, atoms, planets, etc. To focus on artefacts is to focus on things which are *artificial*. It also excludes what we might call *instruments*, naturally occurring objects used unmodified by some human for a purpose, such as a stick seized upon to beat off an attacker.

What else does the definition exclude? Well, artefacts are not just things which have been made or adapted by humans: pieces of slag produced during iron production can be described as things made by humans, as can broken shards of pottery, but these are not artefacts in the sense of the term we're interested in. Although pieces of iron slag can result from human actions they are not *intended* results of those actions, but are mere by-products. And while the vase was an intentional product, the shards are not. The shard's being made of Chinese porcelain results from human intention in some way, but no one intended the *shard* to be made of porcelain. Assuming the breakage was accidental, no one intended the shard to exist at all. It was the vase that was intended to be made of porcelain. Relatedly, though a village might be constructed intentionally, it could also emerge as an unintended result of the intentional building of individual houses. In the latter case, the village is not really an artefact in our sense, although it might be made up of artefacts.

We have to insist that artefacts be things *made* by humans, because there are non-artefactual objects which have intended features. For example, someone might brand a wild animal to indicate ownership. The brand is an artefact, but the animal is not. The *animal* has not been made by humans. So there are objects which have intended features without being artefacts, according to our definition. Similarly, though I have the property of being seated intentionally, I am not an artefact, because there was no intention to create a seated person.

Finally, not all *social objects* are artefacts in our sense. For example, cowrie shells have long been used as a form of currency in many parts of the world (example due to Epstein, 2013). Relative to such practices, cowrie shells have a certain monetary value even if they have never been encountered by a human. Similarly, undiscovered oil deposits can be the property of some company. Such objects do not owe their socially constructed properties to human attitudes towards their particular *instances*, but to general attitudes regarding those *kinds* of things.

Our definition does and should allow that some natural objects *become* artefacts. For example, if I take a pebble from a beach and place it on some loose papers in an office, then the pebble is a paperweight. I have created a new artefact by taking on an existing natural object without modification. This process is referred to here as *minimal creation*, of which more below. (What is the difference, then, between the paperweight and what I referred to above as 'instruments'? I answer this question in §2.2 of this chapter.)

The definition also includes things like curium atoms and GM tomatoes, since these are generally the intended products of human activity. Now, it's fine that these be treated as artefacts, but we should note that there is an important difference between a curium atom and artefacts such as wheelbarrows, Frisbees, words, etc. Curium atoms are sometimes created intentionally in nuclear reactors, making them artefacts, but their dependence on human intentions is just causal, and not constitutive. Curium atoms are what they are in virtue of their internal essences, and could come to exist independently of human activity. From the perspective of the philosopher, the most interesting artefacts are those which depend constitutively on human intentions, those which are what they are in virtue of a distinctively mental contribution made by their artificers. Nothing, it seems, could be a Frisbee if no one ever intended it to be a Frisbee or regarded it as such. Co-opting a term used by Thomasson, I refer to the artefacts which depend constitutively on human intentions as *essentially intentional artefacts*. Giving an account of this distinctively mental contribution to the characters of essentially intentional artefacts is one of the goals of this chapter.

To summarise, the focus here will be on essentially intentional artefacts. These are things which are intentionally made by humans under some description and which have at least some intended features, at least one of which is partially constituted by human intentions regarding those very objects. Within this class of artefacts, there is room for another distinction: some of them can be characterised as artefacts which have an *essentially communicative function* (perhaps in addition to more practical functions). This distinction roughly corresponds to Searle's distinction between social facts (e.g. that x is a hammer) and institutional facts (e.g. that x is a dollar bill, a wedding ring, or a chess pawn). These communicative artefacts are terrifically interesting in their own right, but especially important here since I'll be arguing in the rest of this thesis that words fall into this category.

Finally, we should note that our definition of artefacts conflicts with some of the ways in which 'artefact' is commonly used. An archaeologist might well say that the piece of iron slag or the shard of pottery is an artefact. Additionally, 'artefact' is usually used to refer to medium sized dry goods. In our sense it can include tiny objects such as the products of nano-engineering, huge objects like the Great Pyramid of Giza, as well less tangible, more ephemeral objects like clouds of poisonous gas or utterances. Perhaps one can also make one's body into an artefact, as may be the case for a semaphorist. These divergences from common usage are of course no objection to the proposed definition. The definition aspires to capture an intuitive concept corresponding to a particularly interesting class of objects; it is not intended to apply to all ordinary uses of 'artefact'.

1.2 Intended functions or intended features?

This way of defining artefacts is extremely widespread. Possibly even more widespread is a stricter definition: it is often held that artefacts are objects with an *intended function*, and that what makes it the case that two artefacts are of the same artefactual kind is that they have the same intended function (Kornblith, 1980, Baker, 2004). But not all artefacts have an intended function. Thomasson

(2014) gives the example of idle doodles. She also points out that it is perfectly conceivable that someone could create a statue which is not intended to have any particular function. Even for artefacts which do have an intended function, it is often the case that having this intended function is neither necessary nor sufficient for being an artefact of that kind: Bloom (1996) notes that a boat might not be intended ever to see water (if it's for show, say); we can add that two groups of adventurers might end up making a boat and a raft respectively, though each artefact was intended to do no more and no less than transport them downriver.

Thomasson thus reasonably argues that we should liberalise the traditional definition of artefacts. Artefacts are objects which are intentionally made and have at least some *features* by intention, where these can include not only functional properties, but also "structural properties, sensory properties (flavour, color, sound), aesthetic properties, and so on" (Thomasson, 2014:49).

1.3 Success conditions

Though necessary, an intention to create an object with properties relevant to membership of a given artefact kind is not on its own sufficient for creating an artefact of that kind. The knife-maker has not only to intend to make a sharp blade, but also to succeed (more or less). If things go badly wrong in the workshop then what gets produced might not even qualify as a badly made knife. Success need not be complete: a bent knife may still be a knife, but this tolerance has limits. If you are barely successful in imposing the features you intend on your product then it may fail to be an instance of the intended artefactual kind.

As Thomasson (2003b:24) notes, the requirement that the artisan's creative intentions need only be largely successfully realised introduces some vagueness into the account. Whether the defining intentions are successfully realised enough for a given artefact to count as an artefact of the intended kind is unlikely to be a binary matter. There may well be borderline cases of knives,

cups, etc. Thomasson points out, however, that this is a virtue of the account, since it should come as no surprise that ordinary kinds have vague application conditions and are subject to sorites style paradoxes.

Our artefact-regarding intentions are not self-fulfilling. This is quite clear in the case of the knife maker – success requires the object produced to have certain physical features – but there are cases where we have to be a little more careful. For example, if I take a pebble from the beach it looks like I just have to form the intention to use it in a certain way for it to become a paperweight. However, note that not just anything could be a paperweight: a distant galaxy or a puff of smoke is simply not up to the task. The pebble has to have certain properties in order for my paperweight intentions to be fulfilled. Later in this chapter (§4) I'll be discussing artefacts like chess pawns and traffic lights where the successful realisation of the artisan's intentions consists in something more delicate: it consists in *successfully communicating those intentions* to other users of the artefact.

1.4 Metaphysical assumptions

Baker (2004:4) argues that "artifacts are constituted by aggregates of things," – ultimately atoms – and that constitution is not identity. This is supposed to make sense of the intuition that the career of an artefact does not coincide with that of the aggregate which constitutes it at a given moment in time. For example, an ancient aggregate of grains of sand can come together under the supervision of an artist to constitute a sandpainting of a volcano. One way of looking at this is to regard it as the creation of a brand new entity, a new sandpainting. The very same sandpainting can then survive the destruction of some of the grains of sand, but the aggregate cannot. Adopting such a view would also enable one to say that the pebble constitutes a paperweight, the latter being a brand new object which only comes into existence when the pebble is found and an intended function is imposed upon it.

I find this a natural and perspicuous way to think about artefacts, and this thesis is largely written in that vein. It goes without saying that these are controversial

assumptions. If there are two objects – both a pebble and a paperweight – then how can they occupy the same space? If each weighs one kilo, why don't the scales show a combined weight of two kilos? Such problems are well known and the subject of much debate. Employing this framework here does not seem to introduce any *new* problems that weren't known about before.

These assumptions track Epstein's (2009) "liberal model of objects." As he points out, if someone objects to these liberal metaphysical assumptions about constitution, it should be possible to translate the account into something more neutral. Instead of regarding the paperweight as a brand new object over and above the pebble, we can just have the pebble (an aggregate) and the various relations it stands in to other entities. Thus, when the pebble is taken from the beach it acquires the extrinsic property of being a paperweight. As far as I can see, most of what I say in this thesis could be translated into the more conservative ontological framework without serious loss.

2. MOTIVATING INTENTIONALISM

The intentionalist approach introduced in the previous section is intended to capture a view which represents something of a consensus in philosophy at the present time.[6] It also, I hope, distinguishes knives, Frisbees, cars, etc. in a very intuitive way from natural objects, natural instruments, social objects like cowrie shells, by-products like iron slag, and the like. The purpose of this section is to provide additional motivation for the intentionalist conception of artefacts. The first two sub-sections argue that intentionalism captures the intuitive way of thinking about artefacts. The third tries to explain why this kind of conceptual analysis is a reasonable method to employ.

2.1 Swamp artefacts

According to Thomasson, it is a conceptual truth that artefacts are products of intentional human activity. One way of supporting this claim is by pointing out that it offers a good explanation of a fairly widespread reaction to certain kinds

---

[6] This is not to say that it is undisputed nor that it has no rival theories.

of thought experiments. For example, suppose that there are naturally occurring objects currently on Mars which are intrinsic duplicates of the chair I'm sitting on. Since those objects are not the products of intentional human activity but have come about purely by chance we can call them *swamp chairs*. The vocabulary is intended to evoke Davidson's swampman thought experiment, in the sense that swamp words are freak products of nature, not arising as a result of any human intention. Are swamp chairs chairs? Intuitively not, one would have thought.

In fact, Bloom (1996) offers experimental evidence which suggests intuitions provoked by such thought experiments are split. Should we then abandon the promising idea that it is a conceptual truth that artefacts are the products of human intentions? I would suggest that an explanation of the intuition that swamp chairs are chairs is that people are subject to a kind of design bias: things which have straight edges, right angles, complexity, orderliness etc., as well as things which look like familiar objects which we know to be usually intentionally produced, are more likely to be categorised as artefacts. If this is correct, we should see an increased tendency to attribute artefacthood as the complexity and neatness of the objects increases. I think this is likely to be a good prediction. People are more likely to say that a swamp Chippendale chair is a chair than that some roughly chair shaped piece of rock is a chair. But the roughly chair shaped piece of rock would be a clear case of a chair if it had been intentionally produced.[7]

2.2 Minimal creation and exaptation

I have been suggesting that artefacts are objects which have some features by intention. This should not be taken to mean that artefacts are necessarily the result of some physical process in which a natural object is sculpted or modified.

---

[7] Dipert probes similar intuitions by nothing that an eroded brick might be an intrinsic duplicate of a natural sedimentary rock, while a natural piece of metal might be an intrinsic duplicate of a gear. He draws the following conclusion: "[A]rtifactuality does not consist in any present physical qualities of a thing… Such physical qualities at most give us symptomatic evidence that an object is, or is not, artefactual; the real basis of its artifactuality must lie elsewhere…An object is, or is not, an artefact based on what its history was" (Dipert, 1993:15).

A naturally frost-split rock could be made into a knife as soon as some wandering human finds it and forms an intention to use it in certain ways; a pebble can become a paperweight; a piece of driftwood can be a work of found art. Such cases involve what Thomasson calls *minimal creation*. There is a choice about whether to conceive of such cases as involving the creation of a new object in addition to the naturally existing one, or to say that there is just the original object which acquires a new extrinsic property. As noted in §1.4, I'm thinking about these things in the former way, though I'm hoping that most of what I say could be translated into the vocabulary of the conservative ontology. In any case, one's ontological preferences in this regard are orthogonal to one's stance on minimal creation. The idea that some aggregate of things constitutes but is not identical to an artefact is a feature of the view independently of the phenomenon of minimal creation. After all, the classic puzzles about the Statue and the Lump are concerned with ordinary, non-minimal creation.

The idea that a brand new artefact can be minimally created should be no more problematic than the idea that a brand new artefact can be brought into existence by carving a piece of stone. Minimal creation is just a kind of artefact creation which does not require any physical modification of a naturally occurring object. It can occur when an object is found whose natural properties make it well suited to the role intended by its discoverer. In cases of minimal creation, a natural object becomes an artefact, perhaps in virtue of an intended function which the artisan intends it to fulfil. Even here, the artisan's intentions are not self-fulfilling. Intending to use a pebble as a paperweight results in a new object, but intending to use an existing object as a paperweight is not sufficient for that object's being a paperweight: the rock has to fall within certain physical parameters in order for the creator's intentions to be successfully realised (e.g. weighing less than 500 kilos, not being made of candyfloss).

Related to the phenomenon of minimal creation is that of exaptation, in which an existing artefact becomes an artefact of a different kind. For example, as Thomasson (2014) notes, a shipment of chopsticks could find their way to an

isolated community who know nothing about chopsticks and end up being used as hair sticks. It is not unreasonable to describe this as a case of minimal creation of new artefacts. Just as you can find a frost-split rock and turn it into a knife in virtue of some intended role which the rock happens to be naturally fitted to fulfil, it appears that you can find take some human artefacts and turn them into other kinds of artefacts. A single object may be continuously converted from one artefactual kind to another during its lifetime, even without changing its intrinsic form.[8] Again, there's a choice about whether to think about this as one object being destroyed and a new one created, or in terms of an aggregate of things which persists through changes in its extrinsic properties.

What's interesting about such cases is that they illustrate how creator intention can be partially constitutive of the nature of an artefact. Take the pebble-turned-paperweight. Suppose the pebble has existed more or less unchanged for a million years. When a human with certain intentions comes along, something happens. Suddenly, the pebble constitutes an instance of a particular kind of artefact, a paperweight. What has changed? Nothing but the beliefs and intentions of the human. In this case, a certain kind of intention is the final ingredient needed for the pebble to be a paperweight. Bloom makes a similar point:

> Given that the creation of members of these kinds need not involve any physical changes, our judgments are driven here solely by our intuitions about intentionality. We construe a penny as a pawn only if we intuit that a person sincerely intends for the penny to fall into the class of pawns… Note that one does not have to do anything to the penny for it to become a pawn… [W]hat makes this penny a pawn (as opposed to a queen, say) is the mental state of the person who is considering the chess problem. (Bloom, 1996:18)

A reasonable explanation of our attitudes towards cases of minimal creation is that artefacts are taken to depend constitutively on their creators' intentions.

---

[8] An extreme version of this view could hold that each user of the artefact creates it anew. This recalls van Fraassen's (1993) suggestion that an artwork is not identical to a painting or sculpture, but consists in an individual's imaginative interpretation of it: "the material work is the temporally persisting focus of a sequence of imaginative creations, which can alternately be viewed as a single imaginative work evolving through centuries, liberated from the confines of any one individual mind, including the artist's."

Moreover, the phenomenon illustrates how the specific content of a creator's intention contributes to the character of the artefact created. In relation to the penny which becomes a chess pawn, Bloom writes:

> The right sort of intention is essential; someone who does not know the rules of chess could not do this, as she lacks the right understanding of what it is for something to be a pawn. (Bloom, 1996:18)

A few puzzles arise. For example, why is it that can we make a hair stick by using a chopstick to hold one's hair up and yet if we place a teapot on a stack of papers we still just have a teapot? This problem about exaptation can be resolved (see §4.4), but that resolution makes use of some conceptual machinery introduced later in this chapter. For now, let's focus on a puzzle about minimal creation: why consider the taking of the pebble to result in the creation of a paperweight while saying that a stick seized upon as a weapon in the heat of the moment is merely a case of using a natural object as an instrument?

The distinction between artefactual instruments and non-artefactual ones is no doubt gradual and blurry, but the contrast has to do with the ways human agents conceive of the objects they use, and the extent to which an object satisfies that conception in a non-accidental way. Let me explain the point by example. A particular stick may be used as a weapon because it was the only object within reach, or it may be selected amongst many available sticks for its natural usefulness as a weapon. Suppose it has been carefully chosen – after a thorough examination of promising looking sticks – for its strength, shape, hardness, weight, distribution of weight, smoothness and graspability. In this case the human has a rich conception of the kind of object they need and a particular natural object has been selected as largely satisfying that conception. We can give a detailed explanation of the features of the weapon which refers to the human's intentions. To my mind, it is natural to describe this as creating a new artefact even while insisting that the same stick seized upon blindly would have remained a mere instrument. Similarly, the pebble-turned-paperweight can be assumed to have been selected for its size, weight, aesthetic properties, etc. Of course, such

distinctions are blurry: the stick seized upon blindly was presumably selected in preference to any grains of sand or clumps of mud nearby because the hardness, size, shape, etc. of the piece of wood were conceptualised as making it a more effective weapon than those other items.

2.3 Reference to artefacts

The previous two sub-sections were intended to support the proposed definition of artefacts by showing how its adequacy would explain fairly widespread reactions to thought experiments. This argumentative strategy is common in both the psychological and philosophical literature on artefacts. But one might object that this is no way to begin an inquiry into the metaphysics are artefacts. We could just be wrong about what artefacts are.

Thomasson (2007) provides the following response. First, she argues that facts about how artefactual kind terms are grounded support the intentionalist account of artefacts. Suppose we ostend some region of space-time and coin a new word to refer to the kind of thing which is present there. Thomasson (2007:55) notes that "a speaker who seeks to ground that term's reference must have at least a very high-level background conception of what sorts of features are relevant to being a member of this sort of kind," such as "chemical, biological, artefactual, cultural, legal, etc." In other words, the speaker has to have some conception of the kind of thing they intend to name, otherwise it is indeterminate what kind of thing the new term refers to. Thomasson writes:

> Those who ground the reference of any artefactual kind term must themselves have some conception of what general sorts of features are relevant to determining the kind's nature and the term's extension, for it is this that establishes the term as a would-be artefactual kind term. About this, they cannot be proven to be in error through later empirical investigations, since this establishes the sort of nature that is relevant to the reference of the term, if it refers at all." (Thomasson, 2007:55)

Thomasson thinks that the speaker who grounds the reference of an artefactual kind term is intending to refer to the kind of entity of which creator intention is

37

partially constitutive. They can be wrong about whether or not the term thus grounded refers, but if it refers at all, it refers to an intentional artefactual kind.

In turn, the *maker* of a new kind of artefact grounds the reference of a new artefactual kind term in specific features which the prototype artefact is intended to have. These intentions delineate an artefactual kind. The speaker cannot be wrong about the kind of artefact they intended to make, although they can be wrong about whether or not they have succeeded in creating an artefact of that kind, (and they can be wrong about whether artefacts of that kind are correctly referred to using some public language word).

The idea is that we can begin our inquiry into artefacts by focussing on artefact concepts. If the proposed definition does a good job of capturing the concept of an artefact, then we can justifiably ask whether any such entities exist, but we cannot ask whether that is what artefacts really are. We already know what artefacts are, *if* they exist. Similarly, the inventor of a new kind of artefact can't be wrong about the kind of artefact they are trying to make because the content of their intentions amounts to a stipulation of what kind of artefact they are trying to make.

This response is not entirely satisfactory. Suppose it turns out that there are no entities which depend constitutively on human intentions in the way suggested above. Something could qualify as an artefact by satisfying our artefact conception imperfectly. Moreover, as noted in the introduction, Elder (2007) defends a non-intentionalist account of (a class of) artefacts based on an interpretation of Millikan (1984). He calls such entities *copied kinds*:

> [T]he members of any copied kind are characterized by what Ruth Millikan calls a 'proper function'… That is, the members are produced by a process or mechanism which copies them from previous members similarly shaped, and does so as a causal consequence of performances, by those previous members, of certain functions – productions by them of certain effects. (Elder, 2007:38)

On this view, it is denied that intentions have any role to play in a theory of artefacts:

> [T]he creation does not begin with the artisan's intending what he does. Rather the essential properties which his product will inherit stem from a history of function and of copying that began well before the artisan undertakes his work. This history reaches forward through the artisan's motion – it shapes his shaping. Its existence and its efficacy are independent, largely or even entirely of the artisan's will. (Elder, 2007:39-40)

If this is right, the study of artefactual kinds is more akin to the study of natural kinds. They constitute joints in nature which have to be uncovered using the usual empirical methods, while the metaphysics of artefacts would be a chapter in the philosophy of science.

This approach to artefacts is a serious competitor to the standard view I've been canvassing here, but it is largely set aside in this thesis. By way of excuse, note, first, that Elder (2007) does not intend his account of copied kinds to be applicable to all artefacts. It might be that some artefacts are copied kinds while others fall within the remit of the intentionalist account. Later in this chapter I will be zeroing in on artefacts which have an essentially communicative function. Perhaps these would always remain beyond the purview of the copied kinds account. Second, Thomasson (2007:57) tries to argue (though I am unsure how successfully) that the notion that artefacts have derived or proper functions still requires that these functions be intended by their creators. Third, since this thesis is primarily about linguistic artefacts, and not concerned primarily with general problems in social ontology, it seems appropriate to continue to use the standard view. Whether or not this is a fruitful way of thinking about words will become clear in later chapters. In any case, applying the intentionalist account to the case of words is a potentially fruitful thing to do, both for the light results in social ontology may throw on language and for the extra scrutiny the standard account receives in virtue of being applied to the case of language. Fourth, Bloom (2007:4) makes the psychological generalisation that while "[n]atural kinds are

understood in terms of internal essences" artefacts are conceived in terms of "creator intent, characteristic function, and the social and cultural context of the artifact's creation and use"; if this is correct – if it is part of the very concept of an artefact that it be partially constituted by intentions – then we can construe the present adoption of intentionalism about artefacts as an attempt to sustain a metaphysics which reflects intuitive conceptions, with the copied kinds approach as a fall back option.

3. ARTEFACT KINDS

So far I've tried to speak mainly about individual artefacts, keeping discussion of artefact kinds and artefact categorisation to a minimum. In this section I address these topics head on.

Thomasson argues that the creation of an artefact depends on its creator's having a fairly rich conception of the artefactual kind of which the created artefact is to be an instance. Hilpinen has much the same view:

> When a person intends to make an object of a certain kind, his productive intention has as its content some description of the intended object, and the author's intention "ties" to an artefact a number of predicates which determine the *intended character* of the object. (Hilpinen, 2011)

For example, the person who stipulates that a penny is to be a pawn in a game of chess has to have a fairly rich conception of what pawns are, including what they can and cannot do on a chess board. Thomasson proceeds to explicate sameness of artefactual kind in terms of (largely successfully realised) matching intentions on behalf of their creators. Thus, the reason why two chess pawns are pawns is that they were produced with the same chess-regarding intentions in mind. This is the view I'll be exploring in this section.

It's worth recalling that this tour through the philosophy of artefacts is not undertaken purely for its own sake, but to develop the tools we need to offer a theory of words in subsequent chapters. Questions about word individuation

have been central to philosophical discussion of words. The following discussion of artefact categorisation yields insights which we can bring to bear on the case of words.

3.1 Matching conceptions

Thomasson proposes that, in the case of prototypes, the artisan starts with certain features in mind, functions that the artefact is to perform, structural features it is to have, etc. If the would-be inventor does not have a substantive idea about the central features of the thing they are making, they are just messing around, hoping for a felicitous discovery. If they do, then this conception provides criteria for success:

> [The inventor] must have some goals to direct her activity. Including some features she intends to impose on the object created. Thus she must have a substantive idea of what sort of a thing it is she intends to create (say, a K), where that idea incorporates certain features relevant to being a K, so that she can judge her activity's success in terms of the degree to which the product instantiates the relevant features. In this case, clearly, there is no question of the artisan getting it right or wrong about what it would take to be a K, what features are K-relevant. At this stage, what is relevant to being a K is purely a matter for invention or stipulation by the artisan based on her goals or intentions; she is not trying to discover what makes something a K (so that she could be said to get it right or wrong), instead, she is delineating a new kind by establishing success criteria for her activity. Thus, she creates not only an artifact, but delineates a new artefactual kind, complete with normative success conditions for creating something of that kind. (Thomasson, 2007:60)

With respect to later artisans, either they know about Ks (i.e. they have a conception of Ks which matches that of earlier creators) or they do not. If they do not then they are in the same position as the inventor. If their conception of the object is the same (they intend to impose the same K-relevant features) then they are intending to make a K:

> So a later artisan succeeds at making a K only if he has a substantive, and substantively correct, concept of what a K is and succeeds at imposing on the object all or most of the features relevant to executing that concept. Having a substantively correct

> concept, in turn, must be a matter of substantially matching the prior concept of Ks, since inventors' concepts were originally definitive of what counts as relevant to kind membership. (Thomasson, 2007:62)

To take an example, suppose I try to make a knife. Either I have seen knives before, or I haven't. If I haven't then I'm in the same situation as the maker of the prototype. If I strike upon the same essential features (sharp blade, handle, used for cutting food, etc.) then I may succeed in making a knife. In either case, it just depends on whether my conception of the thing I want to make is a close enough match for the conceptions that other people have of knives (and whether or not my intentions are largely successfully realised.) Artefacts can thus be multiply invented. (It is of course, perfectly plausible that isolated individuals or groups could develop knives from scratch. This is a positive feature of the present approach. Some theories of artefacts have trouble securing this intuitive result. See chapter four (§1.6) for discussion.)

3.2 Ignorant artisans

Against Thomasson one might object that it makes perfect sense to say that I could take a horseshoe to a blacksmith in a country without horses and get them to make me another one. The blacksmith may have no conception of horses and understand nothing about horseshoes, but is nonetheless able to produce an intrinsic duplicate of the one I brought. The objection is thus that an artisan could be entirely ignorant of the essential features of an artefact they have produced.

A good option here is to say that *my* intentions are the ones which make the blacksmith's product a horseshoe. From the blacksmith's perspective, they are creating a piece of metal with certain intended proportions, one which has the intended function of satisfying the client. The blacksmith is like a machine which I have intentionally set in motion. They can replicate all of the intrinsic, physical properties of the horseshoe but cannot confer upon it the socially/mentally constructed properties which make it a horseshoe (such as being intended to

shoe horses). That can only be done by someone with a rich conception of what it takes for an object to qualify as a horseshoe.

The rejoinder to this might be that we don't need to appeal to any third-party intentions. The blacksmith has been presented with an unknown artefact and is able to reproduce it faithfully by replicating its physical properties. Why can we not maintain that this is at least sometimes sufficient for creating a horseshoe? The reason we can't do this can be illustrated by example: suppose that some artisan has no conception of Frisbees or of plates. On Monday, Anne brings them a Frisbee and asks them to make another one the same. They do so. On Tuesday, Brian brings them a plate – which happens to be intrinsically indistinguishable from Monday's Frisbee – and asks them to make another one the same. They do so. In virtue of what is Monday's creation a Frisbee and Tuesday's a plate? My answer is that it depends on Anne's intentions and on Brian's intentions respectively. Anne intended for the artisan to create a Frisbee, and Brian intended for them to make a plate. This is what settles that Monday's creation is a Frisbee and that Tuesday's is a plate. We cannot appeal to the artisan's own intentions. From their perspective they have made identical client-satisfying artefacts on successive days. Nothing the artisan thinks or does could decide whether the objects are Frisbees or plates.

It won't help to argue that on Monday the artisan makes a Frisbee because they take a Frisbee as their model, and that on Tuesday they make a plate because they take a plate as their model. Suppose that Anne's and Brian's belongings have become mixed up so that Anne has mistakenly taken Brian's plate to the artisan as the exemplar for her desired Frisbee, and that Brian has mistakenly taken Anne's Frisbee as an exemplar for the plate he wants. The artisan still makes a Frisbee on Monday and a plate on Tuesday. This suggests that what matters is not whether the artisan's exemplar is a Frisbee or a plate but what the client intends to be caused to be made by the artisan.

43

For a slightly different case, suppose someone notices that finger spinners are selling like hot cakes but has no idea what they are. They could still manufacture and sell a batch of finger spinners merely by carefully producing intrinsic duplicates of existing finger spinners. In this case, there are no finger spinner savvy clients directing production, but we can perhaps appeal to the notion that the artisan is acting in a way which is deferential to a certain practice of making finger spinners. There are, in the vicinity, people who have the right kind of rich conception of what finger spinners are, and the artisan is deferential towards those people.

3.3 Self-referentiality

What makes a plate a plate and a Frisbee a Frisbee? A tempting thought is that part of the story is that Frisbees are intended to be Frisbees and plates are intended to be plates. This is a natural way of expressing the idea that the content of a creator's intentions make a contribution to the artefactual kind of the objects they produce, but it is not adequate as an account of that contribution. Searle (1995, 2007) confronts a similar problem when he notes that social concepts often exhibit a troubling kind of "self referentiality":

> If part of the content of the claim that something is money is that it is believed to be money, then what is the content of that belief? If the content of the belief that something is money contains in part the belief that it is money, then the belief that something is money is in part the belief that it is believed to be money; and there is, in turn, no way to explain the content of that belief without repeating the same feature over and over again. (Searle 1995:33)

The problem can be dissolved by clarifying that talk of something being money because people believe it is money should be construed as mere informal presentation of the view. This way of speaking creates the appearance of circularity where there is none.

Searle's approach in social ontology is discussed in chapter three, so detailed presentation of his view can wait until then, but a quick glimpse is useful here in

order to see our way out of the present problem. On Searle's view, social objects like dollar bills have the social properties they do in virtue of the fact that some community collectively accepts constitutive rules whose form is given by the following schema:

X counts as Y (in C)

Suppose we want to give an account of what makes something a one dollar coin. The X term in Searle's schema is supposed to be filled out with terms which refer to a kind of object, presumably one with a specific shape, weight, material constitution and provenance (e.g. being issued by the *Bureau of Exchange*). The Y term is supposed to be filled out with "status functions" such as being a medium of exchange, a store of value, etc. ('C' stands for a context in which Xs are to count as Ys.) On this view, what it is to be a one dollar coin is explicated in terms of the cluster of properties in Y. Thus, a dollar is not defined as something believed to be a dollar; a dollar is defined as anything which counts as having the properties in Y. As Searle puts it:

> [Y]ou don't have to use the word 'money' in order to define money. The word 'money' functions as a summary term or as a placeholder for being a medium of exchange, a store of value, a payment for services rendered, a measure of value of other currencies, and so on. And if something performs all of those functions, then it's money. So we do not have a vicious circularity or infinite regress. If I say that in order for something to be money, people have to believe that it's money, there is no circularity, because they can have that belief without having the word 'money'. The word 'money' here just is a place-holder for a large number of other functional expressions.

To be sure, this is not a reduction of social facts to non-social ones: such an account does not explain what it is to be, for example, a medium of exchange at a more basic level. What's at issue is how socially constructed properties get attached to an independently specifiable substrate. For Searle it's through collective acceptance of constitutive rules of the form illustrated by the XYC schema. The passage from X to Y is ampliative. In virtue of collective acceptance

of an XYC rule, the Xs end up having new properties they didn't have before. We can understand how these new properties attach to Xs, even if we can't see how to reduce the Y properties (or even the X properties, for that matter) to non-social, non-intentional states of affairs. New socially constructed properties get attached to entities of type X, and the concept of money is analysed in terms of some cluster of Y properties. There is no circularity here.

There are various problems with Searle's XYC schema, but further discussion of the account can be postponed until chapter three, (where I explain why the XYC schema is ill-fitted to explaining the mental/social construction of many artefacts). Here I want to translate Searle's response to the circularity objection so that we can pursue the Thomassonian conception of artefacts and begin to make some headway with the question with which we started: what makes a plate a plate and a Frisbee a Frisbee?

If a Frisbee is defined as something which is intended to be a Frisbee, then we haven't defined Frisbees any more than we define blargs by claiming that blargs are things which are intended to be blargs. But although we might informally present the Thomassonian view that way, that is not really what's going on. Instead, what it is to be a Frisbee is explained in terms of intended features such as being intended for use in a certain kind of recreational activity, intended aerodynamic properties, etc. So, artefact kinds are understood in terms of clusters of intended features. They are not defined self-referentially. (The objection also goes awry in assuming that this kind of social ontological theorising is expected to yield *definitions* of artefactual kinds. For many kinds of artefacts it may be impossible to provide a definition in terms of necessary and sufficient conditions.)

3.4 The structure of artefact concepts

What kinds of features are central to the natures of artefactual kinds? Intrinsic physical features and appearances play some role in artefactual categorisation, but artefactual kinds cannot generally be defined in terms of such features. For

example, not all chairs look alike (e.g. folding camping chairs, chairs carved out of logs, Chippendale chairs, dentist's chairs, orthopaedic chairs, etc.), though none of these are borderline cases of chairs. At the same time, some things that duplicate the physical properties of paradigmatic chairs are not chairs (e.g. swamp chairs).

An alternative would be to define artefacts in terms of their actual causal powers, but this won't work either. One might think that clocks are things which can be used to tell the time, but a broken clock is still a clock, while watches and stars are not clocks. Similarly, a Frisbee is not a plate, even though it could be used as one. A general problem is that artefacts have far too many causal powers. A pebble turned paperweight could be used to hammer in a nail, hold a door open or grind pepper in a mortar, but it is not a hammer, doorstop or pestle. It is a paperweight. We can't make sense of that just by talking about the pebble's brute causal powers. Nor is actual use generally either necessary or sufficient for membership of an artefactual kind: boats which never make it to the water are still boats, and desks which people sit on are not chairs.

In the previous section I canvassed the idea that artefacts are objects produced to serve some purpose or intended function. Against this I noted that some kinds of artefacts may have no function or have a function which is neither necessary nor sufficient for being a member of the kind (e.g. a statue). Even for those artefacts where intended function does seem relevant, e.g. boats, intended function may be neither necessary nor sufficient: a boat might not be intended ever to see water while two groups of adventurers might end up making a boat and a raft respectively, though each artefact was intended to have the function of transporting them downriver. Bloom (1996) notes the results of experiments in psychology which test the relative weight which people accord to physical form and intended function in artefact categorisation. These experiments suggest that intended function is not generally either necessary or sufficient for being a member of a given artefact kind since there is a tendency to judge (i) that a rubber sphere attached to dolphins which is intended to carry people over bodies

of water (i.e. something which has the intended function associated with boats but atypical physical properties) is not a boat, and (ii) that a boat shaped object intended exclusively for unmanned retrieval of marine biological data is a boat.

It turns out to be exceedingly difficult to give a general account of the kinds of features which are taken to be relevant to membership of an artefactual kind. It is no easier to give a specific account of the features which are taken to be relevant to specific artefact kinds. This is partly due to the fact that there is diachronic variation within artefactual kinds: the suits of the 1930s, 1970s and present day are not quite the same: they don't look the same, they're not made the same way or of the same materials, and they have different social roles. But there is also synchronic variation. Within a given artefactual kind there tends to be enormous variety, making attempts to define an artefactual kind subject to immediate counterexamples.

What is more, some members of an artefact kind tend to be judged as more or less typical exemplars of the kind. For example, chairs shaped like a hand or suspended from the ceiling on chains are thought to be rather atypical chairs. These kinds of judgements are best explained by taking artefact concepts to have a family resemblance structure instead of a classical structure defined in terms of necessary and sufficient conditions.

Thomasson states that we should not expect artefact kinds to be definable in terms of necessary and sufficient conditions. Artefact concepts, she suggests could be 'cluster concepts'. Presumably, the thought is that an artefact kind is associated with a range of properties, but something can be an artefact of that kind without having all of those properties; at the same time, there may be no single property which is necessary for being a member of the kind.

Thomasson proposes a distinction between *strict* and *loose* artefactual kinds. At least some artefact kinds are strict in the sense that they have strict application conditions, perhaps because they are associated with technical expertise. She

gives the examples of double breasted waistcoats and Peking duck. The thought here may be that since these kinds are technical artefacts associated with professional standards, very precise criteria for what counts as instantiating the kinds should be forthcoming. To cover these kinds of artefacts, Thomasson proposes the following principle:

> Necessarily, for all x and all strict artifactual kinds K, x is a K if and only if x is the product of a largely successful intention that (Kx), where one intends (Kx) if and only if one has a substantive concept of the nature of Ks that matches that of prior makers of Ks (if any) and intends to realize that concept by imposing K-relevant features on the object. (Thomasson, 2003b:24)

Whether or not double breasted waistcoats or Peking Duck can really be provided with strict application conditions in terms of necessary and sufficient conditions is debateable. What is certainly the case is that many, many artefactual kind terms have much looser application conditions. For one thing, there are many different ways of being a key or a bottle. Moreover, as Thomasson notes, the properties associated with many such artefactual kinds are subject to change over time. She gives the example of modern dresses which may differ markedly from what a dress was understood to be by Victorian dressmakers. The problem is that the dependence principle for strict artefacts does not allow for drift in the nature of an artefactual kind over time: it specifies that later artisans must have a conception which (exactly) matches that of earlier artisans. To remedy this, Thomasson proposes this more inclusive version of the above dependence principle, which requires only that later makers have a conception which largely matches that of earlier makers:

> Necessarily, for all x and all artifactual kinds K, x is a K *only if* x is the product of a largely successful intention that (Kx), where one intends (Kx) *only if* one has a substantive concept of the nature of Ks that *largely* matches that *of some group* of prior makers of Ks (if there are any) and intends to realize that concept by imposing K-relevant features on the object. (Thomasson, 2003b:26)

This weaker condition requires only that later makers' conceptions largely match those of some other group of makers. This allows for gradual drift over time as,

for example, adding machines evolve into computers (or the Old English word, 'docga', evolves into 'dog'). But it also preserves continuity since you can't make an artefact of a certain kind if your conception of the kind does not even remotely match that of at least some prior maker of that kind of artefact. No one can produce a top hat and claim it's a barrel of beer (just as no one can produce a monosyllabic utterance with no consonants and claim it's an instance of 'otorhinolaryngologist'.)

The revised dependence principle differs from that formulated for so-called 'strict' artefactual kinds in specifying only a necessary condition on being an instance of a given artefactual kind. This is because we have to make space for some intricate phenomena. For one thing, someone could produce an artefact according to an artefact kind conception which differs a little bit, but not too much, from some prior group of makers of artefacts of that kind, and be only largely successful in realising their intentions. The result of this could be an artefact which shares relatively few significant properties with prior artefacts of the kind.

Second, artefacts can evolve radically over long periods of time to such an extent that we may be unwilling to admit that artefacts on opposite ends of the spectrum are artefacts of the same kind. Thomasson gives the example of early adding machines and computers. As Thomasson notes:

> Artifactual kinds are notoriously malleable and historical in nature – indeed the possibilities for this are built into our description above… [O]ver time, the concept of K's, spelling out which features are K-relevant, may gradually change… [T]he process of stipulation has become much more gradual and diffuse, as it is responsive to the intentions of a great number of makers over an extended period of time. (Thomasson, 2007:62-63)

Third, artefactual kinds are subject to fission and fusion (though Thomasson mentions only fission.) Thomasson mentions the case of 18th century knickerbockers which apparently gave rise to two different kinds of garments worn today, one being a kind of underwear and the other a kind of outerwear

used for sports. Artefactual kinds are also plausibly susceptible to fusion, such as, for example, when an existing piece of technology takes on an extra role (e.g. smartphones could take on the role of keys, credit cards, etc.). Thomasson's weaker dependence principle is silent on questions like these, though she could probably again insist that that's just how things ought to be.

Recall that the shift from strict to loose artefactual kinds involves abandoning the attempt to define artefactual kinds. Still, there are two things we can say. First, it is often relatively easy to specify sufficient conditions for being a member of a loose artefactual kind. For example, anything which is the product of a largely successfully realised intention to make a small cart with a single wheel at the front, supported by two fixed legs at the rear, which can be lifted or pushed by two horizontal handles, and is to be used for transporting awkward loads in gardening and building…is a wheelbarrow. Having these intended properties seems to be sufficient for being a wheelbarrow, even though none of them is necessary. Second, if anything is a wheelbarrow, then it must be the product of a largely successfully realised intention which largely matches this conception.

4. ARTEFACTS AND RECOGNISABILITY

Essentially intentional artefacts depend constitutively on human mental states. In that sense they are like the social entities such as laws, schools, cocktail parties, etc. But one might wonder if artefacts should necessarily be considered to be *social* entities. What seems most important in the case of an artefact is that its creator – who is often a single individual – have certain intentions. For example, it might seem plausible that when someone makes a knife their very own intentions are what do the work: if someone is intentionally (and largely successfully) honing a sharp edge to produce a cutting tool then that is plausibly sufficient for creating a knife. However, Thomasson (2014) argues that many artefacts are distinctively social in character, depending not just on a single individual's intentions but on those of a multiplicity of individuals.[9] Coins are an

---

[9] In her (2003b, 2007) Thomasson assumes that artefacts do not in general depend on anything but the intentions of individual makers. By her (2014), this has changed.

obvious example. No piece of metal could have a monetary value of one dollar, or have the function of being a medium of exchange, just in virtue of a single individual's intention. For this to work, there has at the very least to be a multiplicity of individuals with beliefs and intentions regarding coins.

Thomasson proposes to make sense of the publicness of some artefacts in terms of a technical notion of *recognisability*. A similar notion can also be found in Dipert (1993). The matter is also related to Searle's distinction between social and institutional facts. On views such as these, there is a certain class of artefacts which have symbolic or communicative features. These can include physical features which serve no practical function directly, but which enable the artefact to be recognisable.

This aspect of artefact theory is especially important for present purposes. This is because it calls attention to a distinctive mechanism of social construction which helps to understand a wide variety of artefacts. In the next chapter I'll show how this mechanism can be used to explain features of words. Here, I'll provide a more detailed account of these recognisability features than has been provided before.

4.1 Thomasson on recognisability

Thomasson argues that among the properties of artefacts which may be criterial for their being members of a given kind are "what might be broadly construed as receptive and normative features, involving how the object created is to be regarded, used, treated or behaved in regard to" (Thomasson, 2014:47). These include being recognisable as having various features. For example, a shop-front will be designed with the intention that it be recognisable (by a certain intended audience) as indicating a place where such and such products are being sold. Without other people with the right beliefs and intentions there can't *be* a shop front, since anything which wasn't widely recognisable as indicating that products are to be procured within couldn't really be counted as a shopfront.

The basic idea is that many artefacts have structural, detectable features which do not serve any purpose directly, but which serve to make the object recognisable as having various socially constructed properties. For example, the red and white pole traditionally positioned outside a barber's shop is intended to make the place recognisable as somewhere you can ask to get a haircut in exchange for money. Thomasson's idea is that a crucial property of many artefacts is that they have certain detectable features which enable the object to be recognisable as a kind of thing which is intended to be used or treated in certain ways. This illustrates one sense in which an artefact could be described as a *social* or *public* artefact, as opposed to a purely individualistic artefact. The features which contribute essentially to the artefact's makeup may depend upon the beliefs and intentions of a wider community, not only on the intentions of the individual creator.

Thomasson qualifies this account in three ways. First, since a hermit might make themselves a teapot we should characterise receptive features not as involving an intention that the object be *recognised* as having such and such a feature, but only that it be *recognisable*. Second, since a rubber shark may be intended to be recognisable as a rubber shark only by the film crew, and not by the audience in a cinema, the receptive criteria should involve recognisability relative to an intended audience. Third, since some artefacts may be entirely personal creations not intended to be recognisable by anyone other than the maker (or not even by the maker), Thomasson states that the essential recognisability of artefacts is assumed to apply to just a subclass of the artefacts, those which are *public artefacts*.

Thomasson (2014:56) considers two objections to the notion that receptive features are key to the identity conditions of many artefactual kinds. The first is the worry that someone alone on a desert island could create an artefact such as a knife. To this Thomasson replies that if the person has previously lived in a human society then their activities are still infused with the conventions and norms of home, so that their creating a knife can be described as intending to create something which adheres to those public criteria. If, on the other hand, they have not, and have miraculously survived infancy and managed to fashion

various useful objects, Thomasson thinks we might admit that there can be *private tools*, artefacts in a broader sense of the term, but not *public artefacts* like knives and chairs. (So, even if the castaway creates something that looks like a knife and is used for cutting it wouldn't be an instance of the public artefactual kind, *knife*. This does seem a rather infelicitous result. Happily, my own take on these matters (see §4.3) does not have this consequence).

The second objection concerns prototypical artefacts: when a kind of artefact is first invented, it looks like only the intentions of the creator can possibly count. Thomasson (2014:56) replies by noting that many new artefacts are recognisable as being subject to existing norms: "the Wright Brothers may well have intended their creation to be subject to at least some of the norms for treatment of transportation devices." Additionally, if someone were to insist that the first knife or the first hammer was really not intended to evoke any social conventions or norms then we could reply that that first knife was a private tool, and that the public kind, *knife*, emerged from that early state of affairs gradually.

4.2 Clarifications of Thomasson on recognisability

Thomasson tends to present recognitional features as just a further kind of feature which can be relevant to artefactual categorisation: just as an artefact kind could be described in terms of functional features or formal features, or some kind of mixture of these, recognitional features are just treated as another kind of feature to throw into the mix. However, this way of thinking of the view fails to do it justice. Instead, we should understand intended recognisability as a *mechanism* of social construction. In this sub-section I'm going to try to make it clear how this mechanism works. (I'm going to work through this in a cumulative way. This will involve posing some objections to formulations of the view that don't really work. These are not intended as objections to Thomasson, but as stepping stones to understanding her view).

One way to present Thomasson's view is to say that an instance of a public artefactual kind must be the product of a largely successfully realised intention to

make a thing recognisable as an artefact of that kind. This is fine as a casual characterisation of the view, but phrased this way it of course invites the kind of circularity worry I discussed above. Less casually, then, the recognisability criterion could be stated as follows:

RECOGNISABILITY: An instance of a public artefactual kind must be the product of a largely successfully realised intention to make a thing recognisable (by a certain audience) as having $Y_1…Y_n$.

In this schema, $Y_1…Y_n$ is some cluster of features associated with artefacts of that kind. They could be functional features, physical features or – as Thomasson (2014:47) emphasises – normative features including ways in which "the object created is to be regarded, used, treated or behaved in regard to."

Naturally, being the product of a successfully realised intention to make a thing which *just looks like* it has $Y_1…Y_n$ isn't generally sufficient for making a thing which actually has $Y_1…Y_n$. So we still need to say something about how an artefact gets its Y properties. To the extent that the Y properties are physical properties, there's no mystery. What is essential to gold spheres is being spherical and made of gold. One succeeds in one's intention to create a gold sphere by physically shaping a piece of gold. But properties like being subject to certain norms, having a certain function etc. cannot be imposed just by physically shaping the object. As a result, we need something like the following:

RECOGNISABILITY´: An instance of a public artefactual kind must be the product of a largely successfully realised intention to make a thing recognisable (by a certain audience) as being intended to be $Y_1…Y_n$.

What is distinctive to public artefacts, on this view, is that they are products of a certain kind of second-order intention. The artisan intends a piece of wood to be a pawn and intends that *this intention* be recognisable. The success of such an intention generally requires structural features of objects (or nearby objects)

55

which communicate the intended Y features. What is key here is that for at least some Y properties, getting people to realise that something was created with an intention that it be Y is sufficient for that thing being Y. For example, a craftsman may intend to make a thing which is recognisable as being intended to be a pawn (where being a pawn is analysed in terms of conventions in the game of chess). One way of signalling their intention that certain pieces be pawns is to give the pieces of wood that characteristic pawn shape, though not all pawns are like this. In games of chess played with real people, types of chess piece can be signalled by writing letters on people's backs, giving them coloured hats and so on. Similarly, recall Bloom's example in which a penny becomes a pawn. What seems to be crucial to being a pawn is not some particular shape (some particular way of signalling pawn-ness). These structural features of pawns are merely there to indicate their creators' intentions. Getting people to recognise those intentions is then sufficient for the pieces of wood (or pennies) to be pawns.

Actually, that can't be quite right, since you might expect that a creator's intentions have not only to be recognised but also *accepted*. If I stipulate that a particular penny is a pawn, it seems that I only succeed in making a pawn if the other player accepts that it is. But requiring creator intentions to be actually accepted seems too strong. If I buy a chess set and refuse to accept that such and such is a pawn, and if I treat the pawn as firewood, I'm still burning a pawn, even though no one ever accepted the creator's intention that it be a pawn. Similarly, if the creator hides a new set of chess pieces in a vault so that no one ever sees them, they're still pawns, bishops, etc. Perhaps what's required is not that the creator's intention be accepted, but that it be *acceptable* to some significant set of individuals: the creator's intention that the object be a pawn can be successful if the object is such that it communicates the creator's intention that it be a pawn, and the object would be accepted as a pawn by a certain significant set of individuals in virtue of their recognition of the creator's intention. So the recognisability criterion can be stated, with added verbosity, as follows:

RECOGNISABILITY´´: An instance of a public artefactual kind must be the product of a largely successfully realised intention to make a thing recognisable (by a certain audience) as being intended to be $Y_1…Y_n$ and the intention that it be $Y_1…Y_n$ is acceptable (to that audience).

This mechanism of social construction won't work for all Y properties. Successfully intentionally making something that looks like it is intended to be used for cutting isn't sufficient for making a knife. It also has to have certain causal powers which enable it to cut. But for some Y properties – e.g. that a line in the sand is the baseline for a game of volleyball, the success of the intention to make a baseline consists in that intention's being recognised (and being accepted/acceptable) by the intended audience. Clearly, this mechanism of social construction depends on mental states not just of the individual maker, but on wider attitudes in the community. People need to be disposed to recognise and accept the thing as having certain (intended) Y properties.

Dipert (1993) makes a related distinction between *tools* and *artefacts*. Tools are objects intentionally made with some purpose in mind, while artefacts are more like what I've been calling public artefacts. Dipert writes:

> An artefact is an intentionally modified tool whose modified properties were invented by the agent to be recognized by an agent at a later time as having been altered for that, or some other, use. In other language, an agent has intended that an(other) agent comes to believe, on the basis of perceiving the presence of one or more (intentionally modified) properties, that (1) the object is an especially suitable means for achieving an end and that (2) the object's creator intended to cause this belief by these properties…Hence, artifacts are, unlike tools, distinctively "social"… They require us as agents to think of other cognitive and acting agents, their attitudes and thought and emotional mechanisms, and the content of their thought and attitudes. (Dipert, 1993:30-31)

To my mind it is extremely awkward to define 'artefact' in such a way that a hammer is not an artefact. But that is just a terminological complaint. Dipert's

account is clearly closely related to Thomasson's. Dipert highlights what he calls the "communicative purpose" of (what *he* calls) artefacts:

> According to my definition, all artifacts have a communicative purpose. By "communicative purpose" I mean that they were made with an intention to bring about a belief in an(other) agent. (Dipert, 1993:102)

At this point I would like to introduce a piece of terminology which I will use throughout this thesis. What Dipert calls "artifacts," I will call essentially communicative artefacts, or ECAs. What is characteristic of ECAs is that an object only gets to be one of these kinds of artefacts via a process in which the creator's intention regarding that object is recognisable and acceptable to an intended audience. That is not the case for wheelbarrows, knives or hammers. It *is* the case, I would suggest, for artefacts like voting slips, newspapers, computer displays, traffic lights, foot-high fences, price tags, police tape, wedding rings, the conch in Lord of the Flies, trophies and medals, uniforms, crucifixes, churches, monarchs' crowns, money, gang tattoos, judges' wigs, car registration plates, tax discs, war memorials, cattle brands, do not disturb signs, etc.[10] Artefacts such as these have an essential, symbolic or communicative role.

As noted, I think it is best to think of "recognisability" or the "communicative purpose" of an artefact as a mechanism of social construction: that some object makes clear its creator's intentions regarding its function or the norms to which it is subject is what turns a lifeless object into an ECA.

4.3 Recognisability and publicness

Sometimes we want to say of an artefact (or kind of artefact) that it is public or social, in some sense. What might we mean by this? In this section I've been discussing artefacts whose character depends on the recognisability to an audience of its creator's intentions. Thomasson (2014) takes this kind of recognisability and casts it as the mark of artefact publicness. She also argues that *most* artefacts are public in this sense, including, for example, hammers and chairs.

---

[10] I suspect that artworks are generally public artefactual kinds in this strict sense.

To my mind, this attitude confuses being a public artefact with being an ECA, and also obscures the distinction between artefacts which are only contingently intentionally recognisable and those which are essentially intentionally recognisable (the ECAs).

Let's start with the second point. It is certainly true that hammers and chairs *are* generally intended to be recognisable by a large audience. Thomasson (2014) emphasises that even a pebble-turned-paperweight may be intended to be recognisable as subject to paperweight norms: if someone took the pebble and used it to mash potatoes, the owner of the paperweight could justly complain that a clearly recognisable norm has been violated. However, although hammers are typically intended to be recognisable by a large number of people as intended to be used for hammering, this is not obviously an essential feature of hammers. One could make a hammer which is only intended to be recognisable to a single individual, or which is not intended to be recognisable at all. It does seem conceivable that one could create a hammer merely by successfully intending to make something for hammering. A second point is that the features of hammers which make them recognisable are just the same physical features which make them apt for hammering.

What distinguishes artefacts like wheelbarrows and paperweights on the one hand, and purely symbolic artefacts is that the former have to have physical properties which play a *direct causal role* in fulfilling their assigned function. Whether or not they need in addition to be recognisable as intended to play that role is up for debate. A brick wall fulfils its role of blocking access to an area by being an insurmountable physical obstacle. It can play its role whether or not anyone is aware that there is a wall there (somebody wearing a blindfold might charge into it by mistake.) Police tape is not like that. The only physical constraints on what plays the police tape role are those which require the artefact to be *recognisable as playing the intended role* (as well as being portable, robust enough to withstand a breeze, etc.). Invisible police tape won't work, but sombreros on washing machines could work as long as they were recognisable as being

intended by the police to block off access to an area (if, that is, people knew that sombreros on washing machines meant what police tape usually means). The police tape is part of the causal story explaining why passers-by keep away from the crime scene, but that causal story goes from perceptible features of the artefact via mental recognition of those features and the knowledge that those features are associated with an intention to keep members of the public away from crime scenes. High brick walls are also generally recognisable as boundaries that we might get into trouble for crossing, but even if they fail to be recognised as having this function, they can still do their job in virtue of their physical characteristics (height, solidity, etc.). If passers-by do not recognise that the police tape has a certain intended function and authority then they are liable to walk straight through it.[11]

On the second point, (that is, Thomasson's and Dipert's assertion that essential recognisability is the hallmark of *social* artefacts) my view is that there can be ECAs which are not intended to be recognisable by anyone other than an individual creator. In other words, there can be ECAs which only get to be the artefacts they are in virtue of the fact that the creator's intentions are recognisable and acceptable to a certain audience, where that audience is restricted to a single member, namely the creator of the artefact. For example, a secretive journal writer might write in using a cypher known only to them. So the fact that an artefact depends on its creator's intentions being recognisable does not necessarily make it public or social in character.

No doubt there are different things we might mean by talking of public artefacts. Here are a few of the things we might mean:[12]

---

[11] This is not to deny that some ECA kinds could be essentially associated with some practical, physical function in addition to their communicative function. For example, a military uniform is an ECA but it also has the function of keeping protecting its wearer from the elements.

[12] A further sense in which an artefact (or kind) could be held to be public would be if its existence depended on collective intentionality. In this thesis, I have chosen to bracket issues concerning collective intentionality, so let's just stick with the five senses above.

(i) The artefact belongs to a kind which is widespread. For example, there are hammers and paperweights all over the world.

(ii) The artefact kind is the work of many hands. In other words, there is no single inventor of that kind of artefact. Rather, it is the product of successive modifications over time. For example, the modern computer evolved over decades requiring the work of many different people.

(iii) People agree (even if the agreement is never expressed) about roughly what it takes to be a member of that kind. For example, there is a broad agreement about which things are and which things are not hammers. This is more or less the idea that the concept corresponding to the artefact kind is one which is shared.

(iv) The artefact is subject to agreed norms, conventions or laws. For example, the monetary status of a ten pound note is explicitly written into the laws of the UK. Hammers can't be taken into aircraft cabins.

(v) The character of the artefact is determined relative to an activity which requires the participation of at least two people, such as ten pound notes or footballs. For such artefacts, that they be successfully intended to be recognisable by at least two people becomes a constitutive requirement.

A hammer can be a public artefact in each of senses (i) to (iv), so it may make sense (at least in some contexts) to talk of hammers as public artefacts. Moreover, most ECAs are probably also public artefacts in sense (v). After all, symbolic artefacts are usually associated with activities which require coordination between more than one individual. For the most part, but not without exception, the existence of ECAs involves the backdrop of a community in which people are able to recognise and accept each other's intentions, and in which they are sensitive to public norms.

Summarising, my first point against Thomasson is that we ought to distinguish between symbolic artefacts which are grounded in a certain kind of intended recognisability (ECAs) and more practical artefacts. Though practical artefacts like hammers may be intended to be recognisable to a certain audience, that is arguably not an essential feature of hammers. Hammers can still be described as social or public objects, in first four senses above, but to be a hammer does not strictly require a mental contribution from anyone but a single creator. If I intend to make something for hammering, and it works, I've made a hammer. I don't need for anyone else to be able to recognise or accept my intentions. In light of this, I can avoid the unwelcome result – to which Thomasson was pushed – that the desert-islander marooned from birth turns out to be incapable of making a knife (even though they could create something which looks and cuts like a knife). My second point is that the publicness of an artefact does not immediately follow from the fact that it depends on the recognisability of the creator's intentions, since ECAs are sometimes only intended to be recognisable by their creator. However, there are various senses we can give to the notion of artefactual publicness.

The point of this discussion is not just to mark the differences between Thomasson's view and my own, but also to bring into focus the notion of an essentially communicative artefact, as well as to begin to elucidate the notion of artefactual publicness. The reason for attending to these categories is that they will help us in the next chapter to cast light on public language words.

4.4 Recognisability and exaptation

Why is a chopstick potentially a hair stick while a teapot placed on a stack of papers never becomes a paperweight? The notion of recognisability can help us to understand this puzzle about exaptation.

Hair sticks and teapots are not obviously public in the sense introduced above. They do not appear to depend essentially on being publicly recognisable: that a teapot is readily identifiable is an apparently contingent feature. It is recognisable

in virtue of its structural properties which (usually) enable it to play the role of brewing tea (or which have an aesthetic function).

While public recognisability is not an essential feature of these artefacts, the fact remains that teapots generally are publicly recognisable. And this recognisability brings with it various norms of treatment. If you take a teapot and use it as a watering can or a chamber-pot, you will be misusing a teapot. The fact that you will be misusing the teapot will be apparent to everyone in virtue of the structural features of the teapot. So while the various teapot-governing norms are not *essential* features of teapots, they can serve to block any intention to reappropriate the teapot as another kind of artefact. When someone puts a teapot on a stack of papers it is plain to everyone that it is a *teapot*. The teapot is being used to play the paperweight role in flagrant contravention of teapot norms. It is possible to determine, I suppose, that the teapot is currently intended by its user to play the paperweight role, but what comes through more decisively is the teapot maker's intention that the teapot be used for making tea.

Imagine a plain chopstick with no writing on it. When used to hold one's hair up, or when stored with other grooming related artefacts the chopstick is no longer recognisable as a chopstick. The structural features of the chopstick happen to be such that when used in accordance with a different intention (the intention that it hold one's hair up), that intention is not undermined: there are no features which signal that the object is a chopstick. (For what it's worth, if the chopstick had distinctive writing on it indicating that it was intended to be used for eating, my intuition is that the process of exaptation would be blocked. You couldn't turn *that* chopstick into a hair stick without modifying it.)

5. OBJECTIONS AND REPLIES

5.1 Automated production

Many artefacts are produced by machines. As a result, particular token artefacts are created despite no individual having any intentions regarding those particular tokens. For example, a bolt which is created by a machine but which falls

between cracks such that no one ever sees it is still a bolt. In response, we can point out that the never-seen bolt is ultimately connected to a creator's intention, though in a rather indirect way: the bolt is the product of a machine which has been designed by some artisan with the intention that it produce bolts. Thus, the never-seen bolt is connected to a kind of elongated creative intention. Dipert explores a similar theme:

> We often then have beliefs and expectations about how our intentions will or might produce changes in the world. Some of these are beliefs about more extended causal chains – through the mediation of our muscles, of tools or machines, or even through the expected or manipulated actions of other agents. If an event occurs through one of these (once-)contemplated chains or chain patterns, then the action is an intentional action. Similarly, if an intended change of properties occurs to an object through one of these chains, there is no reason to believe the property is less a tool or artefactual property than if the causal chain had been shorter. Thus, I can build a machine that builds and boxes computers, and the produced computers are intended to be recognized by others as having been intentionally made to fulfil the function of computing. Is such a computer an artefact? Yes, and those properties of the computer that I made the computer-making machine impose on its products are "intended." I am the agent whose agency is tied to those properties if my intention that computers have those properties brought about products having those properties, even if those properties were brought about by intentionally constructing a machine that more mediately caused them. (Dipert, 1993:128)

This certainly seems right. From the point of view of human agency, there is no good reason to discriminate between events which are brought about directly and those which are brought about through some longer chain of events initiated by an agent with a good idea about the likely effects. Nevertheless, the existence of artefact-producing machines raises a further problem for the intentional account. Suppose that an engineer programmes the bolt-making machine to produce 1000 new bolts, but that due to an unforeseeable malfunction the machine produces 1001 bolts. Nobody intended the extra bolt to exist, but that doesn't mean that it *isn't* a bolt. First of all, which of the 1001 items is the one which is allegedly not a bolt? It is not obvious that it is the last one. Allowing that all 1001 items are bolts, perhaps what we should say is that the designer of the machine intended

that it should produce a bolt whenever it is in a certain state. Although the machine might be in that state accidentally, what happens when it gets into that state was perfectly well intended by the designer.

5.2 Over-intellectualisation

Epstein (2013, 2015) has objected to much work in social ontology on the grounds that it over-intellectualises the processes of social construction. Taking the example of money, he argues that what money is can't be settled by appealing to the contents of people's beliefs and intentions regarding money since people are in fact fairly ignorant about money. What money is, Epstein suggests, is closely tied to the functioning of the banking sector, something which is poorly understood, even by experts.

This is a robust challenge to the intentionalist account, and to the standard model of social ontology more generally. Since my goal here is to apply the standard approach to words, I won't try to answer this challenge in full, but I will make two brief points (one of which I come back to at greater length in chapter five). First, it is not part of my project to claim that all social facts can be explained by the kinds of mental/social construction which I've been talking about. Nor do I necessarily need to claim that all *artefacts* can be explained in such a way (see §5.3). Moreover, *money*, as such, is not an artefact. Perhaps the concept of money is something like the concept of a recession. No one need have any conception of a recession for there to be a recession. But dollars are plausibly not like this. Of course, a dollar bill has various complex properties in virtue of poorly understood relations to the banking sector, but these need not be what makes the piece of paper a dollar bill. Second, on my account of artefacts, one should *expect* certain kinds of widespread ignorance about artefacts. Since I make this point in chapter five, I'll say no more about it here.

5.3 Not all artefacts

Epstein (2015) has stressed that attempts to give unified accounts of whole swathes of the social landscape often suffer by failing to comprehend the

intricacies of particular cases of social construction. He suggests that just as in biology where some kinds are teleonomic and others are just causal role functions, we should expect to find artefacts which are defined just by causal role. Elder (2007) has stressed the importance of attending to continuities between natural kinds and artefactual kinds. For example, beaver dams could be described as artefacts, but beaver dams seem to be more a manifestation of beaver biology than intention. Perhaps some of our human artefacts have a nature which reflects a biological function rather than the content of our own person-level intentions. Relatedly, Sperber (2007) has pointed out that human made biological artefacts are ubiquitous (e.g. cats, many trees and flowers, many types of fruit, seedless grapes, etc.). As noted above, the characters of such artefacts are not dependent on human mental representations. Finally, Lowe (2014) argues that technical artefacts (e.g. machines) depend not on human intentions, but on mind-independent natural laws which govern their operation.

Throughout I have been working towards the notion of an essentially communicative artefact. Strictly speaking, I don't need an entirely general account of artefacts as long as I can understand the class of artefacts to which words belong, since that is the focus of this thesis. Along the way, I've defended intentionalism about various kinds of artefacts including things like hammers and wheelbarrows. I do indeed think that intentionalism is a plausible account of the characters of these artefacts but I also recognise the kind of challenge posed by less intellectualised approaches to social ontology. What I have most confidence in is that the intentionalist approach is right for a certain class of symbolic artefacts – the ECAs – and in particular for words. In this chapter I have presented a plausible theory of how a material object comes to constitute an artefact of a given kind in virtue of (typically) structural features which communicate an artisan's intentions regarding that object. One of the reasons I'm sceptical that a biological/teleological approach could be right for these public, communicative artefacts is that the association of, say, the cross above a church door and the intended purpose of the building is entirely arbitrary. Similarly, the formal features of words are only arbitrarily associated with their

linguistic functions. As Millikan (2003) notes, this kind of arbitrariness marks a disanalogy with biological kinds.

6. CONCLUSION

I began this chapter by characterising *artefacts* as objects which are intentionally made and which have at least some intended features. Some such artefacts are *essentially intentional artefacts*. That is, they depend constitutively and not just causally on their creators' intentions. Curium atoms are artefacts but not essentially intentional ones. Within the essentially intentional artefacts, some are *essentially communicative artefacts*. These are associated with a distinctive mechanism of mental/social construction involving the intended recognisability of the artisan's intentions. Hammers are essentially intentional artefacts, but not essentially communicative ones. Examples of ECAs include chess pawns, traffic lights, and – as I will argue in the next chapter – words.

CHAPTER 2: WORDS AS ARTEFACTS

The first chapter presented an account of an important class of artefacts. I've been calling such artefacts ECAs, short for *essentially communicative artefacts*. This chapter applies the account of ECAs to words.

In §1 I show how to apply the account developed in chapter one to the case of words, and point out some of the advantages of the view, namely that there is a deep and fruitful analogy between ECAs and words, and that the view offers a neat explanation of widespread intuitions about swamp words, minimal creation and exaptation. In addition, I argue that the ECA view can be embedded within the philosophy of science and mind associated with generative linguistics in a way which recruits extra support for the view.

In §2 I address two major objections which hold, for one reason or another, that the intentions my story needs to get off the ground are not available. In response, I argue that even spontaneous, everyday speech is intentional - perhaps not in a full-blooded conscious and deliberative sense, but in a way which is analogous to playing a musical instrument, and which contrasts with e.g. what's going on in your immune system. Subsequently, I attempt to leverage insights arising from generative linguistics in the account of speakers' linguistic intentions.

In §3 I indicate the kind of approach to questions of word individuation which is suggested by the ECA view of words.

In §4 I deal with problem cases such as malapropisms, word-producing machines and accidental word production, ignorant artisans, and involuntary speech as well as a range of other objections.

1. APPLYING THE ACCOUNT TO WORDS

1.1 A fruitful analogy

If someone accepts the account of ECAs presented in the first chapter, then they are likely to feel tempted to apply the account to linguistic entities such as words. Indeed, words might seem to be *paradigmatic* examples of such artefacts. Applying the Thomassonian account to wedding rings and chess pawns on the one hand, and words on the other yields similar theoretical virtues and leads to similar problems with similar solutions. Moreover, there's a reason why the account is well-suited to words. ECAs have a symbolic or communicative function: they are in some sense language like. This is what makes the account well suited to words as well as non-linguistic artefacts such as sirens, shop fronts, etc. In short, part of the appeal of the ECA account of words is that there is a deep and fruitful analogy between words and other (non-linguistic) artefacts. To the extent that the ECA account is the right theory for such artefacts, there is some *prima facie* plausibility to the idea that words are ECAs.

1.2 Swamp words, minimal creation, exaptation

Further support for the ECA view of words accrues from its ability to explain fairly widespread responses to certain kinds of thought experiments. Consider the following two cases:

(i) A swamp emits a gaseous belch that sounds just like the word 'lake'.

(ii) On a distant planet which has never been visited by any intelligent life forms there is a lake, and near the lake there is a rock whose surface has eroded in such a way that it is an intrinsic duplicate of some Earthly carving of the word 'lake'.

Do the scenarios described in (i) and (ii) feature genuine tokens of the word 'lake'? I have a clear intuition that they do not, and anecdotal evidence suggests that I am not alone. Moreover, similar intuitions have been widely expressed in philosophy of language (references below). It appears to be part of the concept

of a word that its instances are products of intentional human activity. This, it is worth emphasising, is exactly what the ECA theory of words predicts. But why, we might ask, should this intuitive, everyday notion of words be taken seriously? Why should we pay any more attention to folk-linguistics than physicists pay to folk-physics? In response, attending to the ordinary conception of words is appropriate if what we're doing is trying to make sense of everyday conceptions of language as an aspect of human culture, and if we assume that the beliefs and intentions of ordinary language users can be partially constitutive of social-level linguistic phenomena. Thomasson, in her various papers on artefacts, explores the idea that the folk are – taken as a whole – protected from massive error regarding certain aspects of the social world. As Searle (1995:32) puts it, with regard to certain social facts, "you can't fool all the people all the time." If that is right, then careful conceptual analysis can be an appropriate guide to the theorist of social phenomena.

We can describe the phenomena in (i) and (ii) as swamp words. According to the ECA view swamp words are not words at all (just as fake money is not money and toy ducks are not ducks). To these cases of swamp words we can add a further case:

(iii) Suppose that human pioneers arrive on the planet. They notice the rock and comment on the coincidence. There even appears to be an arrow shaped indentation pointing the way from the rock to the lake. Then they make a decision: they decide to treat the naturally occurring rock as a sign. The improvised signpost goes on to serve generations of settlers looking for an out-of-the-way swimming spot.

My intuition is that once the human pioneers have decided to treat the eroded rock as a sign, or perhaps a little later when the community starts to believe that the rock is a sign, the rock really is a sign. What has changed? The physical properties of the eroded rock do not change. It is not necessary for anyone to modify the eroded rock by tidying up one of the letters or underlining it. What

change are the beliefs and intentions of humans. Case (iii) is an example of minimal creation (discussed in the previous chapter), whereby a natural object is transformed into an artefact with intentional features by what appears to be a pure act of will.

All of this suggests that speaker intentions play a constitutive role with respect to word utterances. Similar cases have been described in print before. Putnam (1981:1-5) describes a case in which an ant crawling in the sand happens to trace a likeness of the words 'Winston Churchill'. He notes that, intuitively, the lines left in the sand do not represent Winston Churchill. On the other hand, if the lines had been traced by a human being who had seen Churchill and had the skill write in English, and if they had intended for their creation to represent Churchill, then the lines would have represented Churchill. The lesson, for Putnam, is that for a physical object such as an inscription or an acoustic blast produced during speech to represent another object, it is necessary that it be intended to do so:

> [S]uppose the line had the shape WINSTON CHURCHILL. And suppose this was just accident (ignoring the improbability involved). Then the 'printed shape' WINSTON CHURCHILL would not have represented Churchill, although that printed shape does represent Churchill when it occurs in almost any book today. So it may seem that what is necessary for representation, or what is mainly necessary for representation, is intention. (Putnam, 1981:2)

> A discourse on paper might seem to be a perfect description of trees, but if it was produced by monkeys randomly hitting keys on a typewriter for millions of years, then the words do not refer to anything. (Putnam, 1981:4)

The case of the monkey and the typewriter introduces a wrinkle which I will iron out in the objections below (§4.2). For now, just concentrate on the ant trail which looks like an inscription of the words 'Winston Churchill'. I share Putnam's intuition that the ant trail does not represent Churchill, but I also think the thought experiment supports the view that the ant trail isn't even a word. In the quotation above, Putnam appears to allow that the monkeys on typewriters

do produce tokens of English words. It's just that they don't refer to anything (because they were not created with the intention that they refer to anything.)

Stebbing's distinction between a mark and a mark-as-used expresses something like the same point as that made by Putnam, but she, like me, draws the conclusion that unintended objects cannot be word tokens:

> A spoken token-word is a sound to which someone has attached a meaning. A sound is not a word unless someone has attached a meaning to that sound. (Stebbing, 1935:11)

Here, Stebbing is assuming that having a meaning is essential to being a token of a word. This is not obviously true: there could potentially be token words which do not have a meaning, such as grammatical words, interjections, frivolous words like 'jabberwock', or a list of made-up words such as 'blarg' which are expected to be used to name as yet undiscovered sub-atomic particles or galaxies. I think that what the thought experiments strictly support is that in order to be a word token, an utterance has to be the product of some human intention, not that it has to have some intended meaning.

I think that this intentional criterion succeeds in drawing a line between genuine word tokens and non-tokens in a sharp and intuitive way. (Various putative counterexamples will be discussed and defused in §4).

Dennett (1990) objects to intentionalism about artefacts (and specifically about words), claiming that creator intentions cannot be as important as the account makes them out to be because artefacts can be repurposed by later users. Such phenomena were described in chapter one (§2.2) as cases of exaptation. The crucial thing is to recognise that using an artefact can be a kind of minimal creation, where an existing artefact is recreated as a new artefact of a different kind in virtue of the creative intentions of the user. We can imagine that a Swedish inscription of 'god' (which is etymologically unrelated to the English word 'god', and means *good*) could be removed from its original context and used

in an inscription in English of 'a Roman god'. Plausibly, in its new context the inscription is a token of the English word 'god'. It now has a different set of linguistic features from that intended by the original creator of the inscription. Still, we can describe the creation of the English inscription as a case of minimal creation. In virtue of the intentions of the artisan of the English inscription, what was an inscription of a Swedish word becomes an instance of a new artefactual kind – the English word, 'god'. In light of discussions in chapter one, we can reply to Dennett as follows: the intentionalist approach to artefacts predicts that exaptation can occur and explains it in terms of minimal creation, a phenomenon for which there is already a precedent among non-linguistic artefacts.

1.3 Bringing the internalist and externalist perspectives together

Some of the advantages of the ECA theory of words have already been advertised. For one thing, I've suggested that there is a deep and fruitful analogy between words and other ECAs such as traffic lights, coins, wedding rings, etc. If someone accepts that there are ECAs at all, then they may be tempted to regard words as ECAs. Relatedly, viewing words as ECAs helps explain intuitions about swamp words, minimal creation and exaptation. In this sub-section I'm going to suggest that the ECA view receives additional support from insights in generative linguistics.

This is a claim which requires extreme caution. After all, generative linguistics is strongly associated with many forceful critiques of public language. To be clear, my claim is *not* that generative linguistic theory *requires* the articulation of an ECA theory of words. What I mean is the following. First, the ECA theory of words is compatible with a non-revisionary interpretation of generative linguistic theory. Second, the ECA theory of words can be fruitfully embedded within the philosophy of science and mind which animates generative linguistics. Third, the best Chomskian critiques of public language are those which assume something like my ECA theory of words. Attributing this assumption makes these critiques plausible and pertinent. These critiques do not motivate any kind of eliminativism regarding words construed as ECAs. Rather, they recommend

shifting theoretical attention away from such artefacts an inwards upon the natural features of human mind/brains. The nature of the Chomskian critiques of public language is the topic of chapter five, so we can set them aside for now.

Let me explain why I think my ECA view is in harmony with theorising in generative linguistics. The linguist insists that their focus is on cognitive states of individuals, not on some mind-external linguistic reality. This focus is on cognitive states which are invariable across changes in the physical signal. So, for example, there is a linguistic state associated with the sentence 'cows eat grass' which is invariable whether or not one is confronted by an inscription, or an utterance, or one is just thinking 'cows eat grass' in interior monologue. Whether or not some acoustic blast or patch of ink is a token of 'cows eat grass' cannot be determined just by inspecting the acoustic blast. Being a token of 'cows eat grass' is not a property of the blast *qua* blast. Instead, speakers *project* linguistic properties on to externalia.

Collins (2010:48-50) (whose work we will return to in chapter five) makes remarks to the effect that externalia "are invested with linguistic significance," and that the language faculty enables the "projection of structure onto sounds and marks." The point of his discussion is precisely to show why a naturalistic perspective on language will pursue an internalist approach: the source of linguistic properties, the story goes, lies in mental structures; to the extent that externalia can be viewed as having linguistic significance, this is because mental structure is projected onto them. The trouble is, at this point, talk of "projection" is highly metaphorical. What exactly is "projection"? Does it result in the creation of a new kind of (mind-external) entity? Under what conditions? Is this a phenomenon which is restricted to language, or does it occur in non-linguistic domains as well?

Without doubt, Collins would reply that there is no need to settle any of these questions from the perspective of generative linguistic theory. Exactly right. Whether we take seriously the idea that externalia really do have linguistic

properties in virtue of some kind of "projection" is of no consequence for the internalist approach in the cognitive sciences of language. But one way of construing the relation between the ECA theory of words and standard assumptions in generative linguistics is to see the former as providing a substantial account of the notion of "projection" as it is informally used in linguistic theory. My position is that the projection of linguistic properties onto externalia is just another case of making something – an artefact – and that the ways in which an utterance acquires linguistic properties is analogous to the ways ECAs get their distinctive characteristics.

Recall that, on my account, a chess pawn is what it is roughly because it is intended to be used in a certain way on a chess board. No physical analysis of chess pawns could reveal what makes them pawns independently of the beliefs and intentions of humans. We could say that a certain kind of chess-significance is projected onto inanimate pieces of wood. In short, the ECA theory of artefacts involves an account of how minds can impose properties on inanimate objects in a way which seems tailor made to explain the linguist's notion of "projection" of linguistic properties on externalia.

Sadly, it appears that I am not the first person to think of this marriage of linguistic theory with the idea that human intentions can project properties onto inanimate objects. Barber summarises his approach to these matters as follows:

> This approach takes the mental representations that are instanced during the performance or perception of an utterance to represent real aspects of the event, i.e. actual linguistic entities and features, including referential features; but this reality is constituted of intersubjective intentions rather than acoustics. A PRO (and not just your representation of a PRO), for example, genuinely figures in the event that is your uttering *Mary told John to eat*. It does so by virtue of your (tacitly) intending your audience to represent the utterance-event as instancing PRO, though not in the sense that you intend PRO to be audible in the acoustic stream… For you to intend your audience to interpret your act as instancing PRO *is* for that utterance to instance PRO… And if a hearer represents the utterance-event as instancing a PRO, or a reference to food

> ingestion, she does so *correctly*, since your intending her to do so (assuming for simplicity that she is your intended audience) is what it is for the utterance-event to have these linguistic characteristics. Our internal representations are thus subject to the bar of mind-external, or at least intersubjective, reality, and to that extent are representations properly so-called. (Barber, 2013:974)

Barber seems to imply that intending for an utterance to instance PRO is sufficient for an utterance to instance PRO. That cannot be quite right, and it can't be quite what he means either, since he writes elsewhere:

> [T]he acoustic stream…cannot have a particular syntax solely because the utterer intends it to do so... Something extra must happen… [I]n order for a hearer to understand a speaker, she must recognize which sentence the speaker intended the hearer to treat as the object of interpretation. This is an epistemic claim, but there is nothing to prevent us from supposing that a constitutive fact underpins it, namely, that the identity of the sentence uttered turns on the intention of the speaker to be recognized by his or her audience as having produced that sentence. (Barber, 2006:27-28)

This view is strikingly like my own. Nevertheless, Barber's treatment of the view leaves many questions unanswered. Chief among these is the objection that the kinds of mental states posited in linguistics are too 'thin' to count as intentions. This is one of the problems to be dealt with in the next section.

Summarising, I believe that my ECA theory of words is precisely that which a Chomskian ought to adopt, if they care at all about language as a mind-external phenomenon. They are not obliged to care. They are already doing good work on the nature of I-language and theorists do not have to be interested in everything. But there is no incompatibility here. Chomskians can allow that we have person-level beliefs and intentions about language. They can allow that linguistic properties are "projected" onto externalia. Adopting this view does not prevent them from pursuing the internalistic inquiry. In fact it agrees with them that an internalistic inquiry is the correct way to explain facts about public language. The main differences between me and the public language deniers are these:

(i) While generative linguistics quite rightly aims for integration with neurobiology, I believe we should also pursue integration between generative linguistics and the social sciences.

(ii) Generative linguistics, again quite rightly, seeks to get theoretical traction by abstracting away from many of the details about public language, and a major part of the Chomskian critique of public is just a sceptical attitude towards the prospects for making any real progress with inquiries into messy, social-level domains. Against this methodological scepticism, I would urge that ECA words *can* be fruitfully studied from a social scientific perspective.

All of this will be explained at greater length in chapter five. In this section I have been sketching the features of the ECA account of words and extolling some of its virtues. In the next section I discuss two major lines of resistance to this approach.

2. INTENDING WORDS

In my experience of presenting these kinds of views to philosophers, there is a single worry which is always raised in various forms. It concerns the plausibility of the claim that word production is a genuinely intentional activity.

One form of this objection targets the idea just introduced that we can treat the mental states posited in linguistic theorising as something like intentions, and as being responsible for imbuing externalia with linguistic properties.

A different kind of objection is that ordinary speech can appear spontaneous and unreflective in a way which strains the description of utterances as artefacts shaped by intention. Take the words a person utters when they ask for the bus ticket which they buy every single day, or when they excitedly relate the details of some anecdote. Do speakers form an intention to produce every single word token that they do produce? We might feel that speaking – or at least the aspect

of speaking which consists in producing words – is something we do automatically, without paying too much attention.

I will tackle the second kind of objection in §2.1. This presents an opportunity to discuss the kinds of intentions which could play the required role in a theory of artefacts. Then in the rest of this section I'll give the background to the first kind of objection and explain my response.

A third kind of objection is that it is possible to imagine a host of problem cases and apparent counter-examples to the idea that word production is always intentional. These cases include malapropisms, word-producing machines, apparently involuntary speech, and the like. These kinds of objections will be considered along with other objections to the ECA view in §4 of this chapter.

2.1 Speedy and spontaneous intentions

According to the ECA view of words, language use is an intentional activity. Speaking, signing and writing are things that we do as agents. To a certain extent we do them consciously and on purpose and we typically have reasons for doing them. We can also be held responsible for speaking the way we do or for not speaking when we should. In other words, people have person level thoughts, beliefs and intentions regarding their utterances, and they typically intend to utter the words they do. But people typically regard speech as a fairly spontaneous activity, and one which is engaged in by small children. It is therefore worth making a case for the claim that speech is intentional under verbal descriptions. This is important because it is a precondition on the account offered in chapter one that our verbal activity be describable as intentional. Speakers could not intentionally impose certain linguistic features on their utterances if uttering was not an intentional activity.

Heck (2006:2) agrees that speaking is an intentional activity, but notes that "[s]aying that speech is a form of rational (or intentional) action leaves open the question under what descriptions it is intentional." A single speech event could

be described as telling a student they've passed a test, saying the words 'you've passed,' using pulmonary pressure to vibrate the focal folds, etc. The question is where to draw the line between the intentional and the unintentional aspects of speech. Heck argues that speech is rational under *verbal* descriptions such as saying the words, 'it's raining.' In this sub-section I'm going to follow Heck's argument for this conclusion.

In the first part of his paper, Heck argues that speech is typically intentional under a propositional description:

> [S]peech—or, more generally, our use of language—is intentional under propositional descriptions, such as: saying that p. So, to return to the example I used earlier, when I tell my wife that I love her, saying that I love her is something I do; my utterance is intentional under that description. (Heck, 2006:12)

In the second part he goes after a different claim, that speech is intentional under a verbal description:

> But this description of the action is not the only one under which it is intentional. It is also intentional under verbal descriptions, such as: uttering the sentence "I love you". Heck (2006:12)

To argue for this Heck introduces a thought experiment involving people who are just like us except that their speech is not intentional under any verbal description (although it is intentional under a propositional description). What we have to do is imagine people who are as much like us as possible except that their speech is not intentional under any verbal description. If the people who are like that are not like us, the argument presumably goes, then our speech is intentional under some verbal description. That is, if in order to clearly imagine people whose speech is not intentional under such a description, we have to imagine people who are different from ourselves, then our speech must be describable under intentional descriptions, because otherwise we would have to have whatever feature it is that the imagined people have which means that their speech cannot be described that way (or lack the feature they lack).

Heck thinks that such people would be *quasi-telepaths*. A quasi-telepath can tell someone it's raining just by intending to tell them. That's all they have to do. After that some non-intentional mechanisms trigger various events in the subject's vocal apparatus which emits a sound which sounds to you and me just like the words 'it's raining' but the quasi-telepaths are stone deaf to the frequencies of human speech, so they don't hear anything, but it doesn't matter because quasi-telepaths don't need to hear the acoustic signal. They have auditory organs which process the sound below the level of consciousness and simply deliver to the subject the thought that it's raining. They have no awareness of moving their mouths to utter words, nor of hearing any utterances. Communication between quasi-telepaths, Heck (2006:13) tells us "would be like telepathy: it would seem to them as if communication were purely between their minds."

We're clearly not quite like the quasi-telepaths. One difference is that their vocal and auditory apparatus doesn't work like ours. We can hear our own utterances and those of our interlocutors, and speakers recognise that they and their interlocutors have uttered certain words. So it feels different to be a telepath, and being a telepath means lacking some knowledge of one's environment – the words being uttered in it – which we ordinarily take ourselves to have. Heck is claiming that one of the key differences between us and the telepaths is that they are not conscious of their verbal behaviour. It is because of this that the quasi-telepaths' verbal behaviour is not intentional. How could they intend to produce an utterance of 'cat' when they have no experience of utterances and no conception of the role they play in communication? The role played by utterances in communication is as epistemically hidden from them as the role of the antibodies in our immune systems is hidden from us. What do you have to do to turn a quasi-telepath into a normal speaker? Roughly, you have to give them a conscious awareness of their utterances and their linguistic properties and the ability to decide how to use words.

On the basis of these kinds of contrasts with the quasi-telepaths, Heck concludes that:

> These reflections suggest that speech is not only intentional under verbal descriptions but that it is by uttering a sentence that we say something. This remark is not intended simply as one about the causal structure of communication. Taken that way, it would apply equally to the quasi-telepaths I discussed above: Communication as they have it also depends upon the production and reception of sound; in a purely causal sense, they say things by uttering things, too. What distinguishes us from them is that, in a rational (not just a causal) sense, we say things by uttering things. (Heck, 2006:14)

The key notion here is that we can rationalise speakers' behaviour under verbal descriptions: the speaker says the words 'it's raining' because they want to say it's raining and know that 'it's raining' means it's raining. But we can't do that for the quasi-telepaths because they aren't even aware of their utterances.[13] For Heck, the key contrast between us and the quasi-telepaths is that we are aware, as they are not, of the meanings of our word utterances, and we have rational control over the utterances we produce. For us then, uttering a word is something that can be evaluated from a rational point of view. If a speaker utters a word we can expect that they mean something by it and that they want to express the thing that the word means.

The conclusion of the argument is modest: language users *know* things about words, principally – for Heck – what they mean. We might add that people also know other things about words, such as how they should sound, what makes a good rhyme, and how to use them in combination with other words. And people *do* things with words, like describe things, think out loud, make promises, give orders, cajole, reassure, entertain, name ships, cause offence, abuse other people, etc. The conclusion may be modest but its significance for the theory of

---

[13] (Heck (2006:15) also notes that the quasi-telepaths "do not utter sentences at all (though noises do get made)" Is Heck claiming – like me – that having certain word-features by intention is essential to being an utterance of a word? He repeatedly states (Heck, 2006:4,13) that the connection between intentions and word utterances is something he is defending on empirical, observational grounds. But the empirical claim that we have intentions towards our word utterances does not entail that the natures of our utterances depend constitutively on those intentions. On my view, the reason why the quasi-telepaths do not produce utterances at all is that since they have no intentions regarding their word utterances (being congenitally unaware of their existence), their utterances do not have any intended features.

communication is great. Dressing it up in slightly grandiose language we might express the idea as follows: our verbal behaviour – unlike the behaviour of our immune systems – is person-level behaviour; it is in the realm of human intentional activity, free will and rationality.

The worry mentioned earlier can now be appreciated in full relief. Heck is telling us that saying words is an intentional activity requiring conscious knowledge of the meanings of words and their possibilities for combining with each other. Some people will worry that this is implausible. After all, speech is often very speedy, spontaneous and unreflective, and small children are good talkers. The idea behind the objection is thus that it requires too much cognitive sophistication on the part of ordinary language users. If a reply cannot be found, then there is no room to extend the ECA account of artefacts to words, because the kinds of intentions which are characteristic of artefact production are simply not available.

I think the solution to this problem is to allow that the sense of intention which is required for artefactual creation may be less full-blooded than the intentions we ascribe to a great chess player moving a pawn. That full-blooded notion of intention may be correctly applicable to the intentions of the chess player putting their opponent in check in a prestigious match, or a surgeon removing a kidney. Such activities are highly deliberate and involve conscious decisions and inferences. But many things we would be prepared to call actions don't have those features, or, at least, not in quite such a full blooded way.

People sometimes report a sensation of realising that they have been driving on 'autopilot', as if they were not consciously, intentionally controlling the car. Though there may be a contrast between the mentality of such a driver and that of the chess player, it would be too much to say that the actions of turning the car around corners and stopping at traffic lights are entirely unintentional. Even in the case of the surgeon or the chess player, their actions will likely be guided by beliefs and desires which they are not fully conscious of. For example, the

surgeon is presumably so experienced that they may avoid certain pitfalls and errors with the scalpel without consciously thinking about those dangers. That doesn't mean that their skilful moves with the scalpel are unintentional.

Even more illuminating analogies are to be found in the form of playing the piano, slip catching in the sport of cricket or touch typing. (Happily, the last example here is one which involves producing tokens of words, albeit written ones.) The actions of a proficient pianist, sportsperson or typist have the spontaneity and rapidity which is characteristic of speech. They do contrast with the deliberate actions of the chess player or surgeon, but they contrast yet more sharply with events in the immune system such as selecting an antibody to deal with a certain microbe. The cricketer catches the ball intentionally, even though they only had a split second in which to react. If they dropped it, they were nonetheless *trying* to catch it. Did the typist deliberately hit the keys in a certain sequence? The natural answer to this is 'yes'. Did they deliberately select a certain antibody to fight off a microbe? Of course not. Catching a ball, and hitting certain keys on the keyboard of a piano or typewriter are things that people do. They are not events over which they personally have no control. I think speech is the same. The way speakers use words may be spontaneous and speedy and may not involve conscious deliberation, but it is intentional activity nonetheless.

In an attempt to characterise the nature of the mental states underlying actions which are intentional, though perhaps less than fully conscious, Heck selects an example from ordinary life:

> If I walk to the refrigerator and open the door, what explains my doing so may be my wanting a beer and my believing that there are beers in the fridge. And if someone asks me why I opened the refrigerator, that is just the sort of thing I might say: I and others explain my actions by adverting to such beliefs and desires; we do so rightly, even if my mind was elsewhere at the time. The point is one that ought to be familiar: The correctness of this sort of rational explanation does not depend upon the agent's awareness of his own practical reasoning (though it may depend upon his having conscious access to his reasons). (Heck, 2006:9)

He explains that a similar point applies to verbal behaviour:

> [I]f one asks why Smith uttered the sentence "The meeting begins at 4pm", the question may be answered as follows: Smith wanted to say that the meeting began at 4pm, and he knew that the sentence "The meeting begins at 4pm" meant that the meeting began at 4pm and so that, if he uttered that sentence in that context, he could thereby say that the meeting began at 4pm…[I]n offering this sort of explanation of Smith's utterance, I do not mean to suggest that Smith must consciously have engaged in any such reasoning. Sometimes we do; more usually, we do not. That fact does not undermine the claim that what explains Smith's uttering what he did is his having the sorts of beliefs and desires mentioned. (Heck, 2006:15)

Heck's suggestion seems to be that the notion of intentional action may require conscious access to one's reasons without requiring conscious, occurrent knowledge of them at the point of acting. Heck remarks that:

> [T]he word "conscious" is significant. It is notoriously slippery, as well, and I would gladly use an alternative if only I could think of one. I use it mainly in contrast with "tacit". (Heck, 2006:10)

In other words, the appeal to conscious accessibility should not be taken to imply full-blooded consciousness on the ordinary understanding of the word. Instead the role of the requirement that one's intentional activity be guided by reasons to which one has conscious access is mainly to steer clear of the notion of *tacit knowledge*, a notion which has picked up technical connotations in philosophy. 'Tacit knowledge' and 'implicit knowledge' are pieces of philosophical jargon. Sometimes they get used to talk about phenomena which are only distantly related, if at all, to common sense understandings of knowledge. For example, Evans (1981), accepts that speakers have tacit knowledge of a semantic theory for their language, but denies that this involves attributions of genuine propositional knowledge of the theory. Evans's account requires only that the inferential structure of a theory mirror the structure of the causal states responsible for meaning attributions:

> Tacit knowledge of the syntactic and semantic rules of the language are not states of the same kind as we identify in our ordinary use of the terms 'belief' and 'knowledge'… Such concepts as we use in specifying it are not concepts we need to suppose the subject to possess… There is thus no question of regarding the information [as] being brought by the subject to bear upon speech and interpretation in rational processes of thought. [The notion that knowledge of the syntactic and semantic rules of the language] is a real species of belief, but with all the relevant inferential processes made by the subject somehow taking place outside his ken…is certainly a mysterious and confused position. (Evans, 1981:133-134)

In short, what seems to be characteristic intentional action, on Heck's view, is that an event be explainable in terms of a genuine propositional attitude, a contentful mental state, on the part of the agent. Such a state need not be fully, occurrently conscious, but it should be the kind of thing that one could bring to consciousness if prompted in the right way. To say that we have knowledge of language in this sense is a *much* stronger claim than to say we have tacit knowledge of language is Evans's sense. That kind of tacit knowledge is far too weak to enter into an account of intentional action.

It is still possible to dispute the idea that verbal behaviour is intentional and based on conscious knowledge in the way Heck suggests. As noted, small children and people who have never heard of a noun can be competent speakers and word producers. How can such people intend that an utterance be recognisable as a e.g. *singular count noun*? However, the fact that someone lacks the vocabulary to express their reasons cogently doesn't prevent them from being the reasons which explain their actions. The idea that we have propositional attitudes and intentions regarding words which are consciously accessible (and much richer than 'tacit' attitudes) does not entail an ability to articulate that knowledge.

2.2 Mental representation in linguistics

In the previous sub-section I worked up a kind of deflated sense of intention which seems to be sufficient for intentional action and which seems to accurately

characterise our verbal behaviour. This means we can now pursue Barber's suggestion that the states posited in generative linguistics can be leveraged in an intentionalist account of words. Is it plausible that, as Barber suggests, speakers intend an utterance to feature an instance of PRO? In line with the previous sub-section, if the suggestion is to work, the PRO-regarding intention will have to be contentful and accessible to conscious though not necessarily easily articulable. (I don't think Barber would contest this. Recall that he says the mental states which imbue externalia with linguistic properties must be "representations properly so-called.")

Bromberger and Halle (2000) tell us that a phonologist's theoretical representation of the phonological event which occurs when a speaker utters a token of 'the merchant sold shelves' is a description of a set of *intentions* on the part of the speaker. For example, in the derivation which Bromberger and Halle provide, each letter of the International Phonetic Alphabet which features in the final line of the derivation stands for a cluster of articulatory intentions.

(*a*) {[ðə], Art … } + {[mərtʃənt], Noun … } + {Q, Sing … } + {[sɛl], Vb … } + {Q, Past … } + {[ʃelf], Noun … } + {Q, Plur … }

(*b*) {[ðə], Art … } + {[mərtʃənt], Noun … } + {Q, Sing … } + {[sol], Vb … } + {Q, Past … } + {[ʃelv], Noun … } + {Q, Plur … }

(*c*) ðəmə·tʃntsoldʃelv z

<div align="right">

Bromberger and Halle
(2000:23)

</div>

They write:

> My production of event (1) [an utterance of 'the merchant sold shelves] was an action. Like other actions, it was therefore brought about by a distinctive kind of mental set – something we will call an INTENTION. But this term, as we use it, is not to be taken altogether literally. We use it to refer to a familiar kind of purposive mental stance… The uttering of [the sentence 'the merchant sold shelves'], like the aiming of a rifle, also required a distinctive mind-set, distinctive intentions on my part, intentions that I could not have formed without certain pre-existing intellectual capacities… We construe [a structural description of a string in terms of phonological features] as standing for a series of intentions that generated [certain articulatory] movements. Each letter…stands for

such an intention, and each of these intentions called for an arrangement of articulators in the expectation of distinctive auditory effects. The m [in a theoretical description of the cognitive state associated with an utterance of 'the merchant sold shelves'] represents an intention to act so as to produce a specific ENGLISH SPEECH SOUND. (Bromberger and Halle, 2000:23-24)

The parallels between, on the one hand, Bromberger's and Halle's conception of the role of phonological intentions in the production of speech, and the kinds of mental activity which underlie musical or sporting performance are striking: as noted above, such mental activity is speedy and spontaneous, and less than fully conscious; Bromberger and Halle (2000:35) observe in relation to phonological behaviour, "speakers are not aware of performing such actions. But then we perform many actions like zombies." Bromberger and Halle hint that they might share my view that speakers' linguistic intentions play a role in fixing the linguistic properties of utterances, shaping them not just causally but also constitutively:

Spoken or subvocally produced tokens are transitory events that are finite in number, that occur in time and space, that can be perceived, that are shaped by their speaker's occurrent intentions, and that are subject to norms fixed in their speakers' mental make-up. (Bromberger and Halle, 2000:20)

Such a view would harmonise with and enrich the intentionalist proposal recommended in this thesis. One problem is that what Bromberger and Halle and Barber call 'intentions' are rather unlike intentions as they are ordinarily conceived. As discussed above, there is a sense in which one can act intentionally without any conscious awareness of one's intentions, but the idea that speakers could intend for their utterance to instance PRO seems to stretch the point too far. Though the insights of generative linguistics are sometimes presented via slogans which talk about 'knowledge of language,' Chomsky and others (Jackendoff, 2002, Pylyshyn, 1988, Collins, 2005, 2008, Egan, 2003, 2012, 2014) have asserted that such talk is loose talk, suitable only for informal presentation

of the view.[14] A grammar drawn up by a linguist is intended to model an aspect of cognitive competence, but the relation between the grammar and the cognitive states themselves may be more abstract than many philosophers have thought. It may be no part of the theory to claim that a grammar is known to language users or that the concepts of the grammatical theory are concepts which individual speakers possess. To model a cognitive capacity of an individual using a theory described in terms of phonemes, for example, is not necessarily to say that the individual has the concept *phoneme*, or has thoughts or beliefs *about* phonemes.

Rey (2003a, 2003b, 2005, 2006b, 2012) stops short of insisting that the theoretical posits of generative linguistics are ordinary knowledge states (on the grounds that knowledge is not a sufficiently robust notion to enter into serious scientific theorising), but he defends an interpretation of generative linguistics in terms of a "computational representational theory of thought", or CRTT:

> CRTTs can be regarded as proposals to treat thought processes as causal/computational processes defined over representations that are encoded in our nervous systems on the model of processes in a computer. Various propositional attitudes, whether they be knowing, believing, thinking, or cognizing are to be distinguished by the different causal relations they bear to those representations as they are processed in the brain's cognitive architecture. In this way CRTT promises to provide the "mechanism" many have sought to link mental and physical phenomena, providing a mentalistic explanation of rational – as well as of much non-rational – behavior. (Rey, 2003b:4)

This would mean that the mental states posited by linguists have representational content. The job of the linguist is to understand what the contents of these representational states are, how it is that ordinary language users conceive the

---

[14] Not all linguists, it would seem, share this scepticism about the representationalist interpretation. Smith (1999:139-140) writes that "[a]t the heart of [the Chomskian] account is a computationalist view that postulates the existence of different kinds of representation and argues that cognition consists in carrying our computations over them… As with his realism, Chomsky's representationalism and computationalism are both construed literally. He talks of "what is actually represented in the mind of an individual" and writes that "the brain uses such notations as …,"… He further makes the general claim that the mental systems allow "something akin to deduction as part of their computational character," a position which presupposes both representation and computation."

linguistic signals around them. However, even this is vigorously denied by Chomskians. In short, one might object to the attempt to leverage the kinds of states posited in generative linguistics in the ECA account of words on the basis that the states posited in linguistic theorising are too thin to be properly considered as intentional.

This means that Barber's account is not adequate as it stands. He wants to take a proposed grammar (a linguist's theory of an individual's linguistic competence, including semantic, syntactic and phonological components) and treat it as a theory of a speaker's intentions. But so much of what is studied in linguistics is completely inaccessible to consciousness and scarcely describable as a person-level. A *person* doesn't consciously obey a C-command constraint any more than a *person* produces immunoglobulin to neutralise some pathogen. A lot of linguistics is about things which just happen in an individual's mind/brain, rather than being about what people do.

2.3 Intending and cognizing

Instead of foisting a marginal interpretation on to linguistic theory, forcing I-language states into the mould of intentions, I suggest we take the posits of linguistic theory as they are. One thing is for sure, people do have person-level beliefs and intentions regarding linguistic externalia. Some use of language is highly conscious, deliberative, even ponderously so. People who write speeches for presidents, or write novels or PhD theses spend hours agonising over the precise words and sentences to employ. Also, consider looking a word up in the dictionary, rehearsing it in your mind to solidify the memory, racking your brains the next day to recall it, carefully dropping it into a conversation in order to impress one's interlocutors, etc. Similarly, consider trying to parse a long sentence in a Dickens novel, or obsessing over whether a friend's comment was intended to create a hurtful implicature.

Nobody who is party to the present debate is denying that such person-level, conscious states are involved in linguistic performance. Wholesale eliminative

materialism about folk-psychological states is not what's at stake here. Another point which Chomskians will agree with is that I-language states have got *something* to do with person-level beliefs and intentions. Whatever I-language states are, they're the kinds of things which have a causal effect on person level intentions and perceptual states, otherwise it would be impossible to maintain that person level judgements about utterance acceptability, synonymy, etc. constitute evidence for the structure of I-language states, an assumption which is ubiquitous in linguistics.

The problem is not that the cognitive processes which fall within the remit of generative linguistic theory are unconscious, unintentional, unlike ordinary propositional knowledge, etc. Some are undoubtedly remote from the states of folk psychology, but others may be more familiar sounding states including communicative intentions. Generative linguistic theory is concerned with the human capacity to pair sounds with meanings over an unbounded range. Some aspects of this capacity no doubt involve ordinary propositional knowledge which is consciously accessible, inferentially promiscuous, etc. And many aspects of linguistic performance involve communicative intentions. But in order to make progress, theorists are obliged to, as Collins (2010:46) puts it, "fractionate the phenomena and idealize away from the massive interaction effects that produce our normal linguistic behaviour." Linguists want an account of linguistic cognition, and they make no assumption in advance of inquiry that ordinary notions such as intention, belief, knowledge will enter into the theory.

This was the point of the old jargon which has it that speakers *cognize* – as opposed to know – a language. The purpose of this is to bracket the common sense associations of words like 'know' in order to pursue inquiry into a stipulated notion which reflects the concerns of theory. The point is that saying we are studying a subject's linguistic *cognizance* is not the same as denying that part of what we're studying could be described as ordinary, consciously accessible knowledge. Strictly speaking, the theory remains neutral on that.

I think there is a tendency to picture the language faculty as a black box, cut off from other cognitive systems. In fact, Hauser, Chomsky and Fitch allow that language "in its broadest sense" may include most of human cognition:

> [A] neuroscientist might ask: What components of the human nervous system are recruited in the use of language in its broadest sense? Because any aspect of cognition appears to be, at least in principle, accessible to language, the broadest answer to this question is, probably, "most of it." Even aspects of emotion or cognition not readily verbalized may be influenced by linguistically based thought processes. (Hauser, Chomsky, Fitch, 2002:1570)

From a biolinguistic perspective, inquiry cannot proceed into "most of" the human mind/brain. It's necessary to idealize away from the messy detail of a whole human mind in order to focus on questions which are precise and which have some hope of being answered (in a way which is explanatory, rather than merely descriptive). As Collins, writes:

> Any endeavour to gain theoretical traction on this manifold is obliged to fractionate the phenomena and idealize away from the massive interaction effects that produce our normal linguistic behaviour… The first move in this 'divide and conquer' direction in recent times was Chomsky's (1965) distinction between competence and performance. A whole range of factors enter into performance, many of them perfectly general, such as memory, attention, and communicative intention. Chomsky's distinction, in part, was an effort to isolate the hypothesised unique linguistic system that underlies certain peculiar features of our performance, and, in the first instance, we are interested in those features simply because they are the ones that submit to theoretical understanding. (Collins, 2010:46)

So one way in which linguistics idealises in order to frame a tractable inquiry is to ignore the communicative intentions which attend specific instances of verbal behaviour. The focus is on competence – the relatively stable state which constitutes an individual's capacity to generate an infinite number of linguistic expressions. Another way in which linguistics idealises is by stipulating a distinction between narrow and broad conceptions of the language faculty. To see this, it is really worth quoting Hauser, Chomsky and Fitch at some length:

> We…delineate two more restricted conceptions of the faculty of language, one broader and more inclusive, the other more restricted and narrow. *Faculty of language—broad sense (FLB).* FLB includes an internal computational system (FLN, below) combined with at least two other organism-internal systems, which we call "sensory-motor" and "conceptual-intentional." Despite debate on the precise nature of these systems, and about whether they are substantially shared with other vertebrates or uniquely adapted to the exigencies of language, we take as uncontroversial the existence of some biological capacity of humans that allows us (and not, for example, chimpanzees) to readily master any human language without explicit instruction. FLB includes this capacity, but excludes other organism-internal systems that are necessary but not sufficient for language (e.g., memory, respiration, digestion, circulation, etc.). *Faculty of language—narrow sense (FLN).* FLN is the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces. FLN is a component of FLB, and the mechanisms underlying it are some subset of those underlying FLB… The internal architecture of FLN, so conceived, is a topic of much current research and debate. Without prejudging the issues, we will, for concreteness, adopt a particular conception of this architecture. We assume, putting aside the precise mechanisms, that a key component of FLN is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system, and into the conceptual-intentional interface by the (formal) semantic system… All approaches agree that a core property of FLN is recursion, attributed to narrow syntax in the conception just outlined. FLN takes a finite set of elements and yields a potentially infinite array of discrete expressions…Each of these discrete expressions is then passed to the sensory-motor and conceptual-intentional systems, which process and elaborate this information in the use of language. Each expression is, in this sense, a pairing of sound and meaning… One goal of the study of FLN and, more broadly, FLB is to discover just how the faculty of language satisfies these basic and essential conditions. (Hauser, Chomsky, Fitch, 2002:1570-1571)

FLN is treated as a specialised module, but one which interfaces with other cognitive systems (sensorimotor and conceptual/intentional). One way of thinking about this is that phonology targets the interface between FLN and the sensorimotor system, while semantics targets the interface with the conceptual/intentional system.

How is consciousness distributed across such systems? Well, it's not for a linguistic theory to say. 'Consciousness' is not assumed to enter into the theory any more than 'knowledge' is. It goes without saying that some FLB states will involve ordinary, conscious knowledge and communicative intention (e.g. trying to remember a recently-learned word, or speculating about what someone meant when they said 'visiting relatives can be boring'). But whether any aspect of FLN is accessible to consciousness is difficult to know. My guess is that most of it isn't, but that is not something which we can read off from linguistic theory.

None of this is to deny that a competent language user has consciously accessible knowledge regarding the correct use of an external system of signs. Nor is it to deny that generative linguistic theory provides insights into such a person-level capacity. It looks like some of the states targeted in linguistic theory are impenetrable to consciousness but not all. As a point of methodology, linguistics just abstracts away from the full detail about what people do and how they use linguistic externalia and whether the states targeted by the theory could be described as ordinary states of propositional knowledge.

In short, linguists do not deny that speakers have some ordinary propositional knowledge of language, or that intention is an important aspect of communication. Nor do they affirm that linguistics studies something which is entirely separate from all of that. The question for the ECA account of words, then, is this: just which aspects of our linguistic behaviour are intentional, and which are not? This is the topic of the next sub-section.

2.4 Intentional under what descriptions?

I don't know of any principled way of determining which parts of language use are intentional, or which involve ordinary conscious knowledge. As noted, this cannot be read off from linguistic theory, which remains neutral on this point. One informal way of approaching things is to proceed by example. Drawing on Heck (2006:13-14) we can immediately give the following examples:

(i) We are aware that words can be misunderstood. For example, we know that someone can intend to say that Jones is angry but actually come out with the words 'Jones is livid'. There might be someone in the vicinity who knows that 'livid' means *pale* and also that people sometimes get confused about this. In a quasi-telepath conversation, one person would simply intend to convey that Jones was angry and one's interlocutor would grasp that that was what you intended to convey. There couldn't be a mismatch of meanings where one person associates 'livid' with *angry* and another with *pale* because quasi-telepaths do not associate meanings with utterances, for they are not aware of utterances.

(ii) We recognise when we and others have misspoken, perhaps by some kind of "processing error" resulting in a spoonerism or malapropism. If a quasi-telepath means to convey 'loving shepherd' but there is a physiological malfunction and they produce a sound we would recognise as 'shoving leopard' the quasi-telepath interlocutor would grasp the unintended meaning and would have no idea about what might have gone wrong.

(iii) We are often aware that our understanding is being impeded by not knowing the meaning of a neologism or foreign word.

(iv) Sometimes we are unable to resolve an anaphoric ambiguity. If someone were to tell you that 'Anne told Brian to cook Carol's dinner and to tell her when it was ready,' you may wonder – quite consciously – whether Anne wanted Brian to tell *her* or Carol that the dinner was ready.

(v) If someone says "fighting administrators can be distracting" their interlocutor might wonder which of two sentences was intended. They might be able to resolve the ambiguity by attending to pragmatic factors or just by asking. The speaker will probably be able to say which sentence they intended.

To these examples I would add the practice of creating or choosing between compound words ('unputdownability,' 'smog,' 'publicity'/'publicness,' etc.),

94

noticing 'grammatical errors' in the speech of politicians or second language learners, noticing the differences in vocabulary and syntax between different dialects of English, etc. These examples (and those in (i) to (v)) illustrate that speakers have conscious awareness of a variety of morphological, syntactic, semantic and pragmatic features.

A slightly different way to approach the question about which aspects of linguistic behaviour are intentional would be to lean on debates about knowledge of meaning. The idea that competent speakers of a language know a theory of meaning has been widely discussed in recent decades. A great deal of discussion has focussed on the *sense* in which speakers can be said to know a theory of meaning. That is, what is at stake is whether semantic competence should be construed as ordinary propositional knowledge, or something else (such as Evans's notion of tacit knowledge, discussed above). Heck (2006), as we observed above, argues that the rational evaluability of speech acts requires consciously accessible propositional knowledge of meanings, though in that paper he remains neutral on exactly what is known. As Barber (2013:965) explains, it is standardly assumed (i) that knowledge of the meanings of linguistic expressions is necessary but not sufficient for knowing what is meant by a particular utterance of that expression, (ii) that the theory of meaning known by competent speakers is compositional, and (iii) that it takes the form of a referential theory of truth. If something like this approach is right, then one aspect of linguistic behaviour which is intentional concerns the intended referential profile of utterances.

One aspect of linguistic behaviour which we have not broached concerns phonology. There can be no doubt that a good deal of phonology enters into the realm of intention and conscious awareness. This is illustrated by our sensitivity to rhyme and alliteration, the pleasure we take in wordplay and punning, our ability to make judgements about where someone grew up based on their accent, or the practice of adopting a prestige accent when speaking on the phone. Equally, we can become aware that our inability to determine what words have

95

been said is due to failing to hear a certain sound. For example, if someone coughs just when you say the 'g' in 'I've bought a goat,' I may be aware that my uncertainty about whether you've said you've bought a goat or a coat stems from not knowing whether the first sound in the word is a 'g' or a 'c'. Jackendoff adds the following examples:

> [N]otice first of all that we experience language as perceived sound. We can intuitively divide utterances into words and syllables with ease (children can count syllables on their fingers by three years of age). Without too much training, we can even divide the speech stream pretty well into individual speech sounds; this is an essential part of learning to read. We have pretty good intuitions about stress patterns as well: most people can easily say where the main stress of a word lies, though they may be uncertain about subsidiary stress. Note however that not all aspects of phonological structure result in discriminable qualia: the decomposition of speech sounds into distinctive features is experientially opaque. (Jackendoff, 2007:81)

In sum, it appears plausible that much – but not all – phonology involves conscious, intentional action on the part of speaker/hearers. Jackendoff is right to caution that the phonological features which compose phonemes are not consciously detectable. *These* do not qualify as intended features of utterances, though stress, rhyme, and phonemic structure plausibly do.

Heck suggests that once we get to the level of describing speech in terms of movements of the vocal organs we are not describing intentional activity:

> [I]t is not intentional under the physiological description, nor under the acoustical one: Neither moving my tongue and lips in that particular way – physiologically described – nor making that particular noise – acoustically described – was something I did intentionally... [E]xactly how my mouth moves, and exactly what sound emerges, are not things under my rational control. (Heck, 2006:2-3)

It is true that the precise frequencies of the vibrations of one's vocal folds are not something one has control over. However, there is no sharp line here between the intentional and the unintentional. I think that the physiological

articulations of words are intentional in a very real sense. In his play *Le Bourgeois Gentilhomme*, Molière captures the experience of learning some basic phonetics:

> PHILOSOPHY MASTER: The vowel U is formed by bringing the teeth nearly together without completely joining them, and thrusting the two lips outward, also bringing them nearly together without completely joining them: U.

> MONSIEUR JOURDAIN: U, U. There's nothing truer. U… Ah! Why didn't I study sooner in order to know all that!

Like Monsieur Jourdain, when it is pointed out to students in an introductory phonetics class that to say 'food' you have to round your lips and push them outwards they are often surprised and intrigued. I suggest that the intrigue which attends the realisation that saying 'food' involves pushing one's lips forward is due to the fact the one is becoming fully conscious for the first time of something that one has been doing intentionally all along. Quasi-telepaths would experience the phonetics lesson in *Le Bourgeois Gentilhomme* quite differently from the rest of us. For them there would be no moment of realisation like that when Monsieur Jourdain tries the vowel out for himself and feels the positions adopted by his lips and tongue. They would be learning about phenomena which are as impenetrable to them as the workings of our immune systems are to us.

I have not mentioned orthography so far. That literate people have conscious knowledge of orthographical systems, and that they use them intentionally is, I think, beyond question. Indeed, that inscriptions are ECAs is probably the most secure part of the present theory.

What this discussion yields is an informal characterisation of the kinds of linguistic properties which enter into the conceptions ordinary language users have regarding linguistic externalia. These can include orthographical, phonological, morphological, syntactic, semantic, and pragmatic features of words. The upshot of this is that ordinary speech can be regarded as intentional for the purposes of the ECA theory of words. What this doesn't tell us is which

of these properties are individuative of word types. This is a difficult topic, and it is the focus of the next section.

3. TYPES OF WORDS

3.1 Word pluralism

One obstacle in providing a theory of words is that common sense and theoretical discourse have yielded a multiplicity of word conceptions. Julien (2006:617-619) distinguishes *phonological words*, *lexical words* and *grammatical words*. Further senses of 'word' are distinguished in McArthur (1992:1120-1121), but these are the three conceptions – along with that of the orthographic word – which crop up again and again in morphology textbooks.

An *orthographic word* is a sequence of letters with a space on either side. This is the notion of word which is often used by word-processing programmes to count the number of words in a document. According to this conception **word-processing** is a token of a single orthographic word while **word processing** features two tokens, each of a different orthographic word; **fysujhjdgoksonfdmgl** is a legitimate token of an orthographic word, while **processing** and **processsing** are legitimate tokens of different orthographic words. These are reasons enough to draw the conclusion that the orthographic conception of words is not what we're after.

Sentences are typically produced as a continuous stream of sound. Pauses are not usually employed to mark boundaries between different words (and pauses frequently occur in the middle of words, such as before a plosive consonant.) Still, there is a notion of a *phonological word*, which consists in discrete domains for phonological effects such as stress. On this conception of word, utterances of 'I'll' (or other contractions such as 'won't') will not be utterances of two words, but of a single word. Even larger phrases such as 'to the shop' may count as a token of a single phonological word. Moreover, a slip of the tongue may result in an utterance which is not a genuine utterance of any English word despite being a perfectly good phonological word (perhaps one never tokened before). Again,

this is not the conception of word which is of interest to philosophers. One problem with the orthographic and phonological conceptions is that we are looking for a sense of word in which *both* an utterance and an inscription can be a token of that word. Adopting only the orthographic and phonological conceptions of words would not allow us to maintain a sense in which a word can have both spoken and written instances.

One prominent conception of words in theoretical linguistics is that of the *lexeme*. A lexeme is an abstract unit underlying a set of words which differ from each other only in their grammatical forms. Intuitively, the lexemic conception corresponds to items which would be grouped together under a single listing, either in a speaker's mental lexicon or – from the point of view of a lexicographer – a dictionary. For example, there is a sense in which the words 'be,' 'am,' 'is,' 'are,' 'was,' 'were,' 'been,' and 'being,' are different grammatical forms of the word 'be'. But there is another sense of 'word' according to which each of these is a different word. This is the notion of the *grammatical word*. This is the perspective which emphasises the properties of words which are relevant to morphological and syntactic processes. In this sense, 'be,' 'am,' etc. are different words. They would be considered as the various grammatical words corresponding to the lexeme associated with the citation form 'be'. Similarly, using the notion of grammatical word we can distinguish between two different 'being' words, one of which is the gerund, the other the past participle. The distinction here is grounded in syntax.

The grammatical word is characterised in terms of a position in the structural hierarchy of linguistic categories between morphemes and phrases. A morpheme is typically defined as the smallest unit of language with its own meaning. A word is made up of one or more morphemes. Julien (2006:620) canvasses two criteria to distinguish words from morphemes. The first is that "an expression that can stand alone as an utterance [aside from in metalinguistic discourse] is normally no smaller than a word." So, for example, 'that' is a word, since it could stand alone as an answer to the question 'what do you want to eat?'. In contrast, the

morpheme 'un-' would be dispreferred by native speakers as an answer to the question 'are you happy or unhappy?'. (Someone who answered this way would be making a kind of linguistic joke.) The second criterion concerns "freedom of position, or independent distribution." For example, the word 'the' can appear next to a noun in the phrase 'the linguist', but it does not *have* to be attached to a noun in this way. It can appear before an adjective in the phrase 'the famous linguist.' In contrast, the morpheme 'un-' can't be so easily separated from whatever word it is bound to: we cannot say 'un-very-happy'. Additionally, as Hawthorne and Lepore (2011:456) note, "words are self-contained phonological units; morphemes are not [always]. Words have at least one syllable, their own stress, and so on. Morphemes need not." Clitics such as plural '-s' in English illustrate this point. '-s' is a morpheme, but not a word.

To distinguish words from phrases, Julien proposes the criterion of "indivisibility or internal cohesion… [E]ven if both words and phrases can be built from words, with phrases it is normally the case that they can be broken up by additional words and phrases, whereas words that consist of words cannot be interrupted in this way." For example, the phrase 'a black bird' can accommodate an insertion to yield 'a black and beautiful bird' while the word 'blackbird' in 'a blackbird' cannot accommodate any such insertion.

Julien (2006:618-619) suggests that "the core meaning of the term 'word'" is the grammatical word, an assumption which is fairly common, I think (see also Matthews, 1991:24-31). For Julien, this is because this is the notion that "has to do with the role that the word plays in the morphology and in the syntax," and it is also the concept of word which is salient to ordinary language users. Something about this seems right. The notion of grammatical word does seem to track the common sense conception reasonably well. However, if the correct understanding of grammatical words involves the idea that they are individuated purely in terms of grammatical properties, then this would seem to be too narrow to capture the ordinary notion of word. Plausibly, speakers make word type

judgements which are sensitive to phonological and semantic properties as well as syntactic ones (as I argue in the next sub-section).

3.2 The intuitive basis for word type discriminations

Kaplan (1990) wanted an answer to the following question: what makes two utterances utterances of the same word? I should say from the outset that my answer to this question is far from comprehensive. It's thin on detail in certain places. But the answer I have to this question is not empty either. I think it gives the structure of an account of word individuation. What the present account amounts to is a framework within which to understand word individuation, a framework which is common to non-linguistic artefacts and which was outlined in the first chapter. As I will argue in chapters two and three, the accounts offered by Cappelen (1999) and Kaplan (1990) don't even offer an adequate framework within which to pursue these questions.

In a slogan, the ECA account of words holds that individual utterances are tokens of a given word type because they are intended to be. Of course, this is the provocative way of expressing the view. As discussed in chapter one (§3.3), this should only be construed as loose talk. If not, the view has a nasty kind of self-referentiality. Wetzel (2008) objects to the intentionalist approach as follows:

> Any account of what an intention-to-utter-'cat' is will probably presuppose some account of what the word 'cat' is... [This] is like defining 'the' as "anything the community accepts as 'the'." We haven't gotten anywhere. Wetzel (2008:69-70)

The response to this objection parallels that given in chapter one. To be sure, defining the word 'the' as anything which is intended to be the word 'the' is no more informative than defining a surf-board as anything which is intended to be a surf-board. The correct response is the one canvassed in the first chapter. We have to give an account of the content of speakers' linguistic intentions in a way which breaks a word concept (e.g. the concept of the word 'cat') down into its component features. The speaker's intention is not best construed as the intention to say the word 'cat'. Rather the speaker associates a range of properties

with a given word. Some of these properties relate to the forms of utterances and inscriptions, others concern the meanings of words and their possibilities for combination with other words. It's by successfully intending to produce something which has (at least some significant cluster of) these properties that the speaker produces a word.

The previous section (especially §2.4) tried to catalogue some of the intended linguistic properties of utterances. These included orthographical, phonological, morphological, syntactic, semantic and pragmatic properties. This is in keeping with the idea that a word pairs a form with a meaning and can combine with other words in specific ways. Plucking an example from a nearby book, the word 'chapter' is a singular count noun; according to the Oxford English Dictionary, it has nine different senses (and numerous sub-senses) including a "main division or section of a book" and a "branch of an organization or society." In the minds of speakers, 'chapter' is associated with one or more pronunciations specified in terms of phonological structure, and possibly with one or more spellings. Now, this is no doubt oversimplified, but it gives a good enough idea of the kinds of properties which are potentially individuative of word types.

What kinds of linguistic properties are individuative of words? Meanings sometimes seem to enter into word type discriminations. This, presumably, is what explains the intuition that there are two different 'bank' words while 'chapter' is regarded as polysemous. After all, the two bank words are indistinguishable from the point of view of syntax, phonology and orthography. However, meanings do not appear to be treated as essential features of words. Centuries ago the word 'meat' meant food in general, rather than referring exclusively to flesh. As long as the pronunciation hasn't changed too much, the judgement that one and the same word used to mean food but now means meat is sustainable.

Nor are syntactic properties of words generally regarded as essential properties: the phenomenon of turning nouns into verbs is ubiquitous in English (e.g. 'to

table a proposition'), and the result is intuitively not a new word but a new usage of it[15]; the word 'that' can be used as a pronoun, a determiner, an adverb or a conjunction, but a given occurrence of 'that' will have only one of these syntactic properties.

Phonological and orthographic forms are clearly relevant to judgements of word type. Suppose (somewhat implausibly) that the Old English word 'scyrta' has the same semantic and syntactic profile as its modern descendent, 'shirt' but that its pronunciation is very different. I suspect this makes them different words. Or, suppose that English speakers take a shine to a Russian word whose meaning has been widely misunderstood, so that it is used in English with a very different meaning and perhaps a completely different syntax. What explains the judgment that English speakers are using the Russian word? One answer is a certain kind of historical continuity, which is Kaplan's view, a view I criticise in chapter four. But amongst the synchronic, linguistic properties of utterances, the only explanation of the judgement is (near) sameness of form. There are, however, formidable difficulties in supposing that words are individuated in terms of their forms. This is the topic of chapter three so I won't say too much here. Problems include: the fact that a word can be articulated in either speech or writing; the existence of dialectal variation in pronunciation (and orthography) both synchronically and diachronically; tolerance to imperfect word production; problems of formal coincidence (e.g. there are two 'bank' words).

In light of the above, the prospects for definitions of words in terms of necessary and sufficient conditions do not look bright.[16] These problems echo those encountered in chapter one (§3.4). There, I discussed Bloom's observations that both the form of an artefact and its function seem relevant to the categorisation of artefacts and yet neither the form nor the function typically associated with an artefact kind need be regarded as essential to that kind. There is no single form

---

[15] This seems to be in tension with the claim that the ordinary conception of word is that of the grammatical word.

[16] See Wetzel (2008:58-71) for a full length defence of this conclusion.

associated with boats and yet form is not irrelevant since the rubber sphere pulled by dolphins is judged not to be a boat despite having the function of transporting people over water. Nor can we state any function which is either necessary or sufficient for being a boat: the boat shaped device intended for unmanned retrieval of marine biological data is typically judged to be a boat.[17]

A further analogy with certain kinds of artefacts is that some utterances are regarded as imperfect instances of a given word type. We can recognise a drunken slur as an instance of a word, but a marginal one. Similarly, we might judge that a modern word was in use in Chaucer's England, though utterances produced back then are marginal instances of our word. These kinds of judgements suggest that words are associated not with necessary and sufficient conditions, but with some cluster of significant properties.

When these problems arose for artefacts generally, I was obliged to consider that many artefact kinds are not *strict artefact kinds* in Thomasson's sense (§3.4). Instead, for many artefact kinds, the best we can do is to characterise them in terms of Thomasson's "Dependence Principle" for loose artefact kinds:

> Necessarily, for all x and all artifactual kinds K, x is a K *only if* x is the product of a largely successful intention that (Kx), where one intends (Kx) *only if* one has a substantive concept of the nature of Ks that *largely* matches that *of some group* of prior makers of Ks (if there are any) and intends to realize that concept by imposing K-relevant features on the object. (Thomasson, 2003b:26)

Assuming, then, that words can only be characterised as loose artefact kinds, the kind of account we can provide of word individuation is rather limited. What we *can* usually do is provide a set of sufficient conditions for tokening a given word. Suppose I produce an utterance which is a successful realisation of my intention to say something which is recognisable to English speakers as being intended to have the following properties: singular count noun; refers to a main division of a book; pronounced /ˈtʃaptə/ with stress on the first syllable. In that case I will

---

[17] Bloom thinks the function of providing transport over water is too general to distinguish boats from other kinds of artefacts.

have succeeded in saying the word 'chapter' Or, take the word 'cat'. Anything which is successfully intended to be recognisable by English speakers as a singular count noun referring to cats and has a pronunciation which starts with an aspirated velar plosive followed by a short, near-open, front, unrounded vowel and then an apico-dental plosive…is an utterance of 'cat'.

It is much harder, however, to say which, if any, of these properties is *necessary* to being an instance of 'cat', but – shadowing Thomasson's account of loose artefact kinds – we can impose the same kind of necessary matching constraint which was discussed in the previous chapter. Having established a set of sufficient conditions for being an instance of a given word, we can insist that anything which is an utterance of 'cat' must be the product of a successfully realised intention whose content largely matches that just elaborated. In short, the present account of word individuation does not offer a characterisation of word types beyond indicating a kind of family resemblance structure.

I confess that this account is barely satisfactory. It's easy enough to state sufficient conditions, and the idea that making something of the same kind means enacting an intention with roughly the same content is intuitive enough. But nothing that has been said so far allows us to pin down any of the details of a theory of word individuation. After all, my utterance of 'cat' might be intended to be very loud or to be said in a singsong voice. Nobody has to match my intentions in *these regards* in order to count as saying 'cat'. The account so far is quiet about just which properties are relevant to word individuation, except insofar as I fall back on plausible assumptions relating to semantic, syntactic and phonological properties. A further problem is that the account relies on the notion of a family resemblance concept, or cluster concept, notion which raises difficult issues in cognitive science which I don't propose to resolve. Nonetheless, I think there are three things we can say in defence of this part of the theory:

(i) In keeping with observations made above, the formal properties of words (especially phonological but also orthographic properties) seem to be accorded more weight than syntactic or semantic properties in speaker's word type discriminations.

(ii) It is quite typical of artefact kinds that we are unable to provide definitions of them in terms of necessary and sufficient conditions. So the fact that we have to put up with characterisations in terms of family resemblances, and the problems of vagueness which ensue, is precisely what you'd expect if words are artefact kinds.

(iii) What we can do is show how a theory of word individuation should go. What the ECA theory of words does is provide a description of the metaphysical framework within which questions about word individuation can be pursued. Rival theories of words fail to do even this much. As I will argue in the next two chapters, not only do the theories of Cappelen and Kaplan fail to yield adequate theories of word individuation, their accounts are fundamentally unable to provide any such account. For example, by forcing words into a framework inspired by Searle's XYC schema, Cappelen incurs a requirement that we be able to provide two independent type descriptions of a given word. And Kaplan's account in terms of shared origins can't explain the very fact that words have vague boundaries. (These two points will be explained fully in chapter three §4 and chapter four §1.5, respectively). In short, though I'm not offering a detailed and realistic account of word individuation, or a way of adjudicating hard cases, I do at least have the resources to describe how a theory of word individuation should go. Vague boundaries are inconvenient, and I can't tell you where the borders are, but when we have clear cut judgements of word type identity or non-identity, I can explain those by pointing to the intended linguistic profiles of the various utterances.

4. ILLUSTRATIONS, HARD CASES, OBJECTIONS

The ECA view regards the association of intentions and words not as a mere empirical generalisation but as a matter of conceptual necessity. Such a view naturally invites criticism by counterexample. If it can be shown that genuine word utterances can be produced where no appropriate intentions are present then the theory is in trouble. In this section I will consider numerous potential counterexamples to the view, along with other kinds of objections.

4.1 Malapropisms and other errors

I read a text recently where the writer used the phrase "to dissuade a fear". Presumably they meant "to assuage a fear", but that is not what they wrote. They appear to have written 'dissuade'. This creates a potential problem for the ECA view because one might think that the person wrote 'dissuade' despite not having the right kind of intentions: how can the utterance be a token of 'dissuade' if there is no recognisable intention to produce a thing which means dissuade?

My preferred way out of this problem is to maintain that the speaker did intend to say 'dissuade', it's just that they had some false beliefs about what that word means. Though they have some erroneous beliefs about the meaning of 'dissuade' (they think it means assuage) they still count as largely matching the conceptions other people have of the word dissuade. There is, after all, more to a word than its meaning.

An alternative response could be that the utterance was a token of 'assuage' despite looking like 'dissuade.' I find this option very unappealing. It would be a kind of Humpty-Dumpty-ism. It would be to ignore that there are conventionalised ways of signalling one's linguistic intentions. A final option would be to deny that the utterance was an utterance of any word. Such a point of view is not indefensible: one could argue that the speaker does not know how to make the 'assuage' intentions recognisable, and that they are don't have an accurate enough conception of 'dissuade' to produce a token of that word either (because they're wrong about the meaning). This account finds intuitive parallels

in the realm of non-linguistic artefacts. If a mason intends to produce a slate tile for a roof but slips and ends up with shards of slate which are intrinsic duplicates of arrowheads, the shards of slate are not arrowheads.[18] Still, in the case of the clumsy mason, the shards of slate are produced by a genuine accident. The mason had no intention to produce the shards. The intrinsic similarity to arrowheads was a freak accident. The utterance of 'dissuade' is not like that. It is the product of the same "targeted gymnastics" (to use Bromberger's (2011) phrase) which characterises normal speech. Moreover, the speaker's intentions concerning the syntactic and phonological properties of 'dissuade' may be assumed to match other speakers of the language closely. I therefore prefer the first response: the speaker did say 'dissuade' in virtue of a successfully realised intention whose content *largely* matches that of other competent speakers of the language. (That the speaker also manages to convey to their interlocutors the content that some fear has been *assuaged* seems to me compatible with their having uttered the word 'dissuade' and not the word 'assuage'. See Predelli (2010) for discussion.)

Similarly, suppose someone believes that 'cat' is spelled 'catt' (example due to Predelli (2010). Their inscription of 'catt' presumably counts as a token of 'cat'. Though the writer has a conception of the word which partially diverges from that of other 'cat' users, their conceptions are probably similar enough. Competent users of 'cat' will easily recognise the linguistic intentions of the writer.

A different kind of case occurs where a speaker is well apprised of the various norms of use for the relevant words but where some kind of mental or physiological malfunction intervenes. These can include slips of the tongue such as spoonerisms or drunken slurring, as well as typing errors. For example, someone who knows perfectly well how to spell 'representation' might slip while typing and end up writing **representatyion**. This, I take it is a

---

[18] Although they could become arrowheads through a process of minimal creation.

token of 'representation'. Competent users of this word will easily recognise the intended word, and understand how the typo may have occurred. Successful communication of speaker intention is largely unhindered.

Take the case of the Sheriff of Nottingham who (in a Mel Brooks film) says "Struckey has loxed again" instead of "Loxley has struck again." The speaker's knowledge of the linguistic properties of 'Loxley' and 'struck' is impeccable, and they intended the utterance to be recognisable as intended to feature instances of 'Loxley' and 'struck'. Now, suppose it *was* so recognisable (in virtue of context and background knowledge). In that case, I think we should allow that the sheriff really did say "Loxley has struck again" despite the utterance's slightly atypical form. Why should we allow this? Why not just say that to be an instance of a given word speakers have to signal their linguistic intentions via one of the accepted (possibly local) forms?

The reason I don't think we can say this is that much of ordinary speech is so rapid and garbled that the resulting acoustic streams cannot be easily separated into word utterances closely resembling any conventionally accepted form. On the ECA view, this will not preclude their garbled utterances from being tokens of the intended words, as long as competent speakers are still able to recognise a speaker's intentions. So, we either have to allow that the recognisability of a speaker's linguistic intentions *by non-canonical means* is good enough for producing an instance of the intended word, or we have to deny that much ordinary speech features genuine tokens of words (or at least allow that there are frequent gaps). Taking the latter option would not be catastrophic (see the discussion of Rey's *folie à deux* view in chapter five, §1.1), but it would be somewhat counter-intuitive. But then, if recognisability by non-canonical means is good enough, then saying (what sounds like) "Struckey has loxed again" can be good enough for saying the words 'Loxley' and 'struck' as long as the speaker's linguistic intentions were recognisable by people familiar with the intended words.

This tolerance to error has limits. When Donald Trump tweeted **despite the negative press covfefe** it would appear that the physical form of the inscription combined with generally known aspects of the utterance context do not suffice for recognition of intended word type. One might *suspect* that the intended word was 'coverage', but this cannot be inferred with a high enough degree of certainty to count as a successful token of that type. (Intuitions about these cases will no doubt vary. It doesn't matter too much if we disagree about specific decisions pertaining to borderline cases as long as we agree on the kinds of criteria which establish the vague boundaries.)

A further tricky case concerns the Reverend Spooner's alleged utterance of something which sounded like 'shoving leopard' (though he intended 'loving shepherd'). The options here are to say (i) he says 'loving shepherd,' (ii) he says 'shoving leopard,' and (iii) he produces non words. Against (i), it would be nice to say Spooner really says 'loving shepherd'. But in fact, here (unlike in the 'Struckey'/'loxed' case) the process is blocked by the fact that the tokens closely resemble forms used to signal a different word. Against (ii), the present case is unlike that of the 'dissuade'/'assuage' example. Here there is a more kind of catastrophic error. The speaker really did intend to say 'dissuade' despite their slightly skewed conception of that word. In contrast, Spooner was perfectly well apprised of the linguistic properties of 'shoving,' 'leopard,' 'loving,' and 'shepherd'. And he seems to have ended up saying a word he really didn't intend.

My inclination here is to say that in such a case the speaker does not in fact say 'shoving leopard' even though it sounds like they do. This case is a bit more like the clumsy mason, though not exactly like it. As in the case of the clumsy mason, one suspects that the process of artefact creation has gone wrong. Control has been lost and the results have more to do with accident than intention. That Spooner's utterance sounds like 'shoving leopard' is just a hazard of the English language. Given his particular struggles with speech, Spooner would have ended up saying something that sounded like 'shoving leopard' even if 'shoving' and 'leopard' had not been words of English. The point is that the kind of accident

which creates spoonerisms is not generally likely to produce things which sound like genuine words except when entirely independent circumstances happen to obtain. The fact that Spooner says something that sounds like 'shoving leopard' is really just a freak coincidence.

Perhaps some production errors are what get called 'Freudian slips' reflecting some underlying thought on the part of the speaker. For example, perhaps the deformation which results in an utterance of 'it is kisstomary to cuss the bride' was triggered by the vicar's underlying dislike for the bride. In that case the phenomenon is akin to cases of involuntary speech considered below, and the solutions provided there will be apt. The vicar may have produced a genuine utterance of 'cuss', but that is explicable in terms of some real intention to say 'cuss' which, let's say, surges suddenly and displaces the intention to utter the word 'kiss'.

Finally, a sports journalist said of an athlete at the recent Olympic Games that "it is hard to underestimate their achievement." There's a sense in which the journalist did not intend to say exactly those words. This, however, is easily explained. What is clear is that the journalist did not intend to say that the athlete's achievement is hard to underestimate, but it can nonetheless be maintained that the journalist did intend to say the words "it is hard to underestimate their achievement" (though perhaps that wouldn't have been their intention if they had had a bit more time to think about it.)

4.2 Automation and accidental productions

In chapter one (§5.1) I discussed an alleged class of counterexamples to intentionalism about artefacts. These concerned the existence of automated machines for the production of artefacts, and the fact that such machines make possible a kind of accidental production of artefacts. (Recall the bolt-making machine which produces one bolt too many.) Since we have word producing machines similar objections are encountered by the ECA view of words. The solution in this case has the same pattern as in the general case.

The mere fact that production of word tokens involves technology is no obstacle to their being appropriately related to the linguistic intentions of the person who operates the machinery, or who orders it to be operated. As I type, many processes are occurring in my laptop which are not intended by me, but that doesn't mean that the words appearing on my screen are not the products of my linguistic intentions. I know how to make the computer produce items with a certain form which will be recognisable as tokens of specific English words, and I am intentionally causing the machine to produce just those words. The use of machines need not introduce any difficulties not already present in cases where someone uses a pen or their mouth to produce words. Still, there are cases which require additional discussion.

First, it is no doubt true that the people who operate printing presses are often unaware of the words they are printing. Similarly, I sometimes print out philosophy articles and then never read them. On the spoken side, a monolingual English speaker can produce Russian words by playing a CD. In such cases the person operating the machinery may or may not intend the printing press, printer or CD player to produce tokens of words, but they certainly don't have specific intentions about *which* words. These are cases of ignorant artisans. Here I think we need to appeal to the intentions of the original authors and speakers. *Their* intentions at the moment of typing or speaking are the important ones. The authors' intentions impose linguistic properties on tokens on their personal computer screens but these will get passed on to tokens which have been reproduced from the originals via a reliable enough process. Each time the CD is played, the acoustic event can be traced back through a causal lineage which connects up at some point with the intentions of the original speaker.

Second, suppose the staff of some newspaper decide to print 1000 copies of the paper, but that due to a mechanical fault the press continues running and produces further unintended copies. The resulting papers still feature genuine tokens of English words (even if one of the papers falls down a crack and is

never seen by anyone.) Similarly, suppose the ringtone on my mobile phone is a spoken-word poem. I mistakenly think I've put the phone on silent but when someone calls me (who doesn't know about my ringtone), my phone produces genuine tokens of English words that no one ever intended to exist. But the fact that the productions are accidental does not undermine the fact that they have been created via a reliable process which takes words produced by some author and reproduces them automatically, preserving the original linguistic intentions.

A different kind of case is the scenario in which a monkey hammering on a typewriter happens to produce something which looks like the complete works of Shakespeare. My intuition is that a monkey couldn't type out a Shakespeare play, quite apart from the statistical improbability. If by coincidence a monkey hammering away on a typewriter happened to produce an intrinsic duplicate of a copy of Hamlet – something that would fool a reader into thinking they were reading Hamlet – I think it would nonetheless fail to be an instance of Hamlet. In that sense, this is just another swamp word case. On the other hand, some people may have the intuition that the monkey really could type out a Shakespeare play. I want to explain that intuition away. One way to do that is to point out that the typewriter has been designed with various intentions in mind. Specifically, it is intended to produce a letter 'e' whenever the typewriter is in a certain state (one of the keys is in the down position). If that is right, then the monkey can cause an 'e' to be produced by unintentionally putting that typewriter in that state. We can thus maintain that the new object is a token of 'e' only because someone intended the typewriter to produce an 'e' under those circumstances. So we can allow that the monkey produces letters, but can we allow that it produces words? I think we cannot. Though the typewriter may have been intended to produce words, it was not intended by its creators to create specific words. So we have to deny that the monkey produces a copy of Hamlet, but we can at least explain the intuition that it does so by allowing that it does produce sequences of letters; these are real linguistic objects and, of course, when arranged in certain sequences they will look for all the world like the words of Hamlet.

4.3 Ignorant artisans

Another objection canvassed in the first chapter (§3.2) concerned the alleged possibility of ignorant artisans. A blacksmith in a country without horses could faithfully duplicate a horseshoe without knowing what a horseshoe is. I argued that in such cases we have to appeal to the intentions of the client rather than the intentions of the blacksmith.

The same problem can be raised against the ECA account of words, and the solution is again the same. If Anne, who is ignorant of Arabic, learns from Bahira, a native speaker of Arabic, to imitate the pronunciation of some Arabic word, then she can plausibly be said to have uttered the Arabic word. Since she has no conception of what the word means, whether it's a noun or verb, etc., this possibility apparently undermines the claim that an utterance is an utterance of a given word partially in virtue of the speaker's intentions. The obvious reply is to appeal to Bahira's linguistic intentions when she provided exemplars for Anne to copy. We can consider Anne as a faithful replicator. She hears Bahira's utterance and reproduces the same sound as faithfully as she can. In that sense she is like the blacksmith who is ignorant of horseshoes. Anne can be treated in the same way as we treated printing presses and CD players. The person who operates the machinery need not intend to produce words: they just need to know how to operate the machinery. Anne is the technical operator of her vocal apparatus, and Bahira knows how to get Anne to utter the relevant sound.

One objection to this (exactly paralleling the case of the Frisbee and the plate from chapter one) would be that Bahira's intentions don't matter: Anne hears a token of an Arabic word and faithfully reproduces it. Isn't that enough for saying the Arabic word? The response to this is just that the original tokens Anne hears are only tokens of Arabic words because of Bahira's intentions. Moreover, if Bahira teaches Anne using a recording of what she thinks is an Arabic word but is actually a different word from another language which happens to sound the same, Bahira still teaches Anne to say the Arabic word. Here we cannot appeal to the fact that Anne has heard an Arabic word and reproduced it. The word she

heard was not Arabic at all, but the one she utters is, and this is explicable only by appealing to Bahira's intentions to get Anne to produce an Arabic word.

4.4 Involuntary speech

Coprolalia is a phenomenon which involves involuntarily uttering certain words, often taboo words, or words which are likely to offend those who are present. (In popular imagination it is one of the key symptoms of Tourette's syndrome.) Another phenomenon which introduces similar problems (for our purposes) is that of somniloquy (sleep talking). In both cases it appears that word tokens may be produced unintentionally, thus undermining my claim that having the right intention is a necessary condition on uttering a token of a given word. Finally, Wetzel (2008:68) argues that intending to utter a word is not necessary for uttering that word, since a speaker might "intend to remain silent, but against her will utter 'cat' because a neurosurgeon is stimulating a nerve in her brain."

Anyone should be wary of speculating about the mental processes involved in cases like these. Wetzel's case, in particular, involves assumptions about the relations between neurophysiology and cognition, matters which are still poorly understood. How can we be sure that what the neurosurgeon does is not to create an intention in the subject to utter the word 'cat'? If, on the other hand, that is not how things go, how can we be sure that the subject really has uttered a token of 'cat'? If the neurosurgeon simply causes an electric charge to be fired through the subject's vocal apparatus – without the mediation of a distinctively mental event such as an intention – then how is the case any different to one in which I seem to say 'cat' by pure coincidence as a result of being punched in the stomach? Such occurrences are not genuine word tokens at all.

The case of talking in one's sleep is also delicate. Could such phenomena be explained in terms of intentions 'within a dream'? Such a possibility is hard to assess, but just as hard to rule out. I once shared a dormitory with a group of British friends and our newly found New Zealand companion. The latter became quite vocal once he was asleep, referring to the rest of us as a "bunch of greasy

cricket playing x's", where x was an expletive noun. Rightly or wrongly, we took this to be revelatory of a genuine state of mind.

What should we say about coprolalia? A complication for cases like these is that it is at least coherent to describe them as involving a genuine intention on the part of the speaker to utter the words in question. The sense in which these actions are unintended could be explicated in terms of conflicting intentions, sudden changes of mind, or perhaps even higher order intentions that a first-order intention be ineffective. As argued above (§2.1), intentions need not be fully conscious.

4.5 The causal efficacy of swamp words

Cappelen (1999) objects to the notion that it is necessary for some concrete item to be an instance of a word that its creator be in a certain mental state (a notion he calls the 'necessity thesis') as follows:

> Suppose I find the following token on the street:
>
> CAN YOU SPARE A QUARTER?
>
> I pick it up and use it (rather than my voice) to ask people whether they can spare a quarter. When I do this, I use a token of the English sentence "Can you spare a quarter?". Now, suppose I find out that the token was produced with the wrong intentions or maybe with no intentions at all (it might be the result of an accidental spilling of ink). A proponent of the necessity thesis would have to say both that the ink mark isn't a token of an English sentence and that I never used it to ask anyone to spare a quarter (because I can't ask a question in English without using tokens of English words). Both claims are preposterous. I have used a token of an English sentence, whatever its production history might be. (Cappelen, 1999:95)

Let's focus on the case where the found piece of paper was produced with no intentions at all as a result of some ink being spilled. Cappelen thinks that the proponent of the necessity thesis is committed to two absurd theses. The first is that the found ink marks are not instances of English words. The second is that Cappelen's fictional counterpart did not use the piece of paper to ask people to spare a quarter.

116

With respect to the first, I agree that the proponent of the necessity thesis is indeed committed to the view that the ink marks (at least while they're still lying undiscovered on the pavement) are not instances of English words. On the intentionalist view, they might look like English words but that just isn't enough. That was what the swamp word cases like the thought experiment about the eroded rock on a distant planet were supposed to show. If Cappelen wants to count the ink marks as instances of words, then he ought also to count swamp belches and eroded extra-terrestrial rocks. I suppose he probably would. We appear to disagree on this point, but the intentionalist view is not "preposterous."

What about the second claim that the proponent of the necessity thesis is supposedly committed to, the claim that Cappelen's counterpart did not use the piece of paper to ask people to spare a quarter? Cappelen considers the following possible reply on behalf of the necessity thesis:

> An entity is a token of a sign only if it is either produced intentionally or is used in the performance of a speech act. Call this the Modified Necessity Thesis (MNT)). (Cappelen, 1999:95)

This has an obvious ad-hoc feel about it. Thankfully, better solutions are available. I'll consider two. The first avoids MNT by maintaining that the found ink marks never become instances of English words. The second avoids MNT by characterising Cappelen's counterpart's actions as genuinely creating a new artefact.

Cappelen's reason for saying that his counterpart did not, according to the proponent of the necessity thesis, use the piece of paper to ask people to spare a quarter is that one cannot ask a question in English without using English words. But one could argue that it is possible to ask a question – even a question in English – without using English words. As long as the ink marks look the part and Cappelen's interlocutors believe that they are genuine tokens he could be

said to have asked a question in English without using concrete tokens of English words. If everyone *thinks* of the ink marks as tokens of English words, then they will have the right English question in mind, as it were. We can all agree that the ink marks on the found piece of paper look like English words. We are assuming that the ink has been spilled so felicitously that the resulting marks are good enough to dupe any competent user of English into believing that they were created by a careful and competent act of English handwriting. So what's to stop someone using them to do what they look like they're supposed to do? No doubt Cappelen believes that if something looks like a given word, then it is that word, but to make that point here is to beg the question against the view he is criticising. The intentionalist has a coherent way of describing the scenario, so this particular objection can be met.

An alternative response maintains the view that Cappelen did ask the question using real, concrete, English words by explaining how swamp words can become English words in virtue of a process of minimal creation. This second option is closer in spirit to Cappelen's MNT. It consists in the claim that picking up the piece of paper and forming the intention to use it in a speech act is a kind of artefact creation. There's nothing *ad-hoc* in this idea: cases of minimal creation are familiar from the literature on artefacts and have been amply illustrated above. I can pick up a rock and make a paperweight merely by forming a certain intention. Similarly, I can pick up a serendipitously ink-stained piece of paper and make it into an English inscription by forming the intention the use the object in a certain way.

4.6 Cappelen's nested intentions objection

Cappelen suggests that it should fall out of any broadly Gricean story of communication that speakers usually intend to utter the words they do. Suppose then, that a speaker intends to say 'cat'. If the word 'cat' is defined as, say, 'something which is intended to be a noun and refer to cats' then S's intention could be unpacked as 'intending to produce an entity which is intended to be a noun and to refer to cats.'

Cappelen seems to think there is some kind of pernicious regress here. He notes that intentions to act have eventually to bottom out in a description of the action to be performed in non-intentional terms. This is correct, but in the present case, that is exactly what appears to happen.[19] There would be a regress if 'cat' were defined as 'something intended to be an instance of 'cat'. Then we would have an intention to produce something which is intended to be an instance of 'cat'. But this kind of circularity was flagged and defused above. In the present case we appear to have nested first and second order intentions, but it's not clear why that should be problematic. Perhaps the objection should be that it is simply implausible that we have these nested intentions. It does seem a clumsy way of describing speaker intentions, but this kind of nesting is not impossible.

4.7 What public?

Philosophers often say that language is a public phenomenon. What does this mean? In the previous chapter I distinguished some of the senses in which an artefact can be said to be public in character. These were:

(i) The artefact belongs to a kind which is widespread.

(ii) The artefact kind is the work of many hands. In other words, there is no single inventor of that kind of artefact. Rather, it is the product of successive modifications over time.

(iii) People agree (even if the agreement is never expressed) about roughly what it takes to be a member of that kind. This is more or less the idea that the concept corresponding to the artefact kind is one which is shared.

(iv) The artefact is subject to agreed norms, conventions or laws.

---

[19] Of course, it's bottoming out in contentious stuff like meanings, but that seems okay. We don't need a naturalistic reduction of meaning. We're attending to the ways meanings are conferred on a substrate.

(v) The character of the artefact is determined relative to an activity which requires the participation of at least two people. For such artefacts, that they be successfully intended to be recognisable by at least two people becomes a constitutive requirement.

I think words are typically public artefacts in all these senses, though perhaps the most important sense is the last one. According to my account, the publicness of language consists in the existence of conventionalised ways of signalling clusters of linguistic intentions. Words hang around and have the properties they do because they have some useful function which causes speakers to intentionally re-use them. Stable public signs emerge because one of the functions of externalised language is to co-ordinate with other people. The upshot of this is that one can't be like Humpty Dumpty and insist that some utterance means whatever you want it to. Ways of signalling linguistic intentions are established not by any single individual, but through the activities of many users of a given linguistic artefact. This creates norms which individuals can infringe but not override.

According to my story, utterances count as tokens of a given word because it makes a certain set of linguistic intentions on the part of the speaker recognisable *to a certain audience*. But how to specify that audience? It can't be *all speakers of English* since an utterance in British English might not be recognisable to a native speaker from Texas. Moreover, utterances using a local slang might be explicitly intended to be recognisable by only a small subset of English speakers. Difficulties specifying the boundaries of a linguistic community within which a language is shared are an often remarked upon pitfall of theories of public language. Indeed, there is probably no way of zooming in on the community relative to which a word is intended to be recognisable, other than by stating that an utterance of a word is intended to be recognisable (in principle) to all and only the past and present users of that word. More generally, a public artefact is intended to be recognisable in normal conditions by people who are experienced in using that kind of artefact.

Echoing a discussion in chapter one (§4.3) one might point out that not all symbolic artefacts are intended to be recognisable by anyone other than the artisan. Why, then, can someone not have a private way of saying 'cat'? Suppose that someone intends **blarg** to be recognisable by themselves (and only themselves) as a noun referring to cats. Would that inscription then count as an inscription of 'cat'? One option here would be to deny that **blarg** is an inscription of any word at all. I reject this option. Though we won't find private words in the dictionary, dictionaries are not authoritative arbiters of word existence. Moreover, some words are introduced from scratch and then become widely used. Was the first apparent utterance of the word not a genuine utterance of the word? Are only later uses genuine tokens? Does the first utterance become a genuine token retroactively as use of the word takes hold? Better to allow that there can be private words. I even think that 'blarg' is an English word in the scenario described even though it is known by only one speaker of English. What other language could it be a word of, assuming the speaker is a monolingual speaker of English? Perhaps it could be a word of no language; against this, note that 'blarg' has English phonology and syntax. For example, the plural – let's say – is 'blargs'. The speaker is also likely to use the word in English sentences.

So **blarg** can be a token of a word meaning the same as 'cat', and this can even be a word of English, even if it is known only to one person. But in the scenario described it cannot be an utterance of 'cat' because whether something is an utterance of 'cat' depends the activities of the many users of that kind of artefact. According to convention, 'cat' is associated with a limited range of signalling devices (utterance forms, and inscription forms). It makes perfect sense to imagine private English words which are intended to be recognisable only to the speaker. For example, in writing this thesis I have created a number of personal abbreviations and pieces of jargon which I use in my notes. After all, communication between individuals is only one of the uses to which we put linguistic externalia. They can also be used for private rituals and games, or as a kind of scaffold for thought (e.g. repeating a phone number out loud in order to

memorise it, writing out a sum on a blackboard). Still, communication is *one* of the uses to which language is routinely put, and an important one. As a consequence, words typically reflect the exigencies of co-ordination between different speakers. In other words, ways of signalling linguistic intentions become widespread, and norms about the correct ways of using a word will emerge as part of communal linguistic practice.

4.8 Unrecognisable in context

Suppose we're talking about the river bank. Without signalling a change of topic, I say "I'm going to the bank," intending to communicate that I'm going to the (financial) bank. My utterance is not recognisable by the intended audience as (financial) 'bank'. Intuitively I have said (financial) 'bank'. We don't want to say that I actually said (river) 'bank'. Fortunately the view doesn't push us to say this. Though the token sounds like (river) 'bank', there was no intention to utter (river) 'bank', and therefore no such intention can be recognisable. But an uncomfortable result is looming: since my linguistic intentions are apparently unrecognisable, I may have failed to say either 'bank' word. I think this has to be resisted. Perhaps the thing to say is that my utterance *is* recognisable as an utterance of (financial) 'bank' despite the fact the few people would actually recognise it. For example, within relevance theory approaches in pragmatics (Sperber and Wilson, 1995) we could describe the case as follows: when I make my announcement, the audience will quickly settle on an interpretation which fits their expectation of relevance. Unfortunately, in the case at hand there is a readily available interpretation which makes perfect sense right up until the moment when I head off in the direction everyone was least expecting. That most people will be blocked from actually recognising my linguistic intentions, does not, perhaps, entail that my intentions were unrecognisable.

5. CONCLUSION

In this chapter I've suggested that the ECA view of words is supported by a fairly deep analogy between words and other kinds of essentially communicative (though not strictly linguistic) artefacts. The general theory which was developed

in chapter one fits the case of words very neatly. Similar problems arise and take similar solutions. The view explains intuitions about minimal creation, swamp words, and exaptation. It harmonises nicely with theorising in generative linguistics, helping to make sense of the fact that externalia are invested with linguistic significance. In response to worries that linguistic behaviour is too spontaneous to count as genuinely intentional, I've described a slightly deflated notion of intention which is rich enough to do service in a theory of artefact production, and which seems applicable to typical cases of language use. Finally, I've put a bit of a dampener on the idea that we can read off a person's word-regarding intentions from a generative linguistic theory of competence, but I've emphasised that many aspects of our linguistic behaviour *are* intentional, and that we have at least some ordinary conscious knowledge regarding the phonological, syntactic and semantic properties of words. I've also explained how questions of word individuation are settled on the ECA account and I have responded to a number of objections and potential counterexamples.

CHAPTER 3: THE FORMS OF WORDS

This chapter reprises a question broached in the previous one: what makes two utterances utterances of the same word? One answer to this question is what Kaplan (1990) called the *form-theoretic* account, FT. This approach seeks to explain sameness of word in terms of physical resemblance. It is something of a default view, sometimes tacitly assumed, sometimes explicitly defended. Adherents include Stebbing (1935), Hardie (1936), Lewis (1983:163), Davidson (1984:90), Devitt and Sterelny (1999), Cappelen (1999), Alward (2005), Predelli (2010), and, arguably, Searle (1995:83).

In §1 I present Cappelen's (1999) version of FT, and show how it is implemented within Searle's (1995) framework for social ontology, which involves his well-known XYC schema. On this view, words depend on conventions which are maintained within linguistic communities. These conventions involve collective acceptance that things of type X (things with a certain form) count as things of type Y (having linguistic properties such as being a noun, referring to cats, etc.).

In §2 I argue that FT's appeal lies in a tempting but misleading conception of the processes involved in word recognition. The intuitive motivation for FT starts from the observation that competent language users are very good at working out what word types are tokened in the speech and writing to which they are exposed, and concludes that each word in the language must be associated with a distinctive acoustic (and orthographical) profile.

In §3 I make the case against FT and respond to the argument based on word recognition. As well as the usual empirical arguments against FT (such as formal variation between tokens of a given word, and formal coincidence between different words), I explain the error behind the word recognition argument for FT. Its mistake is to assume that word type recognition is entirely a matter of observing the form of utterances and inscriptions. Against this 'bottom-up'

conception of speech perception – according to which word type must be detected on the basis of formal features of the linguistic signal before knowledge of syntax and semantics can be consulted – I argue that the forms of linguistic signals are highly unruly and not reliable indicators of word type; language users' success in word type recognition relies additionally on such things as their grammatical competence, background knowledge, and guesswork. Word recognition is therefore not a discrete first step towards understanding what is said in a context, but is accomplished simultaneously with semantic interpretation. Thus, our epistemological success in word type recognition does not require that all recognisable tokens of a word instantiate a specific form associated with that word. In addition, I raise a problem for the notion of form as it is usually employed in stating FT.

In §4 I argue that the demise of FT throws up a problem for anyone hoping to use Searle's XYC schema to give a social ontological theory of words. Briefly, the XYC schema requires two ways of specifying the membership of a social kind. Cappelen assumes that the X slot can be filled in with a form-theoretic description of a certain kind of utterance, while the Y slot is filled in with syntactic and semantic properties. If, as I argue, no appropriate form-theoretic description is available, the XYC schema begins to look poorly suited to explaining the social construction of words. (Establishing this firmly would require ruling out other ways of filling in the X slot. In this chapter I gesture to what these other ways might be, and express scepticism about their prospects.)

Finally, though having a certain form is neither necessary nor sufficient for being a token of a given word, a theory of words ought to say *something* about the formal features of utterances and inscriptions: Kaplan (1990) was wrong to think that form was *irrelevant* to word type. In §5 I explain the role played by form in the ECA theory of words: I argue that the forms of utterances are localised ways of signalling word-type without being in any sense constitutive of it.

1. THE SEARLE-CAPPELEN VIEW OF WORDS

For Cappelen, words can be defined in terms of the formal properties of utterances and inscriptions. So, for example, there is an acoustic form and an orthographical form associated with the word 'cat'. All and only the things which instantiate either of those forms count as an utterance of 'cat'. Of course, it is not a brute fact about the world that those things are tokens of 'cat'. According to Cappelen, what makes it the case that a person producing a certain kind of utterance is producing an instance of 'cat' is that they belong to a community in which certain linguistic conventions are maintained. He writes:

> It is a matter of convention that what you look at right now are tokens of words. They would not be word tokens if it hadn't been for the presence of certain kinds of conventions. Such conventions are upheld by intentional linguistic activity. So, there can be word tokens only if there is intentional linguistic activity. It's a mistake to infer intentionalism from this. The conventions we have are, very roughly, of two kinds:
>
> (C1) Entities with such-and-such properties count as tokens of the same sign.
>
> (C2) Entities with properties P (where entities with P count as tokens of the same sign according to some C1-type convention) count as tokens of the same sign as tokens with properties P´ (where P´ entities count as tokens of the same sign according to some C1-type convention.) (Cappelen, 1999:99)

The role of C1 conventions is to determine classes of utterances or classes of inscriptions, respectively. According to Cappelen, the 'such-and-such properties' locution can be filled out with terms which refer only to formal properties of tokens. C2 conventions relate form-theoretically determined classes of utterances with form-theoretically determined classes of inscriptions. Further C2 conventions could be introduced for Braille, Semaphore, etc. C2 conventions can serve not only to bridge different media of communication, but also to bind differing physical types within a single medium (something Cappelen doesn't point out): after all, we can't have a single C1 convention which binds together **ɑ**, **A**, and **a**. If there is no single physical description that describes all three of

these shapes at the right level of detail we'll need a separate C1 convention for each of them, and a C2 convention to bind them together.

Cappelen doesn't say much about what he means by 'convention', nor what is involved in being tokens of the same sign beyond having a shared form. We can fill in the gaps with a little imagination, and by drawing on Searle's approach in social ontology, (an approach within which Cappelen's suggestion is clearly based). Searle (1995) proposes to explain the existence of social and institutional facts in terms of features which are conferred on existing objects via collective acceptance of constitutive rules. The form of these rules is:

$$X \text{ counts as } Y \text{ (in context C)}$$

So for example, if we are in a community that collectively accepts that certain pieces of metal have a monetary value of one dollar, then those pieces of metal count as one dollar. There are a few infelicitous features of this schema. For one thing, the parenthesis is dispensable, since any contextual specifications can just as easily be inserted under the X slot. Thomasson (2003a:274) and Epstein (2013) criticise various other aspects of the XYC schema, and Epstein offers the following reformulation:

> For all objects z, if z is X then there is an object u such that z constitutes u and u is Y. (Epstein, 2013:6)

Epstein's reformulation of the XYC schema, and his associated distinction between *grounding* and *anchoring* are helpful sharpenings of Searle's view, but we don't really need to go into these details here. The problem I'm going to outline afflicts the view whichever way it is formulated. I'll continue to refer to the view informally as the XYC schema.

What should we make of the notion of collective acceptance? Searle (1995:23-26) insists that collective acceptance involves a kind of collective intentionality which is not reducible to individuals' first-person beliefs, desires and intentions.

His account of this is enigmatic: it has to do with individuals having beliefs and intentions which can be characterised as first-person plurals. So collective acceptance of a rule would involve a plurality of individuals whose acceptance of the rule consists in an attitude towards the proposition *we accept that X counts as Y in C.* As Searle himself notes, this is no more than a trivial notational manoeuvre, not a rich account of what collective intentionality is supposed to involve. Thomasson (2009:546-547, 2014:55-56) leaves open whether public artefacts require an account of a distinctive kind of intentionality, a kind of *collective* intentionality. I propose to do the same.

Applying the Searleian schema to the case at hand, Cappelen's view is that the existence of words depends on the maintenance of conventions according to which things of *acoustic type X* count as Y. Cappelen doesn't indicate what kinds of Y features are relevant to words, but they presumably include things like being a noun, referring to cats, etc. On this view, it would be misleading to say that words are just kinds of sounds and shapes in just the way it would be misleading to say that money is just paper. The kind of social construction which the XYC schema models should be understood as an ampliative process. We start off with some things which are X and we add some Y properties by convention. In the case of money, we start off with pieces of paper with a certain form and provenance, and we impose on such pieces of paper some extra properties to do with their monetary value. In the words case, we start off with sounds and shapes with a certain pre-agreed form, and we impose upon them a set of linguistic properties. Thus, it would not be quite right to say that words are just sounds and shapes, on Cappelen's view, although things which have the right shape or sound do *constitute* word tokens.

On such a view, the forms of words play an epistemological role: the form of a token is what enables it to be recognisable as having its various syntactic and semantic properties. But word forms also play a constitutive role: having a certain acoustic form is constitutive of being a token of a given word. How do these two roles fit together? In some instances of the XYC schema, the X properties make

a direct causal contribution to an object's ability to perform a certain Y function. Taking a non-linguistic example, suppose we have a convention which establishes that a certain kind of knife is a fish knife. The items singled out under the X description have a certain form and structure which makes them effective tools for eating fish. Those very same features also serve to indicate that the knife is *supposed* to be used for fish.

Sometimes, however, the properties picked out under the X term have a *purely* communicative role. They serve merely to indicate what Y properties have been agreed for the things which have the X properties. For example, there is nothing about the form or structure of traffic lights which makes them good at stopping traffic. Traffic lights are communicative artefacts whose appearance serves to indicate that cars are supposed to stop. In the case of purely communicative artefacts, having the form which *signals* a certain Y property is necessary and sufficient for *having* that Y property.[20] Applying this to the case of words, Cappelen's view is that having the form which signals 'cat'-ness is necessary and sufficient for being an instance of 'cat' (by convention). We are, at this point, very close to the underlying motivation behind FT. Articulating that motivation is the purpose of the next section.

2. THE RECOGNISABILITY ARGUMENT FOR FT

The appeal of FT lies in a tempting story about linguistic communication. The story can be told as follows. Suppose Anne wants Brian to know that Carol is hungry. Humans are not telepathic so Anne has to perform some action which is observable by Brian: she says the words, 'Carol is hungry'. This works because Anne knows that Brian knows that Anne knows that those words express the proposition that Carol is hungry.

On this story, word tokens are visible or audible aspects of the environment. Working out which word types have been produced is a matter of looking and

---

[20] Clearly, this part of the Searle-Cappelen account of words is tracking the same phenomenon as my notion of ECAs. But the characterisation of the phenomenon is not equivalent, as I explain in §§4-5.

listening. Interpretation is thought of as a *bottom-up* process: the first thing the hearer has to do is work out what words have been uttered; then and only then can they use their knowledge of language (and background knowledge) to work out what proposition is being expressed in the context. This bottom-up picture of interpretation is at the heart of the appeal of FT, I submit.

Someone tempted by this story might propose the following argument:

> (P1) All word tokens produced in a language L are recognisable as tokens of a single word type by competent speakers of L.

> (P2) If (P1) then each word type in L is associated with a description in terms of detectable, formal properties which all and only its tokens satisfy.

> (C) Each word type in L is associated with a description in terms of detectable, formal properties which all and only its tokens satisfy.

Call this the 'Recognisability Argument' (RA). (C) involves both the idea that having a certain form is necessary for being an instance of a given word ('the necessity thesis'), and the idea that form is sufficient for being an instance of a given word ('the sufficiency thesis').

(P2) is initially plausible. Suppose Anne sorts 100 marbles into ten sets of ten then mixes them up. Brian is asked to sort them into the same ten sets, and succeeds. This seems to require that Anne and Brian are categorising the marbles on the basis of detectable properties of individual marbles. Suppose, now, that Anne makes a recording of ten utterances of each of ten different word types. Brian has to recover the ten sets corresponding to word type. He's likely to do well. How is this possible? When he considers each token utterance there have to be detectable, formal features which reliably indicate that that utterance is of a certain word type.

130

Brian's task in the word-guessing game is unlike real instances of linguistic communication in ways which turn out to be important. I'll clarify and tweak (P2) in the discussion of RA in §3.4, before eventually rejecting it altogether. For now I grant it for the sake of argument.

(P1) might sound far too strong, but note that (P1) does not require that competent speakers never make mistakes: tokens don't have to be *recognised* as tokens of a given type; they just need to be *recognisable*. Someone might classify an utterance wrongly due to tiredness, but this doesn't mean the utterance isn't recognisable. Nor does (P1) require that any competent speaker be able to recognise every token produced in their language. The phrase 'by competent speakers' could be taken to mean (i) by a certain proportion of competent speakers, (ii) by a certain proportion of competent speakers who are linguistically similar to the producer of the token, or (iii) something else.

Even with these hedges, (P1) might still sound too strong. Some word tokens might be completely unrecognisable by anyone, due to sloppy handwriting or slurred speech. If so, then even if all and only recognisable tokens satisfy a certain formal description, we cannot infer that all tokens of a given word type do so.

The defender of RA could insist that there just aren't any tokens of a given word which are not recognisable as such by (at least some) competent speakers of the language; being recognisable as a token of a given word type is *essential* to an utterance's status as a token of that word type. This is not obviously true: many things are reliably identifiable without being essentially so. Suppose tokens of the bird species, oystercatcher, are recognisable by the fact that all and only oystercatchers have a long orange beak, mostly black feathers and a white stripe along the wing. But being recognisable is not *essential* to being an oystercatcher: a featherless, beakless, unrecognisable oystercatcher is still an oystercatcher.

Nevertheless, on the Searle-Cappelen view, words just are conventionally agreed detectable signs. The form of a word signals that it has certain Y properties precisely because that is the form that has been stipulated as *having* the Y properties. The features of words which indicate their Y properties are precisely those which ground its having the Y properties. If something is not recognisable as a Y (by someone experienced with Ys), that can only be because the X properties which signal the Y properties are not present. But if they are not present, then the item is not a Y. If this is right then an utterance which is unrecognisable by anyone would fail to be an utterance of the intended word type. If it doesn't have the acoustic profile which makes it recognisable as an instance of 'cat', then it isn't an instance of 'cat'. So, anyway, I imagine the form-theorist arguing.

Suppose (P1) is accepted. The objection might now be that even if word tokens are essentially recognisable as tokens of a given word type, and even if all tokens are recognisable in virtue of their detectable, formal properties, that doesn't show that words can be defined in terms of those formal properties. However, this objection is unfair. Given the theoretical concerns reflected in RA, the properties which drive word type recognition are exactly the ones we're interested in; that is the way of carving linguistic reality which answers to the context in which RA is offered. There are no doubt other ways of classifying utterances, but this way of doing it reflects an interest in certain phonetic kinds, recognition of which is hypothesised to be the take-off point for interpretation.

If my rejection of RA is to seriously undermine FT, we ought to have some assurance that RA is a key motivation for FT. My version of RA is intended as a worked-out version of an argument which appears in nascent form in the literature. Devitt and Sterelny write:

> Tokens are dateable, placeable parts of the physical world… The obvious examples of word tokens are inscriptions on a page or sounds in the air… Inscription types and sound types are

identifiable by their overt characteristics and so we might call them "physical types". (Devitt and Sterelny, 1999:71, emphasis added)

An argument in the spirit of RA was made by Cappelen (1999), who objected to views according to which a speaker's intentions could be essential to the word type of an ensuing utterance:

> [A] necessary condition for being in a position to interpret u [an utterance of 'Alice is asleep'] is knowledge of semantic facts such as (s):
>
> (s) "Alice" refers to Alice
>
> To use (s) in interpreting u, it must be possible to identify the first token of u as a token of "Alice". How is that done? …[Y]ou look at it. If you recognise its features as being those of tokens of "Alice", you make use of (s). This is how interpretation gets off the ground… Had intentionalism been true, looking at the first token of u would not enable you to determine whether it is a token of "Alice" and hence wouldn't tell you whether to employ (s) in your interpretation… Only someone with knowledge of the intentions and history of the utterer of u is in a position to tell which word the first token of u is a token of. (Cappelen, 1999:97)

Cappelen's point seems to be that there can't be any essential properties of words which are undetectable, for if there were, one could never be certain which word one was faced with because one cannot exclude the possibility that there is an essential property which the utterance lacks. In other words, having a certain form must be sufficient for being an instance of a given word. Alward (2005:180) echoes Cappelen's argument.[21] This argument is weak. After all, it is possible to identify an object as a member of a given kind without observing that all necessary conditions associated with the kind are met. Having a certain genetic profile could be a necessary condition on being an oystercatcher, but you can still identify them reliably without checking for this. In a further argument, Cappelen writes:

---

[21] "Our ability to understand the utterances/inscriptions of speakers is underwritten by our knowledge of the meanings of the words contained therein. And our ability to utilize this knowledge depends on our further ability to recognize occurrences of the words in question. But [FT's alleged falsity] robs us of the ability to do so on the basis of observed features of utterances/inscriptions: the fact that an utterance/inscription looks or sounds like an occurrence of a given word, or fails to do so, does not decide the issue one way or the other" (Alward, 2005:180).

> The identity conditions for sign tokens were developed by us for a reason. We should expect the distinction between things that are sign tokens and things that are not, to be a distinction that says something about how those entities can be used (something about the function they can perform for us). These two ink marks:
>
> **Can you spare a quarter?**
> **Can you spare a quarter?**
>
> are, for all linguistic purposes, functionally equivalent. Whatever you can do linguistically with the first, you can do with the second. (Cappelen, 1999:95-96)

This fact about the two ink-marks holds true whether or not they were produced with the same intentions in mind, or even if they are swamp words. Here is my take on Cappelen's argument. Two tokens are tokens of the same word type if and only if they are functionally equivalent. Two tokens are functionally equivalent if and only if they are intersubstitutable without disrupting communication. Cappelen's idea is that linguistic categorisation should only be sensitive to the causal roles which utterances and inscriptions are able to play in virtue of their physical makeup. What matters are the detectable, formal features of utterances and inscriptions which language users are able to recognise as belonging to a conventional linguistic category. Since creator intentions are not detectable aspects of physical signals, they have no impact on the causal roles a token is able to play. Thus, three recent defences of FT draw on an argument like RA.

3. AGAINST FT

In this section I first consider three well known objections to FT. Each attempts to provide counterexamples to either the necessity or the sufficiency claim embodied in (C), above. The first can be easily responded to, while the others are more problematic. I go on to discuss the flaws in the word recognition argument just presented before presenting a further objection to FT, one which questions the form-theorist's focus on the intrinsic formal properties of word tokens.

3.1 Multiple media

Words can be written as well as spoken, but acoustic and inscriptional instances of a word do not share a common form. Call this the *Multiple Media Argument*. I will show that the problem is not as insurmountable as Hawthorne and Lepore (2011) and Kaplan (2011) have concluded. Stebbing considers the problem:

> Spoken words are sounds; written words are shapes. The same word may be spoken or written… Sounds and shapes are quite unlike; the sound is not identical with the shape, nor have they any kind of similarity. (Stebbing, 1935:3)

Her own solution is to deny that inscriptions genuinely instantiate words. Instead, as discussed in the introduction, she suggests that inscriptions are mere representations of spoken words. If that is the case, then the fact that 'cat' inscriptions do not resemble utterances of 'cat' is beside the point. However, given that it is part of common sense that words can be tokened in writing, and that philosophers typically share this assumption, defenders of FT may wish to find a different response than Stebbing's, especially in light of the fact that there is a perfectly good one available.

On Cappelen's account, there is an easy answer to the Multiple Media argument: his C1 and C2 conventions are tailor made to resolve the problem. As noted above, we can have a C1 convention which determines that certain kinds of sounds are tokens of a word, a further C1 condition to determine that certain kinds of marks are tokens of a word, and then a C2 convention which determines that all these sounds and shapes are tokens of the same word. In fact, the nature of this response emerges much more perspicuously when one frames it in terms of the XYC schema. We simply note that for any given Y property, we can perfectly well have more than one convention. That is, we can have two conventions, thus (where: X_1 refers to a kind of sound and X_2 to a kind of shape):

$$X_1 \text{ counts as } Y$$

Moreover, we do not need to buy into Cappelen's conventionalist account in order to exploit his essential insight. We can remain neutral on the question of what *makes it the case* that certain inscriptions and certain utterances are tokens of the same word and simply modify (P2) of RA as follows (changing (C) accordingly):

> (P2´) If (P1) then each word type in L is associated with one or more descriptions in terms of detectable, formal properties such that all and only things satisfying exactly one of those descriptions is a token of the type.

3.2 Diversity

Kaplan's (1990) principal complaint against FT was that a single word can be pronounced or written in different ways, thus undermining FT's necessity thesis. Millikan (2003), Wetzel (2008), and Hawthorne and Lepore (2011) concur. Call this the *Diversity Argument*. Some kinds of variation can be accommodated: just as we can appeal to separate conventions for the written form and the spoken form of a word, we can also have separate conventions for the different pronunciations or different spellings of the word. Similarly, that 'r' takes different forms in upper and lower case is secured by having multiple conventions (of the form X$_1$ counts as Y, X$_2$ counts as Y, etc.). Unfortunately, there is simply too much variation for this approach to work generally. Multiplying conventions to deal with this variation would amount to imposing an implausibly large cognitive burden on speakers. In other words, though some variation can be plausibly chalked up to the existence of multiple conventions, there remains a large amount of *unconventionalised* variation.

It bears emphasis from the outset that much of the variation between productions of a given word is irrelevant to word-type recognition. It doesn't

matter what colour an inscription is, or what it's made of. To the form-theorist, what matters is the shape. Moreover, we don't care how deep the ink seeps into the paper or whether the grooves carved into gravestones are triangular or square in cross-section. It's the two-dimensional shape the object presents from a certain position which counts. On the speech side, it doesn't usually matter whether an utterance is loud or quiet, fast or slow, high or low pitched, etc. Soprano singers know there are limits to this: by the lights of FT, some intended utterances of words are too high pitched to replicate the formant pattern associated with certain vowels. Nevertheless, very different waveforms could instantiate a shared formal pattern described in terms of formant frequencies and stops, say. However, the idea that there is an acoustic commonality between all the utterances of a given word faces some significant empirical challenges.

For one thing, there is variation across linguistic communities. Chaucer, Hume and Harper Lee can be said to have used an overlapping set of English words, but they pronounced them in very different accents. Similarly, **colour** and **color** are inscriptions of the same word despite being constituted by different letter sequences. Wetzel (2008:5) notes that 'colour' has eighteen known additional spellings: "collor, collour, coloure, colowr, colowre, colur, colure, cooler, couler, coullor, coullour, coolore, coulor, coulore, coulour, culler, cullor, cullour."

However, the existence of inter-community variation is not decisive. While there is a familiar sense of 'word' according to which we can describe a single word as occurring in modern English, Old English and even Proto-Indo-European despite formal variation beyond the point of mutual recognisability, if we press this objection we are saddling words with an explanatory responsibility which the form-theorist never promised to fulfil. FT may be best construed as marking out one important theoretical conception of words among a plurality of differing conceptions. Though it may be part of historical linguistics and even common sense (at least sometimes), to think of words as individuated by their histories, the form-theorist is not doing conceptual analysis or historical linguistics. The form-theorist is just interested in the formal types which putatively play a certain

role in communication within a language community. An etymologist may say that some utterance made by Chaucer is of the same word type as some modern utterance. The form-theorist need not deny this, but if Chaucer's utterance is unrecognisable by members of our community, they can legitimately deny that it is an utterance of one of *our* words – at least not in the form-theorist's sense of 'word' which pertains only to recognisable form-theoretic units. The form-theorist is interested in finding the word types which RA requires, not what some other explanatory project requires, and if it tramples on folksy intuitions about etymology then so be it. They never set out to vindicate folksy intuitions either.

More problematic is the fact that variation exists within linguistic communities, however locally construed, not to mention within a single individual over time. For example, sloppy handwriting can lead to unconventionalised diversity in the forms of letter tokens. Also, fluent speech is rapid and full of short-cuts: Wetzel (2008) notes that 'extraordinary' can be pronounced with six, five, four, three or even two syllables (/strornry/). A word can be spoken at a high or low pitch, lisped, enunciated theatrically, sung, whispered, etc. Such factors ensure that two utterances of a word may vary wildly from an acoustic point of view.

The form-theorist need not seek formal similarities between word instances directly; they can treat word instances as sequences of segments (where the smaller segments are defined in terms of their form). We can call the finer-grained segments of written words *graphemes* (or 'letters') and those of spoken words *phonemes*. For present purposes, phonemes are to be thought of as kinds of acoustic chunks instances of which can be assembled into sequences to make words. (This use of terminology is not universal: many linguists use talk of phonemes as a way of describing internal aspects of an individual's mind/brain at a certain level of abstraction. There is no expectation that these should be mappable to acoustic properties.) Appealing to phonemic segments may filter out some of the noise, but in the end the acoustic signal is just too messy to be carved into chunks. A single phoneme can be pronounced differently by a single person on different occasions. Bromberger (2011:492) gives the example of 'in'

which is pronounced [in] before 'New York', [im] before 'Boston', and [ing] before 'Cambridge'. Jackendoff (1990:58) notes that the vowel sounds in utterances of 'tap' and 'back' differ from an acoustic point of view since they are modified by a speaker's articulation of adjacent consonants. Rey (2006a:247) mentions that speakers of American English hear utterances of 'rider' and 'writer' as containing a /d/ and a /t/ respectively, despite being identical from an articulatory or acoustic point of view. (The difference in perception is related to the lengths of the preceding vowels.) Such effects undermine the idea that an acoustic blast can be considered as instantiating a sequence of acoustically defined phonemes.

This point is widely recognised in linguistics (Chomsky and Halle, 1968, Fodor, Bever and Garrett, 1974, Laver, 1994, and Liberman 1996). Sapir's pioneering essay "The Psychological Reality of Phonemes," compares phonemes to an artefact such as a club:

> [N]o entity in human experience can be adequately defined as the mechanical sum or product of its physical properties… To say that a given phoneme is not sufficiently defined in articulatory or acoustic terms…is, at bottom, no more mysterious than to say that a club is not defined for us when it is said to be made of wood and to have such and such a shape and such and such dimensions. We must understand why a roughly similar object, not so different to the eye, is no club at all, and why a third object, of very different colour and much longer and heavier than the first, is for all that very much of a club. (Sapir, 1933/1985:65)

It goes without saying that I find the analogy to be entirely apposite. The fact that clubs cannot be defined in terms of physical features is something we are well-placed to understand in light of the theory of artefacts presented in chapter one.

3.3 Formal coincidence

Where the previous point was addressed to the fact that having a certain acoustic form is not necessary for being an instance of a given word, the present sub-

section concerns the fact that having a certain form is not sufficient for being an instance of a given word either. Call this the *Formal Coincidence Argument*.

It is frequently observed that English contains at least two 'bank' words (one of which refers to a feature of river geography, the other to a financial institution). FT's sufficiency thesis provides no grounds for any such distinction. There would just be one radically polysemous 'bank' word. In addition, an utterance of 'at all' may contain a string of sounds which duplicates the sound of an utterance of 'tall.' We surely do not want to say that the word 'tall' has been uttered, but FT appears to require that we say exactly that.

A further uncomfortable consequence is that the form-theorist may be obliged to say that a foreign language word which sounds just like 'bank' (but which has a different meaning and an unrelated etymology, let's say) is the very same word as our 'bank' word. Taking a real world example, there is a Swedish word which is written like the English word 'god', but which is etymologically unconnected to the English word 'god' and which means *good*. Suppose it is pronounced just the same as the English word 'god'. Since utterances of the Swedish word are intersubstitutable with utterances of the English word without disrupting communication, it looks like Cappelen is committed to there being a single word here, which just happens to be used independently with different meanings in two different languages. Cappelen could perhaps try to say maintain that there is a Swedish 'god' word as well as an English 'god' word, even though they share all the same tokens. When an English speaker talks about god, they produce an utterance which is a token both of the English word 'god' and the Swedish word 'god'. This won't do, however, because Cappelen's notion that words are defined in terms of brute causal potential means that he has no leeway for a distinction between the Swedish and English words. To make that distinction, he'd have to appeal to something other than the brute causal powers which the tokens have in virtue of their intrinsic physical make-up.

Finally, these kinds of problems apply not just at the level of the word but also at the level of the sentence. For example, an utterance of 'visiting relatives can be tiresome' can realise two different sentences.

At least some of these problems can be resolved within the XYC approach to words by importing a little bit of intentionalism. After all, we can put whatever we like in the X descriptions of our XYC conventions. We can very well specify that it is (all and) only things produced intentionally by a speaker of English (and which have a certain form) which count as Y. Cappelen (1999) doesn't like this option, but it is a coherent position which avoids some of the problems afflicting Cappelen's account. In any case, the present proposal still leaves the problem of the two 'bank' words.

I think Cappelen just wants to bite the bullet here. He could modify his account, insisting only that having a certain form is necessary but not sufficient for being an instance of a given word. Unfortunately, this would at least partially undermine the form-theorist's appeal to RA: if form is not sufficient for instantiating a given word, then observing an utterance with a given form cannot get the interpreter all the way to knowing what word has been uttered.

3.4 The problem with the recognisability argument

In this sub-section I'm going to show what's wrong with the word recognition argument for FT, thus further undermining the view. Recall that the argument says our success in word recognition requires that each word type be associated with a distinctive acoustic profile. How else could we succeed? Updating the argument in light of the discussions above, RA is as follows:

> (P1) All word tokens produced in a language L are recognisable as tokens of a single word type by competent speakers of L.

> (P2′) If (P1) then each word type in L is associated with one or more descriptions in terms of detectable, formal properties such that all and only things satisfying exactly one of those descriptions is a token of the type.

(C) Each word type in L is associated with one or more descriptions in terms of detectable, formal properties such that all and only things satisfying exactly one of those descriptions is a token of the type.

If the argument is good the form-theorist has something to say in response to the Chomskian challenge. As noted above, (P1) is not obviously true, but I've argued that it is defensible in the light of Cappelen's background assumptions. My rejection of (C) will be based on a rejection of (P2´).

In §2 I tried to explain the appeal of (P2´) by using an analogy with Brian's task of recovering Anne's ten sets of ten marbles, but the analogy is misleading. As we saw in §4, the linguistic signals which interpreters have to attend to are highly variable and unruly. Speech, for example, is often rapid and slightly garbled (e.g. 'strornry' instead of 'extraordinary'). Spoken utterances exist only fleetingly and words run into each other without pauses. Were we to excerpt individual utterances of words from a recording of a fluent passage of speech and play them to a test subject there is no guarantee that the subject would achieve a high level of success in recognising the words. In normal circumstances, word utterances appear in a context: they're uttered by a particular person in a particular physical and social environment, and they usually occur not in isolation but in association with other words in sentences and conversations. This kind of context provides additional clues to the word type of a particular utterance, in addition to any formal clues.

How can you tell a word by the company it keeps? One way is as follows. Linguistic competence consists partly in an internal computational capacity which constrains the range of acceptable sentences in the language. Suppose someone says 'I think she's studying' but a loud noise obscures the first sound of the word 'she'. What people hear is ambiguous between 'I think she's studying' and 'I think cheese studying,' but syntactic competence alone rules out the latter interpretation.

In addition, a whole host of pragmatic factors and background knowledge can be brought to bear. Indeed, there don't appear to be any constraints on the kinds of information that can feed into the process of word-type recognition. In deciding whether one's interlocutor has said 'I've bought some chips' or 'I've bought some ships,' one does not have to rely purely on acoustic information, which may be ambiguous. Knowledge about the speaker and context will usually make one or the other interpretation vastly more probable.

The processes underlying word recognition are still a subject of controversy in psycholinguistics, but many theoretical approaches to this issue allow that such things as background knowledge and syntactic competence play a significant role in performing the mapping from acoustic blasts to mental representations of words. For example, Marslen-Wilson (1987) presents data suggesting that words are often recognised before the whole word has been heard. For example, an utterance of the word 'trespass' can be identified as soon as the /p/ is heard because the sequence /tresp/ can be completed in only one way in English. The sequence /tres/ is compatible either with an utterance of 'trespass' or of 'trestle', so hearing just that initial sequence is not sufficient for word recognition. We can thus talk of the *recognition point* for a word. This is the point at which the acoustic cues are sufficient to determine a single possible word type. However, Marslen-Wilson also argues that there is evidence of a phenomenon he calls *early selection*, in which words heard in context are identified before the recognition point. This suggests that people routinely draw on extra-acoustic information in word recognition.

Other studies indicate what kinds of extra-acoustic information are employed. The McGurk effect illustrates that even language users with good hearing lip read. Warren and Warren (1970) discuss the phenomenon of *phonemic restoration*. Subjects are played recordings of words in which a single phoneme is replaced by a cough. In the context of sentences talking about shoes, [*eel] (where '*' signifies a coughing sound) is heard as 'heel' while in sentences about cars it is heard as 'wheel'. This suggests that knowledge of the meanings of words heard

143

earlier in a sentence plays a part in word recognition. Finally, Miller and Isard (1962) presented subjects with strings of spoken words against a background of white noise. Under these sub-optimal acoustic conditions, they found that subjects were better able to recognise the words if they constituted a grammatical sentence (such as 'accidents kill motorists on highways') than a grammatically correct but anomalous sentence (such as 'accidents carry honey between the house'). The subjects performed even worse when what they were hearing was just a random string of words.

These kinds of phenomena suggest that hearers are able to exploit their background knowledge, syntactic competence and even visual cues to make word recognition judgements which are not licensed by the purely acoustic information available. If this is right, then (P2′) is false: recognisability as an utterance of a given word does not require recognisability solely on the basis of formal features of individual tokens. So even if (P1) is true, and it is essential to a word utterance that it be recognisable, this does not require that all and only the tokens of a given word share a distinctive form-theoretic profile.

It is worth reflecting on a system of communication developed for use by deaf-blind individuals, known as Tadoma. The Tadomist places their hand on the face of the speaker in order to feel the movements of the lips, the jaw, and the vibrations of the vocal cords. The linguistic signal to which the Tadomist has access is presumably severely impoverished compared to that which is available to language users with good hearing. They receive some formal clues – a lip-rounding here, a temporary ceasing of vocal cord activity there, etc. – but on the assumption that these clues do not suffice for word recognition, the Tadomist must have recourse to other kinds of clues, including background knowledge, etc.

The form-theorist might query the significance of this, arguing that however Tadomists identify word types, language users who have good hearing rely exclusively on formal features of utterances; the Tadomists have to identify word

types using a non-standard method but this is just because they do not have access to the rich formal clues in the linguistic signal which hearing people have access to. However, this response gets the dialectic wrong. What the existence of Tadoma shows is that it is possible to achieve a degree of success in word type identification even when the linguistic signal is severely impoverished from a formal point of view. The Tadomist faces a more difficult task than hearing people, but the difference is probably one of degree. In the words of Rey (2006a:247), "we're all, as it were, acoustic Tadomists."

3.5 What notion of form?

I want to make one further contribution to the case against FT. It concerns the notion of form which plays a role in the theory. How should the form-theorist characterise the forms of words? FT is commonly understood as focussing on *intrinsic* form of utterances, but this is not the best way of developing the view. RA requires word tokens to be recognisable on the basis of their formal, detectable properties. But there are intrinsic properties of tokens which are not easily detectable and non-intrinsic properties which are: colour is a detectable property of inscriptions which may not be intrinsic, while being formed of a certain kind of ink is an intrinsic property which may not be detectable. Of course, colour and ink type have little to do with word recognition. But if intrinsic form and detectable, formal properties can come apart, we should think carefully about whether intrinsic form is really the notion FT needs.

I suggested above that what matters to an inscription is something like two-dimensional shape. Now, suppose that concrete objects can instantiate two-dimensional shapes intrinsically. The fact that an observer can only perceive that an object has a given two-dimensional shape by getting themselves into the right position does not entail that the object fails to instantiate that two-dimensional shape intrinsically. The trouble is that many inscriptions get to count as genuine word inscriptions (by the lights of RA) on the basis of a two-dimensional shape which they *only appear* to have relative to some observer. This suggests that what

should matter to the form-theorist is not the intrinsic two-dimensional form of tokens, but something more observer-relative.

To see this, consider *Utopia*, an artistic installation by Georges Rousse (figure 1). The work consists in a painted room in which it appears to a correctly positioned subject that the word 'UTOPIA' is floating in mid-air. What is painted on the walls might be a genuine instance of 'utopia', but if that is so in virtue of its form, as FT says, then it is so most immediately in virtue of the form it appears to have from one perspective. It shouldn't matter to RA whether the paint patches on the walls have the right intrinsic shape. They might or they might not.

This image has been removed by the author for copyright reasons.

Figure 1

Marcus Raetz's sculpture, 'Yes or No?' could be described as being *simultaneously* an instance of the word 'yes' and the word 'no', in virtue of the way it appears from different positions (figure 2).

This image has been removed by the author for copyright reasons.

Figure 2

146

Someone could insist that what matters is the intrinsic two-dimensional shape of an inscription and deny that the above creations really instantiate words (or letters/graphemes) at all, claiming that they are non-words which merely happen to *look* like words. However, even mundane inscriptions might fail to have the relevant two-dimensional shape intrinsically. For example, if the ink from a pen seeps unevenly into the page there might be no two-dimensional slice of it which is letter shaped. Engravings in a bulging rock face might also be problematic.

If FT is motivated by RA then the form-theorist should only care that an inscription would appear to an appropriately situated observer to have a certain two-dimensional shape. What is required is a certain kind of phenomenological salience. To be sure, objects can meet this criterion in virtue of their intrinsic shapes, but what matters to RA is not the intrinsic shape which *grounds* its having the right apparent shape relative to an observer: non-intrinsic-duplicates can ground the same apparent two-dimensional shape just as a single object can ground multiple apparent two-dimensional shapes (see the Raetz example above). If FT is driven by RA then it is the way inscriptions look – not what grounds the way they look – which should matter.

On the face of it this applies only to inscriptions, and not to utterances. However, language users deploy a variety of cognitive resources in speech perception. The McGurk effect – in which subjects are played a recording of someone saying 'ba' laid over footage of someone mouthing 'va' and what is heard is either 'ba' or a third, intermediate sound – reveals that auditory perceptions of speech sounds are influenced by visual perceptions of lip movements (McGurk and MacDonald, 1976). If the acoustic form of an utterance and the way it sounds can come apart in this way, then what should matter to the form-theorist is not the intrinsic form of the acoustic blast, but something like the way it appears (or would appear) to hearers who are competent in the language and appropriately sensitive to e.g. visual cues. This phenomenon – where sound and apparent sound come apart – is not restricted

to exotic cases like the McGurk effect. We will encounter everyday examples below.

There is a weaker and a stronger version of this objection. The weaker version just says that there is an issue here for the form-theorist to address: just what is the notion of form which features in the best version of FT, given that it can't just be an object's intrinsic properties, as most form-theorists tend to assume? The stronger version of the objection insists that results in linguistics call into question the very idea that the forms of utterances can play the epistemological role the form-theorist wants them to. Someone who wanted to press this objection might note that perception of word form is apparently cognitively penetrable. Recall the cases, discussed above, of the McGurk effect, the equivalent pronunciation of 'writer' and 'rider' in American English, and the variation in pronunciation of 'in' depending on what follows it. Most people would be *surprised* that 'in' is pronounced differently before 'New York', 'Boston' or 'Cambridge'. That's not an acoustic effect which is usually phenomenologically salient. Similarly, Rey's point was that speakers of American English hear apparent differences between the coronal consonants in 'writer' and 'rider' even though there is none from an acoustic point of view. What such cases suggest is that the way a word sounds is influenced by our expectations, background knowledge, linguistic competence, etc., something which seems in tension with the bottom-up view. To a certain degree, the perceived form of a word is something which we impose upon it.

4. BAD NEWS FOR THE XYC SCHEMA

The unavailability of a form-theoretic criterion for word membership has a deeper consequence for the Searle-Cappelen approach to words. While the XYC schema yields a compelling story about some instances of mental/social construction, it has some severe limitations if it is intended as a general account of social phenomena. I'll argue that the account is inapplicable to some kinds of artefacts, before making the specific argument that it is inapplicable to words.

Recall that Searle holds that status functions are imposed on existing objects via collective acceptance of constitutive rules. The form of such rules is given by the schema 'X counts as Y in C'. In the XYC schema, X can be either a particular object or a kind:

> "We collectively agree that *that log* is a goalpost."

> "We collectively agree that *banknotes with a specific form and provenance* count as five pound notes."

For the words case, the X term has to be a general description of a kind. We can't have collective attitudes towards every last word token. Setting aside cases where the X property picks out an individual object, the XYC schema requires *two* ways of typing artefacts. For example, suppose we explain the monetary value of a one pound coin in terms of a constitutive rule that certain round pieces of metal issued by the Royal Mint have a value of one pound Sterling. That rule presupposes that we have a description of all and only the pound coins in terms of their shape and provenance, *as well as* a description of them in terms of their value. The two descriptions have to be separable because the X properties are taken to be metaphysically prior to the Y properties, the instantiation of the former grounding the instantiation of the latter.

If the XYC schema is to explain the characters of all public artefacts, then for any public artefact kind, we have to be able to provide two independent intensional descriptions which are extensionally adequate. Searle says as much:

> The bifurcation of the imposition of status-functions into the X and Y components has some important consequences for our investigation. First, the status expressions admit of two definitions, one in terms of the constitution (the X term) and one in terms of the imposed agentive function (the Y term). Thus currency can be defined in terms of its origin and structure: certain sorts of paper issued by the Bureau of Engraving and Printing (X term) are US currency. But currency can also be in part defined as, and indeed is described on the face of US currency as, 'legal tender for all debts,

public and private (Y term). A touchdown is when you break the plane of the goal line with the ball in your possession while the play is in progress (X term), and a touchdown is six points (Y term). (Searle, 1995:87)

The trouble is that, for many artefact kinds, we just can't provide the double description. For example, suppose a chess pawn is defined in terms of the moves it can legally make on a chess board. If these facts about pawns attach to certain objects in virtue of an XYC convention then we need an independent way of characterising the kinds of objects which count as pawns. Such a description is unlikely to be available because pawns are so multiply realisable. (This is where pawns contrast with one pound coins whose production is mechanised; additionally, damaged coins are removed from circulation, ensuring physical uniformity).

The problem is not just that no *physical* description is available to fill in the X slot: it's hard to imagine *any* double description of pawns. In the absence of a physical description of pawns, how could we fill in the X term in the schema? In order to identify all and only the pawns, we will inevitably have to specify that we're interested in the things which can move in a certain way on a chess board. So if we pack enough into X to pick out all and only the pawns, then we will already have pawns, and there will be no need to impose any additional Y features.

Since pawns are artefacts, there are at least some artefacts whose existence does not depend on any rule with the structure of Searle's XYC schema. Similar reasoning applies to words. The XYC schema requires that we can characterise a word both in terms of its X properties (form, according to Cappelen) and in terms of its Y properties (syntax, reference, etc.). Just as Searle needs to fill in the X slot, Cappelen needs to fill in the 'such and such properties' slot. Searle implies he expects a form-theoretic fix on the X entities, while Cappelen commits to this explicitly.

But the argument of this chapter has been that no single physical description can be provided which applies to all and only the tokens of a given word. Linguistic properties such as being a noun or referring to cats just don't line up neatly with independently describable types of sounds or shapes. It is always an option to introduce extra conventions, one for each attested variant pronunciation or orthography for a given word. But I have suggested above that there is unconventionalised variation which cannot be accommodated in this way. Besides, given the sheer physical variety across tokens of a word, Cappelen's proposal can only be made to work by positing a huge number of conventions. Not everyone would need to be in tacit agreement with all these conventions, but the cognitive burden on individuals would nonetheless be excessive.

I've said the X slot cannot be filled in with physical descriptions of kinds of utterances. An alternative would be to say that the X entities are picked out by some historical criterion or casual-historical lineage. I argue against this in chapter four, although I recognise that this does not constitute a decisive repudiation of the role of the XYC schema in a theory of words. Still, until some alternative way of fixing the X properties (or the "such and such" properties) is forthcoming, the Searle-Cappelen view of words looks to have some very severe limitations.

To be clear, the objection just made is not the complaint that the X slot shouldn't be filled in with a description referring to an artisan's intentions. There is nothing to stop us putting intentions into the X slot. On my view of artefacts we do precisely that when, for example, we seek to explain the fact that Frisbees are banned in certain parks. On my view, Frisbees are intentional artefacts. What makes something a Frisbee is that it is the product of a successful intention to make something recognisable as a Frisbee. Nonetheless, when we agree to ban Frisbees in parks, what happens can plausibly be modelled using the XYC schema. The X slot will be filled in with a description of Frisbees (which on my view will mention artisans' intentions) which are singled out for the Y property of being banned in parks. The trouble is that we can't explain what makes

151

something a Frisbee using the XYC schema. There is no pre-agreed convention which stipulates that all and only things with a certain form are Frisbees. Similarly, though the XYC schema may be appropriate when we're trying to explain the fact that some word is taboo, it appears wholly inadequate to the task of explaining the word's existence in the first place.

5. THE ECA ACCOUNT OF FORM

Cappelen assumes that the intentionalist shares his background assumptions about the role of XYC conventions in grounding the tokening conditions of words:

> The disagreement between intentionalists and non-intentionalists is over what to put in for "such-and-such" in "such-and-such properties" in C1. The intentionalist says this should include reference to the intentional production history, the non-intentionalist denies this. (Cappelen, 1999:99)

This is wrong. To be sure, there could be an intentionalist position which stands to Cappelen's view in the way he describes. I mentioned such a view in §3.3, and Epstein (2009) speculates that such a position could be attributed to Devitt. But that is *not* my version of intentionalism. Here, I want to distinguish my account from the XYC approach, showing that my view evades problems which afflict Cappelen's while also solving the following puzzle.

Kaplan (1990) was wrong when he said that form didn't matter to word individuation *at all*. Take the problem of failed word utterances. Suppose someone intends to say 'biology' but says something which sounds like 'philosophy', or just grunts. Intuitively, they have failed to say 'biology'. The problem is this. We want to say that form is neither necessary nor sufficient for being a token of a word, and we want to confine forms to a purely epistemological role. But we need to explain why having the *wrong* form is sometimes a deal-breaker.

On the account I have provided, words are artefacts whose characters are constitutively related to their creators' intentions. One way in which my account differs from the Searle-Cappelen view is that there is no need for collective acceptance of constitutive rules. Another difference is in the structure of the view. Recall that the XYC schema requires two ways of typing words. That is not the case for my alternative account. Let us return to the example of the chess pawn. The artisan intends their production to be recognisable as intended to have a certain functional role in games of chess. If that intention is recognisable (and acceptable), then they have made a pawn. There are many ways of making the intention recognisable: it can be done by giving it a certain characteristic pawn shape, but it can equally be done by making an explicit statement or merely by placing the object (a penny say) in the pawn position on the chess board. From a physical point of view, pawns are multiply realisable in a way which defies an XYC account of their characters. There simply is no way of specifying all and only the pawns by means of a physical description. There are too many ways of making a pawn for each of them to be conventionalised through collective acceptance of constitutive rules.

My view of artefacts offers a mechanism of social construction which only requires one way of typing artefacts (one set of K(ind)-relevant features). These K-making features get attached to a physical substrate, not in virtue of anyone having an independent type-theoretic description of the physical entities that are intended to be Ks, but in virtue of an authorial relation between an artisan and their product. On this view, creator intention is what secures the substrate via a kind of demonstrative relation. We don't need an independent type description of the things which are to be pawns, something which is agreed in advance. We can turn any number of things into pawns on the fly just by successfully signalling our intentions to each other.

Similarly, the ECA view of words seeks to explain how linguistic properties are imposed on a physical substrate, but it is not committed to anyone having a general description of the substrate. The physical entities (utterances and

inscriptions) which have linguistic features imposed on them are picked out not by a general description but by a kind of quasi-demonstrative relation between individual speakers and individual tokens. The fact that a certain utterance counts as being an utterance of 'cat' is not down to its having a certain independently specifiable acoustic profile which everyone agrees is necessary and sufficient for being a token of 'cat'. Instead, the concrete entities eligible to be instances of 'cat' are picked out by being the focus of some agent's intention that it be an utterance of 'cat' (where this means 'is intended to be recognisable as intended to be a noun, refer to cats, etc.). Of course, such intentions are not by themselves sufficient for an utterance to be a token of 'cat'. Those intentions have to be largely successfully realised. But the presence of these recognitional features does not entail that all and only the tokens of a given word can be described in form-theoretic terms, still less that any language user need be apprised of such a description.

A speaker's intentions have to be recognisable, but they do not have to be recognisable by any canonical means. The formal properties which contribute to word recognition are not themselves essential features of the word. The form of an utterance provides *clues*, but many different forms can be clues to the same word, and words can also be recognised by the company they keep. In light of this, we don't need a general agreement about the acoustic form that the Y utterances must take. The entities which are candidates for being Y are picked out not via a pre-agreed general description, but by the authorial relation between a speaker and an utterance.

It's a central part of my account of ECAs that they have a kind of communicative function. Typically this involves structural features of an artefact which can indicate what the creator's intentions were regarding that object. So, how do structural objects communicate Y properties? Doesn't this require something like Searle's XYC schema? I do not think so. In the case of the chair or wheelbarrow, the structural features which aid recognition are also the features which enable it to perform its central function. It's enough to be acquainted with

a few exemplars in order to pick up on the visual cues which tell you it's a chair. We don't need any kind of convention to explain this. In the case of the chess pawn, the Y properties are typically signalled by structural features which are only arbitrarily related to the pawn role. These structural features serve no other purpose than to indicate that the piece is a pawn. Similarly, the red pole outside a barber's shop indicates you can get your hair cut for money. In these cases, structural features are arbitrarily associated with Y features. This, to be sure, requires something like a convention in the sense that there has to be some fairly widespread understanding that such and such a structural feature indicates a certain kind of intended function. But notice here that we have X features which *indicate* Y features. They do not *ground* Y features. For something to be a pawn, it has to be recognisable as having various intended properties (such as moving forward only one square at a time). This will have to be indicated somehow, and this may be grounded in a convention of some kind, but the features which make pawns recognisable vary. In individual cases, having a certain shape makes something recognisable as a pawn, and being thus recognisable is part of what makes it a pawn, but there are no universal generalisations about the kinds of things which make pawns recognisable.

Thus, my account of artefacts enables us to tell a story about the mental/social construction of words which is a distinct rival to the Searleian XYC theory. According to my view, there can be localised conventions about the correct way to signal linguistic intentions. These conventions enable local forms to signal word type without being in any way constitutive of it. In sum, the conventions Cappelen needs are not available (see previous section), and my account does away with the need for conventions.

Finally, the above account enables a solution to the failed word problem alluded to earlier: if form is neither necessary nor sufficient for being an instance of a given word, why is it that having the wrong form sometimes a deal breaker (as in the case of intending to say 'biology' but just grunting, or perhaps uttering something which sounds like 'philosophy')? Kaplan (1990) denies that this is in

fact a deal breaker, but that seems a tough bullet to bite. On my view, the forms of utterances contribute to the recognitional mechanism which bestows wordhood on a token. Utterance forms are (local) ways of signalling clusters of intended Y properties, but no one form is essential to being a token of a given word type. But if your utterance fails to meet the signalling standards of *any* local language community, then your linguistic intentions are unrecognisable, and no on-the-fly agreement about word type can be arranged.

6. CONCLUSION

Form-theoretic accounts of words are species of externalism in the ontology of language. They focus on formal features of utterances and inscriptions because they expect these to play a distinctive role in a tempting, bottom-up story about linguistic interpretation.

I raised a number of objections against FT. There was an easy answer available to the Multiple Media problem, but the Diversity and Coincidence arguments show that the form-theoretic way of categorising tokens cross cuts other theoretically important ways of carving linguistic reality, ways which take into account, e.g. a word's syntactic or semantic properties. From the point of view of a syntactician, a semanticist, or an etymologist, the categories identified by the form-theorist appear gerrymandered: there is no obvious way of mapping the word types which are relevant in those explanatory contexts onto formal properties of the linguistic signal. I also raised a worry for the common assumption that what matters to the form-theorist is intrinsic, acoustic form, suggesting that what is needed is something more observer relative.

One major contribution of this chapter has been to articulate and then debunk what I take to be the main motivation for the form-theoretic view: the word recognition argument. My criticisms of that argument show that word recognition does not require interpreters to be able to fix on stable, readily identifiable formal features of utterances. Word type identification is an aspect of the general task of working out what has been said, not a precondition of it. If

the word recognition argument is the main argument for FT, then FT should be considered to be thoroughly undercut. The form-theoretic way of carving linguistic reality yields categories which are of no obvious interest in relation to any pressing theoretical endeavour.

A further major contribution has been to show how the unavailability of a form-theoretic characterisation of word types creates a problem for anyone wanting to deploy the Searleian XYC schema in a theory of words. Finally, I have explained how the forms of utterances enter into the ECA theory of words.

CHAPTER 4: TWO OTHER INTENTIONALIST THEORIES

In this chapter I criticise two other intentionalist theories of words, due to Kaplan (1990, 2011), and Szabo (2000).

Kaplan's (1990, 2011) *common currency* theory of words is well known, influential, and contrasts sharply with my ECA theory, so in §1 I set out his view and respond. Kaplan attempts to answer the question about what makes two utterances utterances of the same word not in terms of formal similarity, but in terms of a causal/historical connection between them, where this involves a constitutive role for human intentions. Kaplan's view is therefore a kind of intentionalism, but the account has some serious flaws: the problem is not just that his account does not capture the nature of linguistic expressions; it has general flaws which can be appreciated on the basis of the account of artefacts presented in chapter one.

In §2 I discuss Szabo's (1999) idea that utterances and inscriptions should be regarded not as instances of words, but as representations of them. On this view, words, properly so-called, are abstract particulars. These abstracta are not eternal, Platonic forms, but *created abstracta*, things which would not exist if not for the activities of humans. This second part of the chapter thus addresses issues which stand somewhat apart from the rest of the thesis. I initially became interested in Szabo's view because it chimes with aspects of my intentionalist approach, and it seemed to offer a new way of thinking about word individuation, one which would deal smoothly with the failure of form-theoretic aspirations. Unfortunately, Szabo's arguments for this alternative conception of the relation between a word and its tokens are not persuasive.

1. KAPLAN'S COMMON CURRENCY ACCOUNT

1.1 Outline of the view

Kaplan (1990) rejects the form-theoretic view of words which assumes that what makes utterances utterances of a given word is the fact that they have a certain form. He takes the demise of the form-theoretic view to entail an abandonment of the type-token conception, which he views as unnaturalistic:

> It seems to me in many ways that this is a sort of updated version of the Platonic notion of abstract forms. The eternal, unchanging Platonic forms (shapes, perhaps) are the types, and their physical embodiments, which reflect these abstract forms, are the tokens. I think that the token/type model is the wrong model for the occurrence/word distinction (i.e. the utterance/word distinction or the inscription/word distinction). The token/type model best fits what I call the orthographic conception of a word, the typesetter's conception. (Kaplan, 1990:97-98)

In its place he elaborates what he calls the *common currency* conception of words. Kaplan proposes that we understand the relation between a word and utterances of it in the terms of his *stage-continuant* model.

> I propose a quite different model according to which utterances and inscriptions are stages of words, which are the continuants made up of these interpersonal stages along with some more mysterious intrapersonal stages. I want us to give up the token/type model in favor of a stage/continuant model. This is not, I think, simply another way of doing the metaphysics of types under the old token/type conception, but a quite different conception of the fundamental elements of language. I think of my conception of a word as a naturalistic conception. Because the interpersonal transmission of words is so central to my conception, I adopt a phrase of Kripke's, and I call my notion the Common Currency conception of a word. (Kaplan, 1990:98)

On this view, words are 'made up' of physical objects such as utterances and inscriptions, as well as mental events, including whatever it is about the mind which accounts for 'thinking' a word and storing it in memory. A word is an earthbound object spread out continuously in a branching pattern through space-time. The career of a word is taken to begin with some baptismal event. People

who encounter the word store it in memory and then go on to use it, thus transmitting it to other people. Two utterances are utterances of the same word just when they are connected by a continuous path of transmission. Crucially, for Kaplan, what is stored mentally is not a mere representation of the word, but a segment of the word itself. He thinks this is necessary to secure the continued existence of a word even when no one is currently saying it. He writes that "continuity comes about (in part) from links between storage in a mental lexicon and utterances that draw upon *that very stored lexical item*" (Kaplan, 1990:514, emphasis added). Given that we could all stop saying 'Hamlet' for a day and then start saying 'Hamlet' again – the very same word - the next day, Kaplan thinks this requires an unbroken chain of contact with the original baptismal event in which 'Hamlet' was introduced:

> Do things like Hamlet and 'Hamlet' have a continuous existence? We do continue to think and talk about Hamlet, but not continuously. We pause, sleep, eat, talk about other things, and perhaps later return to Hamlet. How do we manage to return to Hamlet after a gap? Storage! This is why I insist on including the quiescent periods of storage of a word—inscribed, recorded, or stowed away in a mental lexicon—as among the stages of the word. There are performances of a word (events when a word is in performance, when it is on the move): titterings, auditions, inscribings, readings, and then there are quiescent stages of a word as it lies unread in a dusty manuscript in a dark archive. If the word 'Hamlet' were "ended" by our not only ceasing to perform it but by destroying these stages of storage, I do not see how that very word could be rediscovered. The "end" of 'Hamlet' comes about when there can no longer be a continuous path from the name's creation to a performance of it. (Kaplan, 2011:513)

What determines that an utterance is an utterance of a given word? That is, given that a subject has been exposed to many thousands of words w_1, $w_2…w_n$ (via utterances and inscriptions) what makes it the case that some utterance produced by the subject is a continuation of w_i? According to Kaplan the connection is secured via the process of intentional repetition:

> The identification of a word uttered or inscribed with one heard or read is not a matter of resemblance between the two physical embodiments (the two utterances, the two inscriptions, or the one

utterance and one inscription). Rather it is a matter of intrapersonal continuity, a matter of intention: Was it repetition? (Kaplan, 1990:104)

So a word is a continuous chain with links which are mentally stored and then intentionally extended. The form of an utterance has nothing to do with word individuation: the difference between two words can be "[j]ust about as great as you would like" (Kaplan, 1990:102). What makes all the utterances of 'omnishambles' utterances of 'omnishambles' is not their shared form or any shared linguistic properties, but rather a kind of historical continuity. The word 'omnishambles' just is this continuous, branching concrete object:

> It is this continuity of earthly embodiment that makes two utterances utterances of the same word, not some form of phonological or orthographical resemblance to an ideal form. (Kaplan, 2011:509)

> Words are…like the Kaplan family. We are a disparate bunch; we don't look or sound much alike. But we are all members of a single family, a single kind, if you will. What connects us are certain relations, but they are historical in nature and not apparent to perception… What makes two utterances utterances of the same word is that they descend from a common ancestor. This no more requires them to resemble or *replicate* that ancestor than my children are required to resemble or replicate their parents in order to be members of my family. (Kaplan, 2011:509)

Kaplan does not deny that the forms of utterances guide language users to a solution to their epistemological problem of working out what word their interlocutor intended to utter. But the metaphysical question about what makes an utterance an utterance of a given word is settled entirely by the speaker's intention:

> We depend heavily on resemblance between utterances and inscriptions…in order to divine these critical intentions. If it sounds like "duck", it probably is "duck". But we also take account of accent and idiolect and all the usual clues to intention. It is the latter that decides the matter. (Kaplan, 1990:104)

Now that we have Kaplan's view on the table, it ought to be clear that his view of word individuation can be easily accommodated within a background

metaphysics of kinds and instances. Kaplan (1990) assumes that the type-token conception requires a commitment to a mysterious world of types as ideal Platonic forms whose tokens have their status in virtue of a resemblance to the Platonic ideal. In reaction against this, and in order to provide an "earthly" conception of words, he takes words to be spatio-temporal individuals. But in doing so, he has overlooked a more natural option, which is to construe types and tokens in terms of kinds and instances (as I have done throughout this thesis). To be sure, the metaphysical status of kinds is one of the battlegrounds between nominalists and Platonists, but the point is, one doesn't have to have a settled view on these matters in order to legitimately employ the machinery of kinds and instances. Kind-instance talk is everywhere, both in everyday thinking and in scientific theorising. Moreover, Kaplan's idea that what unites the tokens of a word is some kind of historical connection between them can be embedded within the kind-instance approach to words. To say that some entities are members of a kind is merely to say that they share some salient property. There is obviously no commitment to their sharing the same physical form. On what grounds do we group certain wines and sitcoms together with the other Australian things? Not in virtue of a shared form. Moreover, there's no reason to think that the grouping of certain objects into kinds should not proceed on the basis of those objects' causal histories or their relations to human intentions. Perhaps it is the case that for an animal to be an ostrich is for it to be born of two ostriches. If so, that should not prevent us from saying that *Struthio Camelus* is a kind of animal, and that a given ostrich is an instance of that kind. Perhaps *Frankenstein* is a member of the kind, gothic novel, in virtue of Mary Shelley's intention to write a book which continued the tradition of *The Castle of Otranto*. That is not an incoherent metaphysical scenario.

McCulloch (1991) makes a similar point:

> Here we might represent him simply as having pointed out that type-identity among token words is not determined by geometrical or phonological criteria… Use of the type/token vocabulary…merely implies that the tokens are bound together by

162

> *some* relevant resemblance or similarity. And if we adopt this usage of 'type' and 'token', then, similarly, we can say that Kaplan's point is that token word-occurrences are bound together into word-types by criteria which are neither geometrical nor phonological. This leaves the type/token vocabulary apparently without strain, in place. (McCulloch, 1991:74-75)

In sum, the adoption of the type-token conception is neutral with regard to any account of the criteria by which utterances are grouped into word kinds. Kaplan's reasonable rejection of the strict form-theoretic view led him to an unreasonable rejection of the kind-instance conception. Happily, he later dropped his opposition to the kind-instance conception:

> I have now concluded that the token-type terminology is too powerful and too useful metaphorically to make it the focus of attack. I never meant to attack the abstract notion of a kind versus an instance of that kind. This is a useful idea, although there is an interesting literature on whether types are kinds. What I wanted to attack was the idea that the type was an ideal, Platonic form…and that what made a token a token of that type was that it *resembled* it. (Kaplan, 2011:509)

Unlike Cappelen's account, the common currency account is not troubled by the problems of diversity and coincidence. Instances of a word may vary enormously but they will count as instances of that word just in case they are connected by the right kind of causal-intentional chain. And there is no problem treating 'bank$_1$' and 'bank$_2$' as different words: the chains of intentional repetition reach back through independent historical lineages to distinct baptismal events. That some utterance of 'bank$_1$' is a close intrinsic match of some utterance of 'bank$_2$' is beside the point. The account is also intended to make sense of the fact that a word can be manifested in speech, writing and other media. Presumably the thought is supposed to be that since nothing in the account commits us to the idea that there is a particular form essentially associated with an utterance/inscription of any word, there's no obvious reason why a speaker cannot repeat an utterance by writing it down or repeat an inscription by saying it out loud. However, there are some serious problems afflicting the common currency account, as I will go on to show.

1.2 Cognitive implausibility

Suppose Anne says 'I've bought a goat' and I mishear and later quote her as saying 'I've bought a coat.' Wouldn't Kaplan's account yield the result that the word I actually pronounced was 'goat'? That counter-intuitive result would seem to follow from the fact that I intended to repeat Anne's very words. However, I think a reply could be made here on Kaplan's behalf. Although there was some intention to repeat what Anne said, there need not have been any intention to repeat Anne's utterance. When I quoted her as saying 'I've bought a coat' the crucial intention may have been the intention to repeat some other utterance of 'coat' which I encountered on a previous occasion. When I heard Anne speak I mistakenly took her utterance to be an instance of the same kind as some earlier utterance I encountered in the past, and it was this latter utterance that I intended to repeat when I quoted her. Given this possibility, Kaplan could avoid the counter-intuitive result that I was actually saying 'goat.'

More broadly, Kaplan is not committed to the idea that when a speaker says a word there is *any* particular *utterance* of the word which they intend to repeat. Such a view would be deeply implausible. In response to Hawthorne and Lepore's (2011) suggestion that he is committed to this view Kaplan protests:

> It's hard for me to believe I said this; it is so manifestly false…
> [W]e hear (or read) the word from someone else, store it in our
> mental lexicon, and on a later occasion draw it from the lexicon
> and use it. (Kaplan, 2011:518)

As noted above, Kaplan's view is that when we encounter a new word, a copy of it is literally stored in the mind. Thus, there need be no particular *utterance* of a word which a speaker intends to repeat. Instead what they intend to repeat is the bit of the word which is stored in the mental lexicon.

However, this view comes with its own problems. Kaplan doesn't really explain what he means by his suggestion that a word is literally stored in someone's mind. It is tempting to object that words cannot be in human minds any more than unicorns can: unicorns, of course, have horns, and there are no horns in

human brains. But that kind of argument is less persuasive in the case of words. We can't just point to some essential property of words which could not be instantiated inside a brain since what properties words have is precisely the thing at issue. Does a word have to be made of sound waves or ink? That would beg the question against Kaplan. Also, the idea that the mind contains word-like things has become familiar from work on the language of thought hypothesis. Finally, one might be sympathetic to the idea that a word could be tokened in interior monologue. In short, the idea that an instance of a word could be literally inside a mind is not nearly as far-fetched as saying that an instance of a unicorn could be.

Still, the idea that knowing a word (even when one is not thinking of it) consists in having it literally stored in the mind seems too much of a stretch. A more natural account would hold that what minds contain are representations of words, where this is understood descriptively, and as imposing certain satisfaction conditions. To take an (admittedly slightly flippant) analogy, consider the activity of a professional baker: when making croissants, the baker does not intend to replicate any particular croissant that they have encountered in the past. They just know what counts as a croissant, and they have access to a kind of recipe which tells them how to bring the ingredients together in the croissant way. To my mind, this is roughly how word production works. Speakers have, stored in memory, a representation of the linguistic profile of a word. They have a substantive conception of what it takes to produce a given word, and they can do so without referring back to any particular utterance/inscription of it.

1.3 Good intentions are not enough

On Kaplan's view, a speaker selects a word from their mental lexicon and intends to reproduce it acoustically. The successful realisation of that intention apparently consists entirely in selecting the right word from the lexicon. After that, there are apparently no more constraints. This means that our word producing intentions are more or less self-fulfilling. What this means is that it doesn't matter if something goes horribly wrong from a physiological point of

view. The view entails, for example, that the Reverend Spooner, when he said something that sounded like 'shoving leopard,' was *actually saying* 'loving shepherd' (Kaplan, 2011:520). To my mind, this is not quite right, and I have provided my own account of these kinds of cases in chapter two (§4.1). But perhaps just what to say about these kinds of malapropisms should be considered as spoils to the victor. However, Kaplan's account also predicts that a drunken slur can be a successful utterance of 'otorhinolaryngologist' however unintelligible it is, even if, say, the utterance only has one syllable and no consonants. This is surely a step too far. Citing Hawthorne and Lepore (2011), Kaplan claims to have a nice response to this worry:

> Suppose someone has a terrible accident. We ask him for his name and telephone number. He has the name and telephone number in mind and strives to speak it; he *intends* to be uttering his name and telephone number. But all that comes out is a monosyllabic grunt that sounds a bit like the word 'row'. Does my view that the intention makes it so imply that the grunt is, in fact, an utterance of the name and telephone number? It need not, because we can take advantage of Hawthorne and Lepore's excellent suggestion that the right thing to say in this case is that the injured person cannot speak. He didn't say 'row', and he didn't say what he intended to say, namely, his name and telephone number, he didn't say anything at all. *He cannot speak.* (Kaplan, 2011:519)

As it stands, this response won't do. What the person in Kaplan's example cannot do is *say words*. The effect of their injury is that they are no longer able to do whatever it is one is required to do to say a word, over and above having the right intentions. Saying words clearly requires the speaker to do something in addition to intending to say the word. On my view, what they have to do is produce a structured sound which adequately signals one's linguistic intentions.

On Kaplan's view, a speaker selects a word from the lexicon – an instance of the word literally stored in the mind – and intends to use that word. But this is not as straightforward as selecting a hammer from a toolbox and using it intentionally. If there really is a mental instance of the word stored in the lexicon, that is not the instance of the word which is then employed in speech. Rather some acoustic

copy has to be made of the mental item. But then the question arises, given that some attempts to make acoustic copies of mental words go awry (as in Kaplan's example), what makes the difference between a successful copy and an unsuccessful one? Kaplan might object that talk of copying misrepresents his view. The point of Kaplan's stage-continuant model was to allow that two utterances (or mental articulations) could be utterances of the same word even if they do not in any way resemble or replicate each other. What is key is that they have a shared common ancestor and are part of a continuous, branching, spatio-temporal object. So on Kaplan's view, an utterance is not in any sense a copy of a mental item but the continuation of one. Still, the same question arises. Given that speakers' intentions to produce a continuation of a word can go awry, what makes some utterances successful continuations while others are unsuccessful? As far as I can see, there is no way of providing a satisfying answer to this question without appealing to specifically intended linguistic properties of words, as in the ECA account.

1.4 Transparent intentions

In chapter one I suggested that an artisan has to have a fairly rich conception of the kind of artefact they are trying to make. But on Kaplan's view, there is no way for a speaker to conceptually latch on to a word via a description of its linguistic properties. There is no cluster of linguistic properties which is characteristic of that word since different spatio-temporal chunks of the word may have radically different linguistic properties. They are all chunks of the same word not in virtue of linguistic commonalities but in virtue of their being part of the same network of transmission and repetition. On Kaplan's view, when the speaker retrieves a word from the mental lexicon what they have to do is to intend to *make another one of those things*. We can think of such intentions as *transparent* since they need not involve any substantive conception of any phonetic, semantic or pragmatic properties associated with the word.

However, reprising Bloom's example of the chess pawn, note that someone who does not know the rules of chess can't make a penny (or a shell) into a pawn just

by intending it to be 'one of those' (pointing at an ostended pawn). For this to work, you have to know what pawns can do in a game of chess. Similarly, Bloom suggests that a "madman" might create a pile of dirt intending to make 'one of those' (pointing to a chair) and regard their efforts as entirely successful. The reason we don't want to say that this is a genuine chair is that the creator seems to be employing a concept which is dramatically different from our own substantive conception of chairs.

What this suggests is that is that successfully realising (from one's own perspective) a transparent intention to *make one of those* is not generally sufficient for making an artefact of the kind of which the ostended sample was an instance. In chapter one (§3.3) I discussed the possibility of *ignorant artisans*, such as the blacksmith who successfully makes a horseshoe from an exemplar despite having never heard of horses. But this requires that someone (the person commissioning the horseshoe) has a rich conception of horseshoes.

What Kaplan's account gets wrong is that it rules out the possibility that someone could in good faith *intend to make one of those*, and judge themselves to have succeeded, and yet be wrong about that. Bloom's artisan intends to recreate an ostended chair and is satisfied with their pile of dirt. To the extent that they consider their production successful, the artisan must have some criteria by which to measure their success. And to the extent that we nonetheless judge them to have been unsuccessful, we must have different criteria for what it would take to successfully recreate an ostended chair.

Bromberger (2011) records a conversation he had with his two year old granddaughter which illustrates the point in the case of words:

> Eliza: Me play.
> Sylvain: Eliza, sweetheart, say "I play."
> Eliza: Me play.
> Sylvain: No, no, not "me," "I." OK? Say "I."
> Eliza: Me.
> Sylvain: Say "ayayayay."

Eliza: Ayayayay.
Sylvain: Say "bye bye."
Eliza: Bye bye.
Sylvain: Great! Now say "I."
Eliza: Me.

It would appear that Eliza intended to repeat Bromberger's utterance of 'I,' and that she judged herself to have been successful in this regard, but the word she actually pronounced was not 'I' but 'me.'

All of this suggests that there is something fundamental missing from Kaplan's account. When an artisan creates a new artefact, it is an inadequate characterisation of what is going on merely to say that they intend to make another artefact of the same kind as some ostended artefact. Whether something is a successful reproduction of an artefact can only be measured relative to some descriptive criteria which establish conditions for success.

1.5 Common ancestor not sufficient

On the common currency view, to be an instance of a given word is to be part of a causal-intentional chain branching out from an initial baptismal event. Kaplan says that for two utterances to be utterances of the same word is for them to share a common ancestor. This suggestion rapidly runs into problems. As Kaplan well recognises, words change their forms, meanings and other linguistic properties over time. The Proto-Indo-European word kind, 'swésor' can be considered the common ancestor for the English word 'sister', the French word 'soeur', as well as many other words in modern Indo-European languages. This case illustrates not just that utterances which are connected together historically in the right way can be just too different to be considered utterances of the same word, but also that words can be subject to fission.

One might be tempted to wave this objection away, noting that all ordinary objects are subject to worries about vagueness and persistence. That is, after all, the kind of reply I make on behalf of the approach to artefacts endorsed in this thesis. But there is a sharp difference between Kaplan's account and my own

which means that he can't wave these problems away so easily. On my account, I can recognise that language change often happens in a gradual way such that we can't see where to draw any precise boundary between, say, a word of Old English and its modern descendant. But I *can* tell you on what grounds we make such distinctions. The Proto-Indo-European word 'swésor' is not the same word as the English word 'sister' because they have radically different linguistic properties: they are associated with different sounds, different phonological systems and enter into different grammatical relations. That's why we don't treat certain Indo-European utterances as utterances of modern English words. Kaplan, on the other hand, has eschewed any appeal to the linguistic profile of utterances. Being an instance of a given word just is being on the chain of intended repetition. End of story. Such an account offers no way of distinguishing between a Proto-Indo-European word and its English descendent, no explanation of a boundary, vague or otherwise.

1.6 Multiple invention

Some artefactual kinds can be multiply invented. Knives, axes or hammers may have been invented many times by disconnected individuals or groups. There may be radios and boats on distant planets. Can words be multiply invented? That is, could two completely isolated groups share a word? If they can, then we have a counterexample to Kaplan's claim that being part of a certain causal-intentional chain is necessary for being an instance of a given word.

There are real examples of pairs of words which take similar forms and meanings but which are not etymologically related. These include 'much' (English), 'mucho' (Spanish) and dog (English), 'dog' (Mbabaram). Intuitively, these are pairs of distinct words despite their similar forms and meanings. This might be thought to suggest that tokens of the same word have to be etymologically related, that being a token of a word involves being part of a branching, concrete individual, and that word types cannot be multiply invented.

My view is that words *could* be multiply invented, but for reasons we can well understand, this hardly ever happens. The examples just mentioned are not genuine cases of synchronic linguistic coincidence. 'Mucho' is not written or pronounced exactly like 'much', nor does it behave in the same way syntactically (e.g. it's marked for gender). I think the desire to hold that they are not the same word reflects a sensitivity to the fact that they have different linguistic properties. We might say that they are *similar* words, but not that they are one and the same.

So my claim is that the perceived distinction between 'mucho' and 'much' is not based on a conception of words which requires a shared etymological origin but on our sensitivity to the fact that these words have different linguistic properties. In further support of this explanation, note that 'banana' (English), 'banan' (Norwegian), 'banane' (French) can be considered to be different words despite sharing an etymology as well as many synchronic linguistic properties. There's a sense in which they can't be *quite* the same words because they are embedded in different languages and this inevitably means that they have slightly different pronunciations, orthographies, syntactic properties, etc. If this explanation of the distinction between 'banana'-like words is right, then the distinction between 'mucho' and 'much' can be explained the same way, without recourse to an etymological constraint on word-individuation.

The actual rarity of genuine synchronic linguistic coincidence is what drives the intuition that multiple invention of words cannot happen. But it could happen. According to a Wikipedia article on cross-linguistic onomatopoeia, the following languages have very similar words for expressing pleasure in eating food ('yum yum' in English):

Batak, *nyaum nyaum*
Catalan, *nyam nyam*
Danish, *nam nam*,
Estonian, *näm näm*
Hungarian, *nyamm nyamm*
Indonesian, *nyam nyam*
Korean, *nyam nyam*

Tagalog, *nam nam*

I have no idea if these words share a common ancestor. It is at least conceivable that similar sounding words emerged independently to express the same thing (perhaps because saying these words resembles a sort of chewing action). Suppose these words don't have a common origin, and suppose that they are pronounced exactly the same, and that they are employed in similar ways. I would want to say that a single word had been invented multiple times.

One way you could get a genuine case of multiple invention of words would be a scenario like the following. Suppose that a wall is built along the banks of the Seine, isolating inhabitants of Paris's left and right banks. Suppose that by pure chance the populations of both the left and right banks independently come up with a word with the same pronunciation, spelling, meaning, syntax, etc. In that case I think we should say that a single word has been multiply invented. If you don't say this, there will be costs: What happens when the wall comes down? Do the two distinct word types fuse? My answer: nothing of any metaphysical import happens at all; there was only ever one word type.

Gasparri (2014) alleges that a Kaplan-style causal-historical account is required in order to explain cases like the foregoing. To see the point, suppose that an inhabitant of Paris's left bank finds themself on the right bank and uses the word 'boudon' in conversation with a right-banker. Communication is unhindered, but there is a sense that the right banker's true beliefs about what the left-banker said are lucky. This can be explained, the thought goes, on Kaplan's view: the left-banker uses the left bank word, the right-banker assumes that the right bank word has been uttered and attributes to it the meaning that the right bank word has. As it happens, this is the exact meaning of the left bank word, but only by chance. In response to Gasparri, the luckiness of this case can be equally well explained on the ECA view. On this view, what's lucky is the right banker's inference that the word 'boudon' had been uttered, since the left banker's strategy for getting the right banker to realise their linguistic intentions was only

successful in light of an accidental matching between their individual conceptions of the word.

1.7 The irrelevance of historical properties

One final objection to the common currency view is that the historical properties are largely irrelevant to the roles words play in our lives. When we ask what word has been uttered it's because we want to know if it was a noun or a verb, whether it referred to cats or had a certain argument structure or pejorative force. Whether the utterance is part of some historical chain linking it with utterances which have quite different linguistic properties is beside the point. Moreover, people are mostly fairly ignorant of the etymologies of words and even have many *false* beliefs about etymology. Neither of these predicaments undermines communication. In short, from the point of view of ordinary language use, or of the disciplines which attend to our ordinary conceptions of language, what matters are typically the synchronic, linguistic properties of utterances.

Gasparri (2014) suggests that what he calls Kaplan's account could be defended as just one theoretical conception of words among others, one suitable for the purposes of historical linguistics, perhaps:

> The key lies in acknowledging that [the common currency account] and grammatical typing [i.e. an approach like my own] operate at different levels, pursue fundamentally distinct explanatory goals, and cannot be placed in direct competition with one another…[The common currency account] focusses on historical ancestry and sets out to deliver a *diachronic* lexical taxonomy, whereas the theoretical labour carried out in typing word tokens based on structural-functional equivalence falls in the scope of *synchronic* linguistic analysis. (Gasparri, 2016:130)

This is fair enough. If Kaplan's view is best construed as one complementary to mine and not contrary to it, then I need not push any harder at this point. Nevertheless, it remains to be shown exactly why historical linguistics should need anything like Kaplan's common currency conception of words. After all, that discipline could just as easily be understood as the study not of how words change over time, but of how they change *into one another* over time.

2. SZABO'S REPRESENTATIONALIST ACCOUNT

2.1 Outline of the view

What is the relation between a type and its tokens? Whatever it is, Szabo notes that we can get knowledge about word types even though we only ever interact with word tokens. Thus, whatever the relation is between a type and its tokens, it must be that it can provide an epistemological warrant for knowledge of types. Wetzel (2008:38) also commits to a constraint of this nature. She points out that linguists (and, I would add, the folk) typically attribute properties to word types on the basis of encountering some of their tokens, and that they do so in a principled way. This is something which needs to be explicable in light of the relation that holds between a word and its utterances.

Szabo (1999:147) points out that understanding the type-token relation in terms of that between a kind and its instances would offer a solution to his problem about how we can obtain knowledge about types on the basis of acquaintance only with tokens.

> Predicates whose nominalizations pick out kinds are projectible, so the move we make from properties of tokens to properties of types is a matter of legitimate inductive generalization. This reduces the type/token problem to the general problem of induction. Whatever explains how we can learn about the genus *Panthera Leo* by observing particular lions will also explain how we can learn about the type 'lion' by observing particular tokens of that type. If we accept that types are kinds of tokens, and hence that tokens are instances of their types, we have the following solution to the type/token problem:
>
> Instantiation view. A type T is instantiated by its tokens, and it is in virtue of this that empirical information about a token of T can play a role in justifying our knowledge about T. (Szabo, 1999:147)

The idea is that finding out about the properties of a kind just is finding out what properties its instances have. For example, finding out that *Panthera Leo* is a sexually dimorphous genus just involves looking at particular male and female lions, and noting that males and females generally look quite different. Of

course, instances of a kind typically have many properties which are not essential properties of the kind. Some lions have three legs or are situated less than 200 miles from a rocking chair, but these are not properties of the genus. But Szabo's suggestion is that the epistemic problem of determining which of a concrete particular's properties are projectible across the kind is just the familiar problem of induction. There is no special epistemological problem of the relation between a kind and its instances.

Szabo, however, thinks the kind-instance model of words should be replaced by something else. He defends the idea that words are abstract particulars and that their tokens (whether spoken or written) are representations of them.[22] This allows for a different kind of answer to the problem about how we get knowledge about types on the basis of acquaintance with tokens. After all, everything I know about blue whales is based not on acquaintance with blue whales but on acquaintance with representations of them. Learning about things on the basis of representations is ubiquitous. Of course, we have to be careful. Some representations are dubious or deceitful. We have to have some assurance that a representation is reliable before we can use it to acquire knowledge about the thing it represents. Another issue is that representations are not generally intrinsic duplicates of their representata. Only certain aspects of the representatum is selected for representation (e.g. the portraitist seeks to represent skin tone but not blood type) and the representation typically has properties which are not intended to be representational (e.g. though the portrait is made of canvas and paint, the subject of the painting is not represented as being made of the same materials). But these kinds of problems are common to quite everyday cases of finding out about things (blue whales, say) on the basis of representations of them.

That said, Szabo's representationalist account of words and their instances does have a more unusual kind of problem. Szabo is assuming that we have no knowledge of word types independently of our knowledge of tokens. None of

_____

[22] Szabo's view is given a sympathetic hearing in Ludlow (2011:58-60).

us do. That is just supposed to be a feature of the status of words as abstracta. You might wonder why he wants to presuppose that. After all, we know things about numbers and it's not clear that numbers even have tokens. Even if you think tokens of *numerals* represent numbers, it still seems that our knowledge of numbers is not entirely based on our acquaintance with those token numerals. Though it raises deep and difficult questions in epistemology, we do appear to have *a priori* knowledge about abstract mathematical objects. Perhaps Szabo's thought is that we can know things about those abstracta which conform to the Platonic ideal (eternal, necessary, etc.) via *a priori* reasoning, whereas this is not the case for created abstracta like words are (on his view.) In any case, he *does* make the assumption that we know nothing of words independently of our experience of word tokens, and this assumption raises an interesting issue.

I can find out information about blue whales without being acquainted with them directly. Acquaintance with representations of them suffices. But it only works because *some* people are acquainted with whales (or because they're acquainted with something that gives them some reason to believe in whales). If no one was acquainted with whales (or some tangible evidence for them) then there would be no representations of whales that I could trust (and probably no representations of them at all.) Once at least someone has experience of whales (or experience of other empirical phenomena that allow them to form hypotheses about whales) then they can start to produce the representations which will teach me about whales. Words aren't like this on Szabo's account. The only way *anyone* gets to learn about word types is via encountering representations (tokens) of them. Tokens, Szabo tells us, are artefacts whose function it is to represent abstract words. Who gives them that function? Presumably we do. But if all our knowledge of words comes from representations of them, how do we know enough to start producing the representations? There is apparently a vicious regress here.

Szabo does have an answer to this kind of worry. In the case of the blue whales, our representations of them merely aim to *describe* something which exists

independently of them. On Szabo's conception word tokens are not like that. Tokening a word (i.e. representing it) is a creative act. It is in virtue of the representation that the thing represented exists. He immediately anticipates a worry here:

> It seems that one cannot make something by representing it, for what is represented must exist both before and independently of its representations. (Szabo, 1999:161)

This, as Szabo goes on to point out, is a worry which can be answered, for it depends on an overly restrictive conception of representation. Many representations appear to be copies of actually existing objects. For example, a photograph can be regarded as a copy of whatever was in front of the lens. But not all representations are like this: for example, architects create drawings and blueprints of buildings before the buildings exist. So an option here is to say that the first tokens of a word represent a word that does not yet exist.

> If things go well, the token will spread through replication among speakers of the language. At some point - there is no sharp line here - the conventions that guide the use of these tokens become sufficiently widely known, and a new word type is born. (Szabo 1999:162)

The example of a blueprint for a new building is a nice one because it illustrates a way in which we can have representations of things that do not yet exist. It even goes some way to showing how we can potentially obtain knowledge about a thing based purely on encountering representations of that thing, even when no one has ever encountered the thing directly: arguably, if planning and investment has reached an advanced stage one can know on the basis of observing the blueprints that the future building will be over 100m tall, square, etc. But the blueprint example is imperfect is various ways when it comes to illustrating Szabo's idea about word types and word tokens. For one thing, Szabo thinks that representing a word by uttering and inscribing tokens of it is (eventually) sufficient for the word to exist. So representations of words have a creative power. But no amount of drawings or blueprints can be sufficient for the actual existence of the building. So one might still wonder how a

177

representation of a word can be creative in this way. Szabo has a further example to illustrate how this might work. Someone with the right kind of authority can create a new national boundary just by creating a representation of it, such as by drawing a line on a map. The national boundary example can also help illustrate how representations alone can furnish us with knowledge. By looking at a line drawn on a map we can determine that the new border is 200 miles long, that it follows a mountain range or divides homogenous cultural groups, etc.

It seems appropriate to remark at this point that *drawing a line on a map* is not strictly what creates the new border. A sufficiently powerful person can draw a line to indicate where they intend the boundary to be, but the line seems just to indicate this intention. If the national boundary thereby comes into existence, this is because the powerful person's intention becomes known. What's more, perhaps no one individual can create a national border on their own; perhaps a sufficient number of people need to start thinking the right way before the new boundary actually comes into existence. In any case, all that is needed is that people know about the intended boundary. There does not have to be a wall or a signpost on the ground.

Much of this is, of course, broadly in line with my own approach in social ontology, though there are some marked differences. On my view word tokens represent a speaker's linguistic intentions. In virtue of the recognisability and acceptability of those intentions the utterance actually constitutes an instance of the intended word. It's important to me that there's nothing unfamiliar or unearthly in this account. Words are basically just kinds of combinations of thoughts and utterances. In contrast, on Szabo's view, word tokens represent not mental states (as in my view) but abstract particulars. And utterances do not instantiate word types at all: they merely represent them.

2.2 Representation at best necessary
One potential interest of Szabo's proposal is that it promises to cut through certain problems which afflict other accounts of words. Recall the objections to

form-theoretic views of words which were discussed in chapter three (§§3.1-3.4). The Multiple Media Argument and the Diversity Argument consist in pointing out the formal variation between tokens of a given word type. But if tokens represent types rather than instantiate them, then this is no problem. A photograph and an inscription can both represent a person despite having no significant formal properties in common. The Coincidence Argument poses no problem either: two identical photographs can represent different people (identical twins, say). More broadly, Szabo's approach avoids the difficulties associated with our inability to *define* words in terms of properties of utterances. In chapter two I suggested that, for a given word, there may be no single linguistic property which is common to all tokens of the word. Again, this is no problem on Szabo's representationalist account. What all and only the tokens of a given type have in common is that they represent the very same abstract particular.

However, representing a word type could only be – at best – a necessary condition on being a token of that type, not a sufficient one: almost anything can be a representation of a word. For example, I might draw a red dot on a piece of paper and decree that it represents the word 'cat'. We can imagine that doing so might be useful in a discussion about linguistics. The red dot represents 'cat' but is not a token of it. To take another example, if I'm testing someone's vocabulary by reading out a pre-prepared list of words I might keep a tally of each word type that they recognise. The result will be a series of lines on a page, each of which represents a word type without for all that being a token of any word. So, not all representations of word types are tokens of those types. In that case we may wish to ask, of all the representations of 'cat', what makes some of them tokens? There is a certain *kind* of 'cat' representation whose instances are all and only the tokens of 'cat'. So what properties do representations have to have, in addition to representing 'cat' to be tokens of 'cat'? It's not clear how Szabo can answer these kinds of questions.

2.3 Szabo's Kilimanjaro argument

Szabo asks us to consider the following inscription:

This is the name of the highest peak in Africa: Kilimanjaro.

The name of the highest peak in Africa is not a concrete particular. But the ink patch after to colon is a concrete particular. So the inscription cannot be true. It is false to say that **Kilimanjaro** is the name of a mountain in the same sense as it would be false to say, literally, pointing to a rhinoceros, that *that is the most endangered species in Africa.* The reason is that the rhinoceros is not a species at all. However, if the token of 'Kilimanjaro' is a mere representation of the name, then we can see how the sentence could be straightforwardly true. Suppose there are no rhinoceroses around and I say, pointing to a picture of a rhinoceros, that *that is the most endangered species in Africa.* Here it is understood that I'm referring not to the picture, but to the thing it represents. The idea, then, is that the above sentence is true because the word 'this' refers not to the ink patch after the colon, but to the abstract particular the ink patch represents.

Szabo himself helpfully provides a possible response to this argument. He notes that instances of a kind can sometimes be used to call attention to the kind without being considered to be representations of the kind. Using his example, suppose someone says 'this is the largest living cat in Africa' while pointing to a particular lion. The utterance calls attention to a token lion, and this lion is supposed to make you think of the species to which it belongs, but it does not plausibly do so by representing its species. So perhaps the token of 'Kilimanjaro' in the example above does not represent its word type, but only calls attention to it.

2.4 Szabo's robot argument

Szabo asks us to consider a machine which has been designed to imitate human speech. It does so fairly well. Any competent speaker of English is able to understand the machine's utterances. But the sounds that the machine produces

are such that no human being could ever make *quite those* sounds due to the limitations of the human vocal apparatus. Szabo suggests that, in such a case, the anti-representationalist who thinks that type-token talk should be understood in terms of kinds and their instances would be under pressure to deny that the machine is producing genuine tokens of English words on the grounds that the machine is producing sounds which no human could produce. On the other hand, the representationalist is under no such pressure. After all, it is quite typical for representations of one and the same thing to vary wildly: perhaps no portraitist can represent a human face in quite the way a camera can, but that doesn't mean they do not both represent the same face.

There is, however, an easy response to this argument on behalf of the instantiationist. The fact that the sounds produced by the machine could not have been produced in exactly that form by humans is beside the point. For one thing, the instantiationist is not committed to a purely form-theoretic account of word-individuation. So what makes two utterances utterances of the same word need not be cashed out in purely acoustic terms. But even if the instantiationist wanted to stick with a purely form-theoretic account of word-individuation, there is something they can say in response to Szabo. The relevant claim would presumably be that two utterances are utterances of the same word just when they both have some set of acoustic properties. This would not rule out that some genuine tokens of a word could have some extra acoustic properties not generally shared by tokens of the same word. After all, there may be some English speakers who have such high or low voices that I could never pronounce their words the way they do. That doesn't mean that either of us is failing to produce genuine instances of English words.

2.5 Szabo's inverted word argument

Let's turn to Szabo's showpiece argument for the representationalist view. He calls it the *inverted word argument*.

Karel is a seventeen-year-old who has recently began to study English. In his book, because of a series of unfortunate misprints, there is the following chart:

13 : thirty; 14 : forty; 15 : fifty; etc.
30 : thirteen; 40 : fourteen; 50 : fifteen; etc.

He interprets this chart as any of us would, and consequently comes to believe that 'seventeen' refers to 70, and that 'seventy' refers to 17. The mistake is perfectly manifest in his writing. But it is not detectable when he speaks because in pronunciation he makes the opposite mistake: he pronounces 'seventy' as a normal English speaker would pronounce 'seventeen' and he pronounces 'seventeen' as a normal English speaker would pronounce 'seventy.' (Szabo, 1999:153)

Szabo asks us to consider what happens when Karel utters a token of the sentence 'I am seventeen.' He claims that Karel fails to know what his utterance of 'seventeen' refers to. Presumably, the thought here is that Karel's utterance only expresses the true proposition that Karel is seventeen by sheer luck, and Karel is unaware of the lucky coincidence which yielded that result. The luckiness of the scenario might then be thought to undermine Karel's claim to know what number his utterance referred to. Szabo concludes that:

[O]ur knowledge of reference is mediated by types: not knowing what types the tokens employed in an utterance belong to undermines the speaker's ability to know what those tokens refer. (Szabo, 1999:154)

Szabo says this is just what we should expect if word tokens represent their types. This is because word tokens only represent things like chairs and numbers indirectly. A token represents its type and the type represents chairs or the number seventeen. So the only way Karel could know what number he was referring to is if he knew what word type his token represented and what thing the type referred to (the number seventeen.)

Szabo thinks the same cannot be said for the instantiation account. On the instantiation account all Karel does wrong is misclassify his utterance. That

doesn't explain why he fails to have knowledge of what he is referring to. To underscore the point he introduces another character, Vladimir:

> Vladimir's textbook is free of typographical errors and his pronunciation is excellent. So he has neither of Karel's problems with 'seventeen'. However, he is confused about grammar; he believes that 'seventeen' is a verb, and has correspondingly strange beliefs about what it is for a word to be a verb. It seems clear that we have to say that Vladimir does not know what verbs are, but it seems equally clear that this confusion need not interfere with his grasp of 'seventeen'. He does not know what kind of word 'seventeen' is - or at least he does not know how to classify it with respect to a particular important division - but he is perfectly clear about what it stands for. (Szabo, 1999:155)

In response, *some* kinds of misclassification are harmless. For example, Vladimir may think that 'pig,' 'skull' and 'think' are descendants of Norman French words, and that 'prison,' 'arrive' and 'castle' came to us from the Celts. These kinds of misclassifications obviously do not undermine Vladimir's claim to know what these words refer to and won't interfere with communication. However, mistaking a noun for a verb does seem to be the kind of thing that would cause trouble. I think the most important reply to Szabo is to disagree with his assessment that Vladimir does not know what verbs are. To be sure, he may have been induced into confusion over the correct application of grammatical terms (i.e. over the use of certain public language words), but this doesn't show that he doesn't know what verbs are. After all, he distinguishes between nouns and verbs every day in ordinary linguistic communication. Szabo's argument glosses over the distinction between fully conscious knowledge (which one has acquired the technical vocabulary to articulate explicitly) and the kind of state which is characteristic of linguistic competence. If Vladimir couldn't distinguish verbs from nouns, then he would not be a competent speaker of any human language.

2.6 Are words utterance-like or thing-like?

Szabo seems to be ambivalent about word types. At the outset he gives examples which suggest that word types have exactly the properties we tend to find attached to tokens:

> [T]he English word 'water' is bisyllabic [and] the subject of the English sentence 'The sky is blue' is a definite noun phrase… (Szabo, 1999:146)

Later on, however, he seems to be thinking about things quite differently:

> If inscriptions of the word 'lion' are representations, it seems that their *representatum* must be the eternal, unchanging archetype, a *quasi*-Platonic form of lion-ness.

That is not the position which Szabo actually endorses. But he tries to get closer to a notion of what an abstract word type would be by evoking this idea of Platonic lion-ness and subtracting the Platonic associations to allow for the fact that word types are created at a particular moment in time. In any case, lion-ness, appears to have more to do with things with four legs and a mane than it does with things with syllables and argument structure.

If word types are like their tokens (e.g. they contain syllables), then the grounds for maintaining representationalism are weakened. If word types and their tokens have more or less the same properties (apart from the ones they can't share, like being abstract or being concrete) and if tokens only represent their types by actually having the properties they represent their types as having, then it appears that the relation between word types and their tokens is instantiation and not representation.

On the other hand, if word-types are not like their tokens, if they are more like Platonic forms of lion-ness, then it is not clear how we can find out anything about word types on the basis of encountering tokens. Szabo's insistence that we can learn about word-types by observing their tokens appears unfounded. What can a patch of ink on a page or an acoustic blast teach us about *lion-ness*? If we don't have *a priori* knowledge of lion-ness then surely the only way we could find out about it would be to find some tokens of lions and observe them, rather than some tokens of 'lion'. Moreover, if the word 'lion' is anything like the Platonic-form of lion-ness then it's not clear why the word 'lion' should be considered to

be a human creation (which Szabo does consider it to be), nor why the study of words should fall within the remit of linguistics.

3. CONCLUSION

In this chapter I have considered the merits of two intentionalist accounts of words. Kaplan's differs sharply from my own and is a direct rival, though I hope to have shown that the account has serious flaws, and that these flaws become salient in light of the discussions of artefacts and social construction in this thesis. Szabo's account is not strictly a rival to my view, since it is principally concerned with the nature of the metaphysical relation between types and tokens. Nonetheless, his interpretation of that relation provides a useful contrast to my own, and appears at first to offer a new way of thinking about problems of word individuation. Unfortunately, that promise is not fulfilled, and Szabo's arguments in favour of his view are not persuasive.

So far in this thesis I have presented my ECA theory of words, drawing heavily on discussions of artefacts in social ontology; I have advertised the advantages of my view, notably its connections to ideas in generative linguistics; and I have criticised other extant accounts of words. In the next chapter, I return to the relation between the ECA view and generative linguistic theory, in order to articulate and defuse various sceptical worries which Chomskians have raised against the notion of public language.

CHAPTER 5: RESISTING CHOMSKIAN WORRIES ABOUT PUBLIC LANGUAGE

I've been trying to develop an account of public language words which makes sense of their existence as part of the fabric of the social world, and as intimately related to human intentions. My view that public language phenomena can be fruitfully studied within the social sciences and philosophy – and even the very existence of public languages and words – is challenged in a large literature in linguistics and its philosophy. Sometimes these challenges appear to recommend a severe eliminativism regarding public language phenomena. Sometimes they are couched in terms of a philosophy of science which regards high-level social phenomena in general as ineligible to enter into serious scientific inquiry. Sometimes, the worry seems to flow much more narrowly from insights into language provided in generative linguistics. The purpose of this chapter is to address such challenges.

The sources of these arguments include Chomsky's own writings (Chomsky, 1993, 1995, 2000), textbooks in linguistics (Isac and Reiss, 2013, Smith, 1999) and works by Barber (2013), Jackendoff (1990, 2002, 2006), McGilvray (1999), Rey (2003a, 2003b, 2005, 2006a, 2006b, 2008, 2012), Stainton (2006, 2011, 2014), Stoljar (2014), and others. I will not attempt to provide a comprehensive overview of this literature. Instead, I will draw on it to reconstruct a series of arguments starting with the most radical and controversial challenges, and ending with those which I take to be most illuminating.

To give a sense of the sceptical views expressed by some linguists and philosophers regarding public language phenomena, it will be useful to gather some illustrative quotations. Consider the following passage from a well-known textbook on linguistics in the Chomskian tradition (in which the author uses the term *E-language* to refer to various possible conceptions of public language):

> Concentrating on E-language (or imagining that one is concentrating on E-language) is like concentrating on alchemy. There is a huge amount of data, a long tradition of scholarship which has given rise to a wealth of apparent generalizations, and a dead-end… [M]ost people now agree that there is no such field of inquiry as alchemy, and Chomsky argues that there is no such thing as E-language. He introduced the term…to cover any notion of language that isn't I-language, but with the caveat, widely ignored, that there is no such thing. (Smith, 1999:34)

Elsewhere, Smith writes (in only a slightly less sceptical tone):

> If such social or supra-personal constructs as E-languages were coherent and consistent, they might be the appropriate domain for political, mathematical or logical statements, but if they are supposed to reflect the individual human capacity for language, they are neither coherent nor definable. (Smith, 1999:31)

Isac and Reiss write in another introduction to generative linguistics:

> If we take the mentalistic approach seriously, then we have to admit that there is no entity in the world that we can characterise as "English". There is just a (large) bunch of people with fairly similar mental grammars that they can use to communicate in a way that is typically more efficient than between what we call Japanese and English speakers, because the so-called English mental grammars are more similar to each other. We will continue to use terms like "the English language," "Warlpiri plurals," and "Samoan verbs," but bear in mind that each name is just a practical label for a set of individual mental grammars that are identical with respect to a given phenomenon under analysis. (Isac and Reiss, 2013:16)

Such quotations illustrate the sceptical sentiments often expressed towards conceptions which treat language as a mind-external, public entity. One aspect of this scepticism is focussed on the ordinary notion of word. The closest analogue to the notion of word appearing in generative linguistic theory is that of a linguistic expression, where this is conceived mentalistically, as a cluster of linguistic properties which characterise speakers' mind/brains at a specific stage of linguistic processing. McGilvray sketches the view:

> According to [a Chomskian conception of words] words are mental objects. Think of a person as having something like a mental dictionary. The entries in this mental dictionary constitute a

person's "lexicon," which includes various "lexical items"… [T]he lexical items of the linguist are (ideally) defined in terms of the innately specified features that make them up. That is, they are defined in terms of "phonological," "formal," and "semantic features." The phonological features are those that, after mental computations, lead to the production of a sound, represented in the "phonetic" features at the "phonetic interface" (PHON). The formal features (N(oun), V(erb), A(djective), P(re(post)position)) and semantic features…lead, after mental computations, to the production of a specific meaning, represented in a configuration of features at the "semantic interface" (SEM). (McGilvray, 1999:95-96)

Chomsky indicates that this technical conception of a linguistic expression should be construed as an advancement over the traditional externalist conception:

> I have taken an expression to be a pair <PHON, SEM> constructed from lexical items, LI, each a complex of properties, including I-sound and I-meaning… [C]onceptual discussions of the nature of meaning and reference typically regard words and other expressions as phonetic (or orthographic) units, or as dissociated from sound or meaning; accordingly a word can change its meaning, perhaps even both its sound and meaning, and still be the same word. It is not obvious that these conventions make sense… The simplest thesis is that *an expression E has no existence apart from its properties at the interface levels*, PHON(E) and SEM(E)… (Chomsky, 2000:175, emphasis added)

In keeping with the position adopted throughout this thesis I do not intend to question the reality of the theoretical posits of generative linguistics. That science has yielded powerful insights and is deserving of a scientific realist attitude. What I want to resist is the idea that admitting I-languages into one's ontology forces public language out of it, or that the advent of generative linguistics should signal the end of theoretical inquiry into public language. I should stress that although radical attitudes towards public language are frequently expressed by generative linguists, many of them have expressed conciliatory views similar to my own. At its most cogent, the critique of public language boils down to signalling that inquiries into I-language or public language reflect different explanatory concerns, involve different methodologies, have different criteria for success and different prospects for achieving that success.

But less enlightened critiques also persist, so it is worth setting them out clearly and responding to them. The arguments to be considered in this chapter break down into three kinds:

§1 addresses *no such thing* arguments. These aim to show that mind-external words should be eliminated from any sensible ontology. These are radical and controversial views, and the arguments in favour of them are not compelling, or so I argue.

§2 addresses a more moderate kind of argument than that presented in §1. *No science* arguments allege that artefacts – and in particular words – are somehow inapt for entering into the scientific way of thinking. This raises interesting issues in the philosophy of the social sciences, but no has no devastating implications for inquiry into public language. I respond by pointing out that many successful scientific disciplines do attend to public language words (and other social entities). I also respond to the worry that since words (and other social entities) are partially constituted by human beliefs and intentions – at least on my view – members of a linguistic community are somehow protected from error regarding linguistic entities, something which would seem to make scientific inquiry pointless.

I turn in §3 to what remains of the Chomskian critique of public language, finding a reasonable stance which justifies and explains the Chomskians' focus on I-language and their ambivalence about public language. One aspect of this stance is a commitment to the idea that the significance of linguistic externalia has its source in the linguistic makeup of speakers, a position which encourages an internally directed, psychology-led inquiry into language. In light of this, externalist conceptions of language simply don't enter into generative linguistic theorising in any significant way. Another aspect is a methodological preference for the methods of the hard sciences which require theorists of language to idealise away from the chaotic complexity of social phenomena. I find these features of Chomskian philosophy to be plausible and insightful, and to have a

cautionary value to philosophy of language. They do not entail any kind of eliminativism regarding public language phenomena. In fact, I believe this strand of the Chomskian critique presupposes something like my ECA theory of words. It is by attributing this presupposition that we can interpret the sceptical arguments most charitably. The remaining differences between me and the public language sceptics largely boil down to differences of interest and different levels of optimism regarding philosophical inquiry into the chaotic world of public language.

1. NO SUCH THING ARGUMENTS

Philosophy is replete with arguments against the existence of ordinary objects, including causal redundancy arguments, sorites arguments, problems of material constitution, and others (see Thomasson, 2010 for a survey). It would be impossible to discuss each of these in this thesis. In a sense, such issues are orthogonal to present concerns since they affect non-artefacts as well (e.g. persons). Admittedly, the range of potential solutions is not always the same for artefacts and non-artefacts, but I will set such problems to one side. My focus will be on sceptical arguments which bear specifically on the status of words as human artefacts.

1.1 Rey's eliminativism

As discussed in chapter two, Rey defends a robust representationalist interpretation of generative linguistics, according to which the natural phenomena which linguists are tracking with their theoretical grammars are contentful states of individual speakers' mind/brains. With this assumption in the background Rey argues that sentences, words, morphemes, phonemes, etc. – Rey calls them *standard linguistic entities*, or SLEs for short – are *intentional inexistents* like unicorns and Sherlock Holmes. I-language states represent the speaker/hearer's environment as containing entities with linguistic properties, but there are no such entities. Words and sentences are a mere illusion. The illusion, argues Rey, is harmless: linguistic communication is a kind of *folie à deux* in which speaker and hearer are both systematically tricked into interpreting acoustic

signals as actually instantiating linguistic properties which they do not. I'll describe Rey's arguments and position in more detail before offering my response.

According to Rey's (somewhat marginal) interpretation of generative linguistics, the job of the linguist is to provide a grammar, where this is construed as a theory of a cognitive system which performs computations over contentful mental states. The output of the language faculty is thus construed as a description which represents a speech signal as having a certain structure. For example, a competent speaker would represent the sentence 'John seems to Bill to want to help himself' as having the structure in figure 3 (the example and the phrase structure diagram come from Rey, 2006a:245):
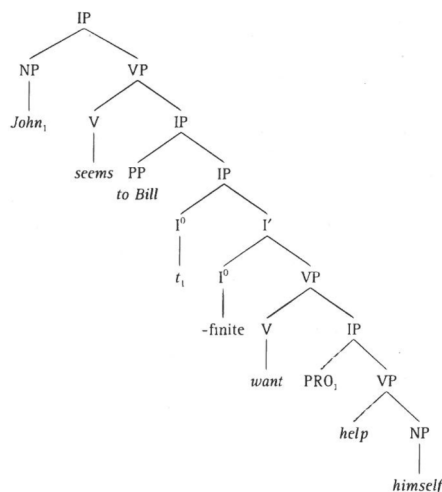

Figure 3

According to this analysis of the sentence it has a specific hierarchical structure. On Rey's interpretation of linguistic theory, the speaker literally represents the acoustic blast as having this structure. His eliminativist argument then proceeds as follows: the speaker represents the acoustic blast as having the structure shown in the diagram. But no entity in the vicinity of the speaker actually has that hierarchical structure. Utterances of the sentence do not. Neither do inscriptions of it. In fact, there just isn't anything in the world which has that hierarchical structure. So the sentence does not exist anywhere, any time.

The second premise – that nothing concrete has that hierarchical structure – requires a little care. In fact, my rejection of Rey's argument will be based on a rejection of this second premise. But there is something which is right about it: the properties which we represent the sentence as having are not intrinsic properties of the acoustic signal, nor of an inscription. A Martian scientist with a Mary-like mastery of acoustic science could determine all the acoustic facts about an utterance, including every energy intensity at every frequency throughout its duration. Would this exhaustive study of the acoustic properties of an utterance allow them to recover the structure in the tree diagram? I think not. In this sense, utterances and inscriptions do not instantiate the properties we take the sentences they are utterances and inscriptions of to have.

Recapitulating, the sentence has a hierarchical structure. The utterance does not. Therefore the utterance is not the sentence. In fact nothing has that structure. So the sentence doesn't exist. Arguments of this kind could be multiplied across an indefinite range of linguistic properties. Let me briefly note another example mentioned by Rey. Rey notes that the sentence in 'John seems to Bill to want to help himself' is analysed as featuring the phonologically null elements, trace and PRO. These phonologically null elements are a well-grounded aspect of current linguistic theory and are as real a part of the sentence as the word 'help'. They are essentially words with no sound properties. They are part of the structure of the sentence. They both refer to John. But these phonologically null elements are not present in the acoustic signal, nor are they inscribed on paper.

You might think that, even if the hierarchical structure of the sentence is not realised within the acoustic stream, at least the phonemes – the basic building blocks of SLEs – are. For example, 'cat' is made up of three distinct segments, each a phoneme of English, and the sentence invoked in 'John seems to Bill to want to help himself' is a string of 34 phonemes. Perhaps an acoustic stream counts as a token of a certain SLE in virtue of being a certain string of

phonemes. However, as Rey explains, the reality of phonemes is as much an illusion as the reality of sentences.

There are many phenomena discussed in the literature which give a feel for this point (Jackendoff (1990), Fodor et al. (1974)), but I'll keep it brief, especially since the issue was broached in chapter three (§3.2). At the phonological level of representation, SLEs can be naturally chopped up into these discrete units, phonemes. Acoustic blasts cannot be so chopped up. To be sure we can chop the stream up into discrete units any way we please, but the Martian scientist wouldn't find anything in the physical signal which determined that it should be chopped up in a way which corresponds to the sequence of phonemes we think are there. Also, even we won't find it easy to chop the stream up in the way we want. Jackendoff (1990:58) makes the following observation: 'tap' and 'back' are each made up of three discrete segments and they share the same middle segment. One and the same phoneme appears in each word. But from an acoustic point of view, there is no distinct element that utterances of the two words share. In each case the vowel sound is modified by the consonants which surround it. To take another example, in English, regular verbs are marked for past tense by a suffix, '-d'. So 'play' becomes 'played'. 'Join' becomes 'joined'. 'Hug' becomes 'hugged'. What about 'kiss', 'pick', and 'hop'? These become [kist], [pikt], and [hopt]. The words 'kissed', 'picked' and 'hopped' are mentally represented as featuring /d/. It's just that it gets transformed by the unvoiced phonemes which are adjacent to it. When we pronounce those words, no /d/ ever gets produced.

Rey takes observations such as these to show that while representations of SLEs play just the role in our cognitive economy which they are taken to play according to generative linguistics, the SLEs represented do not exist. Bits of text and bits of speech are not SLEs, nor tokens of them, for these concrete phenomena do not in fact have the properties we take SLEs to have. Nothing does. Natural language expressions are not tokened in speech and writing and they're not tokened anywhere else either. There are no words or sentences, just as

there are no ghosts and no phlogiston. There are only representations of SLEs, stable contents of computational states. We're subject to a harmless but persistent illusion that the utterances we produce have the properties which are ascribed to SLEs. This may create the impression that we literally hear and see words and sentences, though we in fact do not.

How, then, is human communication possible? After all, you might have thought that communication is possible in light of the fact that a speaker actually produces tokens of words which a hearer is able to recognise and determine the meaning of. Here is what Rey has to say on the matter:

> The hypothesis that I find implicit in at least much phonological research is what I will call the "*folie à deux*" view: the human ability to speak natural languages is based on the existence of a special faculty that includes a system for the intended production and recovery of SLEs. To a first approximation, instructions issue from speakers' phonological systems to produce certain SLEs, and these instructions cause various motions in their articulatory systems, which in turn produce various wave-forms in the air. These wave-forms turn out, however, not to reliably correspond to the SLEs specified in the instructions. All that seems to be true is that when they impinge on the auditory system of an appropriate hearer, this hearer's phonological system will be able to make an extremely good guess about the intentional content of the speaker's instructions, not about any actual SLEs, which, *ex hypothesi*, never actually got uttered. Indeed, this sort of guessing in general is so good, and the resulting perceptual illusion so vivid, that it goes largely unnoticed, and speakers and hearers alike take themselves to be producing and hearing the SLEs themselves. It is in this way that it's a kind of *folie à deux*: the speaker has the illusion of producing an SLE that the hearer has the illusion of hearing, with however the happy result that the hearer is usually able to determine precisely what the speaker intended to utter. Indeed, were SLE tokens actually to exist, it would be something of an accident. Their existence is completely inessential to the success of normal communication and to the needs of linguistic theory. (Rey, 2006a:239-240)

On this view, the phenomena which constitute the physical medium of linguistic communication are not to be conceived as realisations or tokens of SLEs, but they still have a pivotal role in communication. Rey tells us they are like 'clues'

which help a hearer to infer which SLE the speaker intended to deploy, but they are not themselves SLEs. The amount of information conveyed by these clues may vary according as the speaker speaks quickly or slowly, articulates with more or less precision, etc. The physical signal which hearers are required to interpret may even be degraded by the presence of background noise, speaking over a telephone – or in the case of a text – by intentional or accidental damage. But even in optimal cases, the physical signal will not be sufficient to determine all the properties of the intended SLE, for many of the SLE's properties will not be *in* the physical signal.

Suppose we accept Rey's background assumptions about the representationalist interpretation of generative linguistics. Suppose we also accept – as I think we should – that properly linguistic properties are not intrinsic properties of acoustic blasts. This does not entail that the blast is not a token of an SLE. What Rey's argument glosses over is the possibility that the blast could constitute an SLE token in virtue of its relational profile.[23]

Consider the following line of reasoning: a contract is a legally binding agreement between two people. But this piece of paper does not have the property of being a legally binding agreement between two people. If the Martian scientist inspected it they could never determine that it possessed that property. Therefore this piece of paper is not a contract. In fact nothing is. Contracts do not exist. What has gone wrong here? We have overlooked the relational properties of the piece of paper. That this piece of paper is legally binding is not true in virtue of the paper's intrinsic properties. It is binding in virtue of its relational properties. A similar point could be made about Marcel Duchamp's urinal. What makes it a work of art? Surely not its intrinsic properties.

What is plausible in Rey's assertion that bits of text and speech don't have some of the properties which we take SLEs to have, is that they don't have them

---

[23] This objection to Rey echoes Devitt (2006).

intrinsically. But the premise Rey needs in order for his anti-realist argument to go through is the claim that bits of text and speech do not have those properties *tout court*. This latter claim, I suggest, can be denied. Specifically, my ECA theory of words is one way of making good on the assertion that acoustic blasts constitute SLEs in virtue of their relational profile. That is, in virtue of an authorial relation between a speaker and their utterance by which the speaker intends the blast to be recognisable as having been uttered with certain intentions in mind, acoustic blasts can be viewed as having linguistic structure projected onto them.

Rey thinks he can maintain his SLE eliminativism alongside a realist attitude to ordinary objects such as his Honda. This is because while our mental representations of utterances impose linguistic conditions which are not met, by contrast, our concept of a Honda imposes conditions which *are* met (by an object in Rey's garage, for example). What are the conditions which have to be met if a thing is to be a Honda? One thing is for sure, they are not purely intrinsic conditions. Hondas have to have been made in Honda factories, or to have been commissioned by the Honda corporation. Perhaps Hondas are things which are made with certain intentions in mind, as in the theory of artefacts in chapter one. So if Rey is going to allow that non-intrinsic properties enter into the individuation of artefacts it's not clear why he thinks our representations of acoustic blasts impose conditions which are not met. To be sure, they do not meet those conditions in virtue of their purely acoustic properties, but the correct thing to say would appear to be that many aspects of our mental representations of utterances do not impose acoustic conditions in the first place. If we mentally represent an acoustic blast as featuring an instance of PRO, why assume that this imposes an acoustic constraint on the blast? On the ECA view we could argue that for PRO to be instantiated in an utterance is for the speaker's intention that PRO be instantiated be recognisable.

Ultimately, Rey cannot pick and choose in the way he wants. He thinks he can deny the existence of SLEs but maintain the existence of Hondas and other

artefacts. However, if the argument works at all, it will be too powerful and end up ruling out just about all ordinary objects. After all, what would Rey say about chess pawns? There are certain conditions on being a pawn. Presumably, these relate to the moves a pawn can make on a chess board. That a particular piece of carved wood has the relevant chess-functional profile cannot be understood in terms of the piece's material makeup. Does this mean that Rey should be a pawn-eliminativist? It is hard to see what could justify such contrasting attitudes to words and pawns. I recommend my ECA theory for both.

In short, I believe that the ECA view is preferable to Rey's *folie à deux*. One reason for this is that Rey's argument is too strong: it ends up ruling out vast swathes of ordinary objects. In addition, the ECA view helps avoid a further kind of problem pointed out by Barber (2006, 2013), according to whom Rey's view creates a puzzle about the transmission of knowledge through linguistic testimony.

1.2 Idealism about words?

In a popular textbook on generative linguistics, Isac and Reiss present an argument which appears to recommend a kind of idealism about words. On this view, it's not that words aren't real. They are. It's just that they are strictly mental entities. By way of example, they ask us to consider a Kanisza figure, an example of which is shown below:
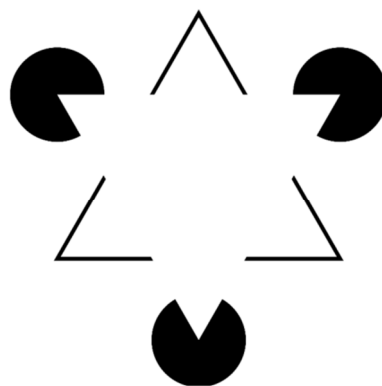

Figure 4

Most people have a perception of a downwards pointing triangle in the middle of the figure. However, from a purely physical point of view we can measure the

light reflecting off the page and determine that there is no physical distinction between the area of white 'inside' and 'outside' the triangle. Isac and Reiss explain this by saying that the mind is organised in such a way as to impose a triangle "on the page" (Isac and Reiss: 2013:22). However, this is just informal talk. They do not believe that the mind imposes a genuine triangle on the piece of paper in front of the observer's eyes. Rather the impression as of a triangle is a property which is imposed on the observer's perceptual experience. In other words, the triangle is an illusion. Noting the analogy with linguistic properties such as word boundaries and hierarchical sentence structure they conclude:

> There is no such thing as Warlpiri or *the* Warlpiri word for "child"… [T]here are just a bunch of humans whose minds contain similar kinds of rules and symbols that we informally group together as Warlpiri. Similarly there is no triangle or rectangle on these pages, but humans…all construct the same percept upon exposure to this page. (Isac and Reiss, 2013:26)

Thus far the argument exactly parallels Rey's, but where Rey explicitly embraces SLE eliminativism, Isac and Reiss make some rather obscure remarks about the 'reality' of the triangle, suggesting that if we choose to deny that the triangle is real then we are using 'real' according to an arbitrarily decided definition. They also claim that it would be "uncomfortable" to describe our typical perceptual experiences of Kanisza triangles and words as delusions. Moreover, they go on to say, it would have the consequence that the science of linguistics studies entities which are not real. What they prefer to say is that linguistic science discovers perfectly real entities which happen simply to be remote from our everyday experience of medium-sized dry goods. What is not clear is whether Isac and Reiss believe that words, languages, and Kanisza triangles are identifiable with mental structures and events. That is the interpretation which is suggested by their discussion of the Kanisza triangle. After all, they believe that the perceived triangle is real but deny that there is a physical triangle. What other candidate for the real triangle is there if not some feature of the observer's mental state?

Such a position would seem to embody a serious mistake. The fact that ghosts or Sherlock Holmes are not real does not entail that they are ideas. Nor does the fact that we have a perception as of a triangle entail that there really is a triangle *somewhere*. Nonetheless, in the previous chapter – in connection with Kaplan's view that words are literally stored in memory – I commented as follows: "It is tempting to object that words cannot be in human minds any more than unicorns can: unicorns, of course, have horns, and there are no horns in human brains. But that kind of argument is less persuasive in the case of words. We can't just point to some essential property of words which could not be instantiated inside a brain since what properties words have is precisely the thing at issue. Does a word have to be made of sound waves or ink? That would beg the question... Also, the idea that the mind contains word-like things has become familiar from work on the language of thought hypothesis. Finally, one might be sympathetic to the idea that a word could be tokened in interior monologue. In short, the idea that an instance of a word could be literally inside a mind is not nearly as far-fetched as saying that an instance of a unicorn could be."

All the same, it's one thing to suggest that words can sometimes be instantiated in the mind and yet another thing to suggest that words are only ever instantiated in the mind, which appears to be Isac and Reiss's view. The real problem with the argument parallels that of Rey's. Rey insists that representations of SLEs impose conditions which are not met and the argument ends up threatening to eliminate vast swathes of ordinary objects from our ontology. Isac and Reiss, on the other hand, make the idealist move of identifying words with mental entities. But the danger here is that if the argument were successful it would bring with it a kind of wholesale idealism about social phenomena.

There are echoes of Jackendoff (1983, 2002, 2006) in this position. Jackendoff (2002:303-304) denies that it makes sense to say that words, fictional characters, social entities and auditorily perceived objects are "objects in the world." Instead they are objects in "the world as conceptualised by [a language user]." On one

reading this could be taken to chime with my ECA approach. But Barber (2006) interprets him as a full-blown idealist:

> Idealists say that, despite never being realized in acoustic events, sentences are realized *in the minds* of language users. Ray Jackendoff adopts this position in claiming that discrete vowels and words (along with shapes, musical structures, and, presumably, syntactic structures) are elements in a 'projected world', not a real world 'out there'. (Barber, 2006:14, emphasis added)

This would be a fairly radical view. It's not just words which turn out to be mental entities on this interpretation of Jackendoff's view but also Wyoming and the Mississippi River:

> There is nothing tangible about Wyoming… It is a purely politically constructed entity, its rectilinear boundaries fixed by a stipulative act. We can touch the Mississippi River, and swim in it. But is the river the water contained in it, the bed of the river, the complex of the two? Exactly where does it end in the Gulf of Mexico, and exactly where does its tributary, the Missouri end in it? One can draw arbitrary lines on a map, but these are understood as matters of convenience and not some sort of "natural truth about the world." (Jackendoff, 2002:301)

I am unsure whether Jackendoff is the idealist Barber takes him to be. For all his talk of "pushing "the world" into the mind of the language user…right along with language" (Jackendoff, 2006:226) he also talks about Wyoming's "rectilinear boundaries." Wyoming is thus an entity with spatial properties and is presumably too big to fit into anyone's head. On the other hand, the following quotation suggests the idealist interpretation:[24]

> I would not blame the reader for being a bit suspicious of this expression "the world as conceptualized by the language user." It smacks of a certain solipsism or even deconstructionism, as though language users get to make up the world any way they want, as though one is referring to one's mental representations rather

---

[24] Also: "[W]e should properly think of "the perceptual world" (or "phenomenal world"…) not as absolute reality but as the "reality" constructed by our perceptual systems in response to whatever is "really out there… Thus the perceptual world is reality for us. Apart from the sensory inputs, percepts are entirely "trapped in the brain"; they are nothing but formal structures instantiated in neurons. But the perceptual systems give us the sense, the feeling, the affect, of objects being out there. We experience objects in the world, not percepts in our heads. That's the way we're built. (Jackendoff, 2006:228-229)

> than to the things represented. And indeed, there seems little choice. [T]he conceptualist position, has no direct connection between the form of concepts and the outside world. On this picture our thoughts seem to be trapped in our own brains. (Jackendoff, 2006:226)

It thus appears that the view defended by Jackendoff and Isac and Reiss is not the moderate view embodied in the ECA approach which considers social artefacts (for example) to be partially constituted by human intentions, but real, mind-external entities nonetheless. Their view appears to be the radical one that words are really inside minds. As these authors recognise, the kinds of features of words which they take to support an idealist theory of words are also features of ordinary non-linguistic objects, so the view amounts to a radical, wholesale idealism. It's not clear why anyone should want to believe that.

It should also be noted that Isac and Reiss's worry about linguistics studying non-existent entities can be resolved without making the idealist move they propose. One can be a word eliminativist like Rey and still think that the mental processes and representations of words are real. What linguistic science studies are the cognitive processes underlying our linguistic behaviour. Whether or not such states are such that their bearer could be held to be in error or to be experiencing an illusion is beside the point from the perspective of science. The mental states themselves are real and a perfectly worthy subject of study. Similarly, a cognitive psychologist studying religious belief is not faced with a choice between believing in God or asserting that they are studying entities which are not real. They study people's beliefs about God. Whatever the facts about God, beliefs about God do appear to be real and might be worth studying from a scientific perspective.

In this section I have considered two fairly radical perspectives on words arising from similar presuppositions in generative linguistics. Rey's eliminativism concludes that there are no such things as words (or sentences or phonemes, etc.) Speakers are deluded into thinking that acoustic blasts have linguistic structure, but in fact, nothing in the world has that structure. This doesn't hinder communication because we are all deluded in the same way, so the mistakes

cancel each other out. And it doesn't involve censuring linguistics, since that discipline is thought of by Rey as studying speakers' mental representations, which are real, even if the things they represent are not. (Although, as discussed in chapter two (§2.2), Rey's strong representationalist interpretation of linguistics is controversial.) Isac, Reiss and Jackendoff think that words and sentences are real, but they appear to identify them with strictly mental phenomena. I've suggested that the problem with these arguments is that they pertain to quite general features of the manifest image, and do not arise in light of phenomena which are specific to linguistic matters. Thus, what these authors recommend is either a wholesale eliminativism about ordinary objects, or a wholesale idealism. These are radical options, and the arguments considered here do not exclude pursuing the ECA approach to words and other artefacts. We will therefore turn our attention to arguments based in philosophy of science, and subsequently to matters more narrowly relating to generative linguistics.

2. NO SCIENCE ARGUMENTS

There is another strand to the criticism of public language associated with linguists and philosophers influenced by Chomsky. It is an approach which accepts the *existence* of ordinary words and languages, perhaps even assuming something similar to the ECA view of words, but which argues that such entities are for some reason not proper objects to feature in scientific discourse. In this section I will canvass a range of such arguments and provide a response.

2.1 Resolvable by fiat

Stainton (2006:919) suggests that public language phenomena are attended by questions whose answers seem to be resolvable by fiat: "Rather than calling out for discovery of something real, these seem matters of decision." For example, Stainton asks whether 'forge', meaning to create a fraudulent imitation and 'forge' meaning to shape metal with fire and a hammer are distinct words or really just one and the same word, and whether 'forge', 'forged', 'forges' are distinct words.[25] One problem with this is that it's not so clear that these

---

[25] This kind of objection echoes Jackendoff's claims about Wyoming, mentioned in §1.2.

questions can only be settled arbitrarily. In formulating a semantic theory, say, it might be necessary to make a decision one way or the other, but then the choice would be made in terms of the contribution it makes to the overall virtues of the theory. This is compatible with the way Stainton puts the point in a later paper (2012): there, he notes that "insofar as [the boundaries between words] get precisified, it cannot reasonably be done in terms of 'physical properties'." This, of course, is something I have argued for (especially in chapter three). Expanding upon Stainton's remark, and in light of the view developed in this thesis, we can say that wherever the joints between words types lie, they are a reflection not of physically detectable differences in speech signals, but of the way humans represent or think about those speech signals.

Similar points can be made regarding the individuation of utterances. Are utterances just the physical blasts of sound and air which come out of people's mouths? Or do certain extensions of these blasts count as utterances too? When we listen to recordings of Winston Churchill's speeches, is what comes out of the speakers or headphones describable as instances of words or just as representations of them? An analogy with sign language would suggest that what come out of the headphones are not utterances but mere representations of them: a photograph of a manual gesture is not a manual gesture but a representation of one. On the other hand, an analogy with inscriptions is less clear: is a photograph of an inscription an inscription, a representation of one, or both? Whether a photograph of an inscription is an inscription or a representation of one may depend on the intentions of the photographer.

In short, words (and utterances of them) seem to be individuated in ways which depend on our decisions and our intentions, while one way of thinking about science assumes that it inquires into how things are objectively, independently of our decisions and intentions.

Ultimately, my response to this argument (as well as others considered in this section) will be to question whether the fact that words depend on our intentions

and decisions (or have a normative character, depend on our interests, etc.) really implies that they are ineligible to enter into scientific discourse. I'll make that response in §2.4. In the meantime, I'll continue to canvass the different forms the no science arguments can take.

2.2 Normativity

Stainton (2006) suggests that certain forms commonly employed by children such as 'runned', 'broked' and 'swimmed' are not genuine English words. He also notes that there is a sense in which 'hopefully' is not supposed to mean 'it is to be hoped that', and that a dictionary reveals the existence of the English word 'peavey', even though hardly anyone knows or uses it. He concludes that what counts as a word depends on normative factors. What words there are in the language is somehow a reflection of the ways people are *supposed* to speak: "Clearly, what rules these words in or out is not how people do speak, but rather something about how they should speak" (Stainton, 2006:920).

One of the triumphs of linguistics has been the adoption of a descriptive and explanatory approach to language which sets aside normative attitudes to language. That methodological stance has permitted a wide range of inquiry into the cognitive processes underpinning our linguistic behaviour. For example, the fact that children overgeneralise rules for past tense forms is often offered as evidence for the existence of an internalised set of grammatical rules. Serious scientific inquiry, the thought goes, benefits by abstracting away from people's normative attitudes to the objects under study.

One can of course reply that a theorist of public language could either attend to or ignore normative attitudes towards words like 'runned' and 'hopefully' without in either case denying that 'runned' is a word of English (though used mostly by children) or that 'hopefully' has a public language meaning equivalent to 'it is to be hoped that.' But a deeper point is that public language, as I have construed it in this thesis, is inherently normative. The main thrust of Thomasson's (2014) work on public artefacts was to show that artefacts have a normative dimension.

On her view, that something is intended to be regarded or treated in a certain way (a crucifix, say) can be criterial for membership of a given artefact kind. On my ECA view, speakers intend for their utterances to be regarded in a certain way (as having certain intended linguistic properties). But not just any intention will be successful. If someone intends an utterance which sounds like 'cat' to refer to dogs, they will fail in their intention. There is a limited range of *accepted* ways of signifying the 'dog'-intentions. These consist in conventional associations between phonological forms and other linguistic properties. These conventions come with norms. The person does something *wrong* when they say 'cat' and mean dog. They fail to utter the word 'dog' because their utterance does not meet accepted standards for uttering 'dog'.

In short, the present version of the argument holds that the normative character of words (and other artefacts) is what makes them ineligible for scientific theorising. Science tries to describe how the world *is*, and to explain *why it is* that way, and ignores people's attitudes regarding how the world *should be*. (Again, my reply to the argument will be provided shortly.)

2.3 Interest relativity

A further kind of argument centres on the fact that words and public languages are individuated according to criteria which reflect socio-political concerns. For example, the languages of Bosnia, Croatia and Serbia are differentiated for cultural and political reasons, and not for any linguistic reasons:

> The name 'Serbo-Croat' was officially adopted with the formation of the Kingdom of the Serbs, Croats, and Slovenes (known as Yugoslavia from 1928)…This situation lasted until the collapse of Yugoslavia in the wars of 1991–1995. Since the establishment of the independent states of Bosnia Herzegovina, Croatia, and Serbia and Montenegro (still officially known as Yugoslavia until 2003), the term 'Serbo-Croat' no longer has any official validity in sociopolitical terms. The language spoken in these countries is now officially known as Bosnian, Croatian, and Serbian, respectively. In linguistic terms, the standard language remains essentially the same, but the sociopolitical reality is that it no longer has a single name. (*Encyclopedia of Language and Linguistics*, 2006:259-260)

205

Suppose, then, that the boundaries between public languages are drawn in ways which reflect socio-political concerns rather than linguistic properties. If it were then assumed (perhaps implausibly) that words cannot belong to more than one public language, then a Bosnian's utterance of 'knjiga' is literally of a different word than a Croat's utterance of 'knjiga'. In other words, word individuation will be infected with the same kinds of socio-political, interest-relative criteria as public languages. These ways of categorising words and languages ignore their properly linguistic properties: 'knjiga' has the same phonology, syntax and semantics in both Bosnian and Croatian, and Bosnian and Croatian speech is mutually intelligible.

One quick response to this argument is that although public languages are *sometimes* counted in a way which reflects political tensions, recent wars, national boundaries, etc., this does not mean that they *are always* or *can only be* counted in that way: that a distinction is sometimes made between Bosnian and Croat does not entail that there is no viable notion of public language which abstracts away from precisely those socio-political parameters. It is, of course, much harder to show *how* to count public languages in a way which abstracts away from socio-political concerns. In this thesis I've largely set issues about whole public languages to one side, focussing instead on words, and I don't claim to have a theory about how to draw the boundaries between public languages. I will therefore pursue a different strategy, that of allowing that the characters of words and public languages can, in some sense, be related to socio-political concerns. What I will question is the idea that this means they have to be excluded from scientific inquiry.

2.4 Responses to the above

The above arguments each have something like the following form:

1. Nothing with feature F can be targeted in serious science.
2. Words have F.
3. Therefore words cannot be targeted in serious science.

Features which have been proposed as explaining the science-inaptness of words have included the fact that conceptions of public language involve elements which are apparently decidable by fiat, the fact they are suffused with normativity, or that they depend on human interests including socio-political concerns. This, I trust, is what Chomsky is suggesting in the following passage:

> The idea [of a common public language] is completely foreign to the empirical study of language... What are called 'languages' or 'dialects' in ordinary usage are complex amalgams determined by colors on maps, oceans, political institutions and so on, with obscure normative-teleological aspects (Chomsky 1993, pp. 18–19).

As McGilvray puts it:

> [M]aking linguistic behaviour the subject matter of serious science is to attempt to make a stable natural object out of what are, in fact, highly variable, interest-dependent, context sensitive, creative efforts of people. (McGilvray, 1999:110)

Since science seeks to abstract away from such matters, taking an objective and value free stance, the argument goes, ordinary notions of words and languages have no place in science.[26] One thing to notice is that the arguments just presented, though framed mostly in terms of linguistic cases, would have implications for artefacts and social phenomena in general. If successful, they would cut a very large swathe through the social sciences.

My response to these arguments has two main strands. The first is that nothing has been said, so far, about *why* things which are norm-governed, interest-relative, etc. are ineligible to enter into scientific discourse. I'll consider two possible answers to this question, finding each unsatisfactory. The second strand to my response is what Stainton (2012) calls the Moorean response, which disputes the alleged scientific inaptness of artefacts and other social phenomena on the

---

[26] Stainton, (2012) puts the point in terms of Sellars' distinction between the scientific and manifest images: "What is metaphysically special about public languages…is that like Hinduism, jazz, Tuesdays, [and] yarmulkes…these are only objects for us. All are value-laden; all are tacitly suffused with human free will. In brief, all pertain to Sellars' (1963) Manifest Image. In sharp contrast…a science aims to afford a human-transcendant, value free, law-governed description of certain phenomena."

grounds that there are a multitude of scientific disciplines which study such phenomena (including linguistic phenomena) with great success.

Just why do Chomskians assume that social entities with the kinds of features I've been highlighting are ineligible to be targeted in serious scientific inquiry? Perhaps one reason relates to the idea that, on the ECA view (and views within the standard approach in social ontology more generally), competent speakers enjoy a kind of protection from error or ignorance regarding their public language creations. The thought may be that since linguistic artefacts have their characteristic properties *assigned* to them by intention or perhaps by collective agreement there is a sense in which we can't be wrong about the properties of linguistic artefacts. One implication of this is that a science of public language couldn't teach us anything we don't already know.

It is true, on my view at least, that people are to a certain extent protected from error with respect to the mentally/socially constructed entities sustained by their own beliefs and intentions. For example, an artificer of a new kind of artefact delineates a set of intended features which constitute an artefactual kind. They cannot be wrong about what it takes for a thing to be a thing of that kind, for that is something they have personally stipulated (although they can be mistaken about whether some object constitutes a largely successful realisation of an artefactual kind, i.e. whether or not a thing has the intended features). But there are many ways in which we can be severely ignorant about artefacts and other social phenomena, in ways which can be fruitfully addressed by scientific inquiry.

First, although artefacts depend for their nature on human representations, they are also partly constituted by physical entities which have properties which can be investigated scientifically without reference to the mental states of language users. The physical properties of linguistic artefacts are studied in acoustic phonetics, articulatory phonetics, etc. It is worth noting, however, that this will not provide much succour to someone defending the aptness for scientific inquiry of words and public languages. After all, I've spent a great deal of time in

208

this thesis emphasising that specifically linguistic properties are not to be found among the intrinsic properties of acoustic blasts and ink patches.

Second, our artefact-regarding intentions are not always easily articulable, as was discussed in chapter two (§2). In particular, speaking is a speedy and spontaneous affair and our linguistic intentions are less than fully conscious. In order to generate linguistic artefacts, these intentions have to be contentful and consciously accessible (in some sense), though artisans need not be able to articulate these intentions in any way. I suggested in chapter two that elucidating our language-regarding beliefs and intentions requires a systematic approach. Generative linguistics itself can be seen as shedding light on our underlying intentions regarding language (without assuming that generative linguistics targets intentions specifically or exclusively), as can some work in truth-conditional semantics.

Third, many artefacts depend not on any single individual's beliefs and intentions but on those of a wider community. No individual's intentions tell the whole story (for many kinds of artefacts). In addition, facts about those kinds of social entities may not be evident to people from outside the specific group or society in which those entities play a role, even if they are evident to those within the group or society. In light of this, it looks like – for many artefacts at least – there is a kind of protection from massive error, but this applies not at the level of any individual, but at that of some larger group. We can't all be wrong all the time, but any individual can be wrong about just about anything. (It's unclear to what extent this response can be pressed in the case of linguistic artefacts. To be sure, no individual has a complete grasp of a public language; on the other hand, we can learn a lot about a nearly-extinct public language if we have access to the last remaining native speaker.)

Fourth, there's a lot about artefacts which cannot be learned by reflecting on people's attitudes towards them. Artefacts are real, they interact with each other,

with people, and the laws of nature, often in unpredictable ways. As Thomasson writes:

> [S]ocial science does not typically concern itself with such issues as whether or not it is in the nature of a pencil to be a writing instrument or what the necessary conditions are for something to be a dollar bill, but rather with such issues as the impact of the printing press on the growth of religion in Europe, the consequences of mechanical production on urban growth and standards of living. (Thomasson, 2003b:34)

Though particular social entities and kinds may have their characteristic properties via a process of mental/social construction, they are embedded in the world in ways which are not evident independently of systematic inquiry. The numerical values of our currency units may be stipulated and not a matter of scientific discovery, but the same cannot be said for the effect of lowering interest rates on property values, exchange rates or immigration levels. In the same vein, historical linguistics inquires into, for example, the effect of the Norman Invasion on English grammar and pronunciation, while other disciplines study the kinds of vocabulary used by sports journalists when discussing male athletes in comparison to when they are discussing female athletes, the ways in which police use of public address systems affects the behaviour of large crowds, etc. These are not matters which any individual has knowledge of independently of systematic inquiry.

I've been considering one possible reason why ECAs (and other things which are dependent on human intentions, norm-governed, related to socio-political concerns, etc.) might be considered the wrong kinds of entities for targeting in scientific inquiry. Now let's consider another. Stoljar (2015) interprets Chomsky's (2000) sceptical stance on referential semantics in light of Lewis's (1983) discussion of natural properties, and the argument is very closely related to our present concerns. I'll first present Stoljar's reconstruction of the argument before making the minor adaptations that are needed for the present context. Stoljar notes that, for Lewis, natural properties are associated with a "package" of features including being sparse, conferring objective similarity on objects which

share the property, figuring in laws of nature, and – along with other natural properties – forming a small set of properties upon which all other (less natural) properties supervene. Finally, naturalness comes in degrees. This metaphysical picture is implemented in a philosophy of science which holds that the aim of science is to discover very natural properties, with physics perhaps aspiring to discover perfectly natural, fundamental properties, and chemistry and biology aiming to discover not perfectly natural but still very natural properties.

Stoljar then imagines the Chomskian argument to proceed as follows: if semantics is a genuine science, then the objects and properties which have to exist if semantic theories are true should be very natural properties and objects. But if semantic theories are true then what has to exist are such things as London (since referential semantics posits a relation between the word 'London' and London), a messy, gerrymandered, social object if ever there was one.[27] Supplementing the argument in the obvious way to adapt it to present concerns, we would note that words themselves (not to mention word-world relations such as reference) are hardly very natural entities. Since naturalistic inquiry in general should only be concerned with very natural entities, it should not be concerned with public languages or words.

This response is helpful, but it doesn't really justify the claim that words can't be studied in the manner of the natural sciences. It does explain why words aren't targeted in fundamental physics, but that doesn't rule out the fruitfulness of a social scientific discipline targeting public language phenomena.

A further problem with the argument is that it does nothing to address a host of counterexamples to the thesis that public language phenomena are excluded from naturalistic inquiry. Stainton (2012) offers what he takes to be a Moorean response to the scientistic argument. He says that though he does not have a philosophical account of how it is that social entities can enter into scientific theorising, nor of how a science inquiring into social entities could occupy a

---

[27] Stoljar notes that related arguments can be found in Bloomfield (1933) and Fodor (1980).

middle position between the hardest sciences, and ordinary, non-scientific discourse, his rejection of the *no science* arguments considered here is based on the Moorean certainty that there *are* such sciences. There just are disciplines which target norm-governed, interest-relative, not-very-natural phenomena. He (2006, 2012, 2014) gives the examples of archaeology, criminology, ecology, economics, epidemiology, gerontology, horticulture, medicine and social psychology, as well as – on the language side – clinical linguistics, computational linguistics, dialectology, discourse analysis, educational linguistics, forensic linguistics, historical linguistics, lexicography and pragmatics. So Stainton's response is that *there just are* sciences which target social entities and kinds, and in particular there are such disciplines which target ordinary languages and words.

Unpacking this response a little, that criminology, say, is a science which studies such phenomena as punishment, deterrence, incarceration, theft, murder, manslaughter, corporate crime, and domestic abuse suggests that social entities whose character is determined relative to human interests, norms, etc. are not necessarily precluded from being studied scientifically. Criminology can also target artefacts, such as in studies of different kinds of weapons used in different types of crimes, perhaps seeking to explain these facts with reference to laws and social structures within a society. Similarly, that forensic linguistics, say, is a science which targets linguistic externalia is a counterexample to the conclusions of the sceptical arguments canvassed above.

At this point, it is worth pointing out a different line of retrenchment that the Chomskian might defend: although there are sciences which study social entities, they could suggest that such sciences differ in important ways from the kinds of scientific endeavours in physics, chemistry, biology and cognitive science. The latter projects seek to explain phenomena in terms of deeper principles, to reveal the joints in nature. In contrast, the argument might go, social sciences and sciences of the complex are merely descriptive or taxonomic sciences which seek statistical regularities, perhaps, but do not offer the promise of deep theoretical understanding.

Evaluating this argument would raise issues in philosophy of science which are beyond the scope of this thesis. One thing to say is that if the assumptions of this argument are correct, it remains an option for the public language theorist to accept that philosophy and the social sciences are not engaged in a science of the deep, explanatory kind and to construe their ambitions in a different way. I'll come back to this kind of response in §3.2.

In this section I've considered a range of arguments which aim to show that words are the wrong kinds of things to enter into scientific discourse in light of their status as human made artefacts whose characters depend on people's intentions, or on societal norms, or on socio-political concerns. However, I have replied that these arguments do not provide an adequate explanation of why such entities can't be targeted in science. I have also argued that there are sciences which target such entities. In light of this, it appears that the attempt to discredit inquiry into public language by pointing to general facts about artefacts is not successful. In the next section I turn to considerations which flow more narrowly from generative linguistic theorising.

3. DON'T LOOK THERE ARGUMENTS

So far in this chapter I've discussed some fairly radical arguments which either seek the elimination of words from our ontology (at least of words construed as kinds of externalia) or which aim to show that social entities such as public language entities are not the kinds of things which submit to naturalistic inquiry. I think all of these radical arguments can be resisted. Nonetheless, there remain – in the most sophisticated Chomskian critiques of public language – two themes which deserve to be drawn out. That is the work to be undertaken in this section. The first point is that public language phenomena are not required to enter into generative linguistic theory in any substantive way. The second is a healthy methodological scepticism about naturalistic inquiry into public language, based on the idea that it constitutes an *interaction effect*. I am in substantial agreement with both of these points. These critiques do not amount to any kind of radical

rejection of common sense linguistic entities. Moreover, they seem to *presuppose* something very like my ECA theory of linguistic artefacts. I will end the section with some comments on what these conclusions mean for the status of inquiry into public language phenomena.

3.1 The irrelevance of public language to generative linguistics

Theorising in generative linguistics is associated with a methodological stance called *methodological naturalism*, (see Collins, 2010). One aspect of this stance is that scientific inquiry should not be held hostage to common sense. While scientific inquiry may sometimes take common sense conceptions as a starting point of inquiry, as progress is made new concepts will be introduced which explain the phenomena. The form such explanations take need not be constrained in any way by common sense conceptions. As Collins notes:

> [T]he concepts employed in successful scientific inquiry are typically invented for purpose and have only a loose correlation with commonsensical notions, which are often eliminated or retained just for motivational purposes. (Collins, 2010:45)

To take a simple example, a scientific inquiry might begin with a common sense conception of *air* as a simple substance which is eventually replaced with detailed knowledge of the different gases in the Earth's atmosphere. From the point of view of atmospheric science, there isn't really any such thing as air, just molecules of nitrogen, oxygen, carbon dioxide, etc.; the notion of air is not required in order to formulate our best theories of atmospheric phenomena. To subsume this phenomenon under a general principle, Collins formulates his *naturalistic reality principle*:

> (NRP) At a given stage of inquiry, a category is taken to be (naturalistically) real iff it is either successfully targeted by naturalistic inquiry or essentially enters into the explanations of such inquiry. (Collins, 2010:47)

Assuming that there is *no* science which takes *air* to be a central concept, NRP entails that air is not "naturalistically real". Collins then alleges that conceptions of public language fail to meet the demands of NRP. The claim is not that public

214

language entities fail to exist. Instead, it is that public language categories are not regarded as linguistic types from the point of view of serious linguistic theorising. That is, it is possible to state the theory without mentioning so-called linguistic externalia. Genuinely linguistic entities (i.e. linguistic states of individuals) interact causally with speech sounds, but there are no *generalisations* associating linguistic types with types of externalia. Collins writes:

> The externalia are not necessary because linguistic structure can be realized in a wholly internal manner, as in private monologue. They are not sufficient because (i) the richness of linguistic structure far outstrips any external signature and (ii) the apparently unlimited heterogeneity of the externalia recruitable in linguistic performance does not admit generalizations mappable onto the linguistic categories; for example, there are no independently specifiable features common between tokens of the sound type /dog/ and tokens of the inscription type 'dog' such that both are ambiguous between verbal and nominal categorisation. In other words, the relevant categories are invariant over external differences and so cannot be identified with externalia without eliding that which does remain invariant over the recruitment of externalia, viz., the cognition of the competent speaker/hearer. (Collins, 2010:48)

That acoustic types cannot be mapped to linguistic types should be a familiar point by now. What this shows is that linguistic theory is not concerned, in the first instance, with acoustic phenomena. Generative linguistics seeks to explain aspects of linguistic performance in terms of human cognitive structures, just as one would seek to explain human visual capacities in terms of cognitive structures. That utterances and inscriptions of 'dog' are ambiguous between verbal and nominal characterisation, as Collins notes, is not primarily a fact about the utterances and inscriptions, but about our response to them. A linguist's grammar is a theory of the internal mental structures which determine this reponse. To see this, consider one of the principle methodologies employed in linguistic theory. A central source of data for linguistics is that of speakers' acceptability judgements. For example, a subject will be shown various written sentences and asked to judge which are acceptable strings in their language. The data are not facts about speakers judgements of grammaticality (whatever that might mean to the informant). Instead, linguists take as data the mere fact that a

certain set of strings are judged acceptable and others not. When we have a set of judgements we then seek an explanation for why these strings are judged acceptable and these not. What is needed is a theory of the workings of the human mind. What is the nature of the internal processes which leads to these judgements and not different ones? Collins writes:

> Thus, the notion of *being grammatical* does not apply to strings in the first instance; rather, a theory of the language faculty (a theory of *being grammatical*, if you will) brings with it a delineation of the structures that the FLN output. How these structures affect the pattern of acceptability, which we take to be a FLB interaction effect, is what we want to discover. (Collins, 2010:50)

In other words, the point is not that words don't exist. We can very well maintain, as Collins (2010:48) does, that "externalia are invested with linguistic significance." But we reach an immediate theoretical dead-end if we treat externalia as the focus of linguistic inquiry. A satisfying explanation of linguistic externalia requires adopting the internalist perspective. Linguistics studies internal representations and the faculty competence which underlies them. These states are the source of linguistic properties. From a scientific perspective these are the proper focus of an explanatory scientific inquiry. Language as an external object only has linguistic properties derivatively. It's by pursuing internalist explorations of human linguistic competence that we may hope to *explain* the linguistic properties of mind-external entities.[28]

One might try to maintain that linguistic properties inhere in linguistic externalia in a wholly mind-independent way. A competent speaker would then be understood as someone who has managed to discern the linguistic structures which are characteristic of their language community. The trouble with this, according to Collins, is that we simply have no way of determining the linguistic

---

[28] A similar argument is endorsed by Rey. According to him, the problem with "social, response-dependent" accounts of words is their failure to take account of the fact that the arrows of linguistic explanation point inwards towards the internal linguistic capacity of individuals: "[T]he underlying error [of accounts of linguistic entities in terms of social, response-dependent proposals] is a failure to appreciate the important shift of the explanatory locus in modern linguistics, from external objects to internal conceptions" (Rey, 2008:177).

properties of externalia without going through individuals' conceptions of those externalia:

> One can always claim the linguistic structure is external…but if the structure is identifiable only through the cognitive resources of the speaker/hearer…then the externalia lose any independence as a proper parameter in any serious explanatory practice. If the structure is to be depicted as genuinely external for our best science, then, lest it become an explanatory dangler, we should be able to identify it independently of the relation it bears to our cognition… [W]e must insist that the externalia be something more than a reflection of the cognitive design with which we have already credited the subject. (Collins, 2010:48-49)

The argument is that even if we affirm that language is some mind-external linguistic system, if we are to explain humans' competence with that system, we will need to credit individuals with cognizance of that system. But since we have no evidence that externalia have linguistic properties other than the fact that we conceive of them in a certain way, the hypothesis that there really is some linguistic system independently of such conceptions is taken to be explanatorily vacuous. Linguistic properties considered as properties of mind/brains do real explanatory work, the Chomskians say. Projecting such properties onto concrete (or abstract) external objects adds no explanatory value.[29]

Summarising, Collins is arguing that linguistic externalia do not enter into generative linguistic theorising in any serious way, and therefore fail to meet the demands of NRP. The conclusion that public languages (and by extension public language words) are not "naturalistically real" suggests an eliminativist attitude towards public language, especially given that the subtitle of the article is "Why there is no such thing as language." However, that is not the main thrust of

---

[29] As Collins notes: "The only thing that could explain the data is that which explains why speakers judge as they do, and such an explanation is, at best, essentially postponed by attributing properties to the strings, for the speaker/hearers must be cognisant of those very properties if they are to be on explanatory duty. An externalism of grammatical properties, therefore, looks to be explanatorily supererogatory; the externalism confuses what the language faculty (FLN and FLB) enables – the projection of structure onto sounds/marks – with the target of explanation itself – the capacity to project, inter alia" (Collins, 2010:50)

Collins's argument. His main concern is to show that public language notions are entirely dispensable from the point of view of serious linguistic theorising:

> My submission is that language, construed in an externalist manner, fails to meet NRP, its identification being arbitrary and non-essential to the explanation of the salient phenomena. If this is so, then while languages so construed might be real in some or other sense (in the same way, perhaps, games or pieces of music might be) they are not naturalistically respectable… [S]uch entities do not offer properties that are either necessary or sufficient for the characterisation of the linguistic structures posited in current linguistics, i.e. they are not linguistic types. (Collins, 2010:48)

> We remain free to posit languages as abstractions and are equally free to have whatever ontological attitude we like towards them, and generate ontological conundrums much as we do about money or games or pieces of music. It is not the aim of science, however, to cleave to our quotidian ontological scheme. (Collins, 2010:55)

In short, Collins is arguing that serious theorising in generative linguistics has no need for public language notions, the latter not entering into the theory in any ineliminable way. Nor does affirming this mean affirming that linguistic externalia have *nothing* to do with the entities studied in generative linguistics. Collins can allow that linguistic externalia in a child's environment constitute the *primary linguistic data* – the environmental stimuli – which play a role in determining parameter settings in a child's language faculty as a natural part of language acquisition. In addition, people's person level judgements about utterances and inscriptions, judgements of acceptability, synonymy, etc., can be a significant source of evidence for theoretical linguistics. Nonetheless, descriptions of kinds of linguistic externalia are not necessary in order to state the fundamental commitments of linguistic theory.

One possible objection to this way of thinking is a kind of *tu quoque*. A longstanding aspect of methodological naturalism is a commitment to the idea that there are no in principle restrictions on the potential explanatory links between disciplines, however unrelated they may appear (see, e.g. Anthony, 2003). One might accuse Collins of a double standard here. After all, he is at pains to

insist that "naturalistic inquiry into language has no need for the notion of [external] language" (Collins, 2010:41). This would, however, be unfair. First, he could rightly insist that that generative linguistics has dispensed with public language notions as a result of extensive empirical inquiry, not by ideological pre-emption. Second, Collins need not deny that new evidence for the internal workings of the language faculty could come from newly discovered, ancient, clay tablets inscribed with an unknown language. Moreover, the generative linguist does not deny the *interest* of sociological approaches to language (or need not). The reason they do not pursue connections between I-language and public language has more to do with methodological preferences and scepticism about the possibility of scientific progress regarding public language. (Such criticisms are clearly related to those considered in §2, but there is a way of framing the Chomskian position in a way which relates specifically to public language. This will be discussed in the next sub-section.

This is one point at which something I take myself to be in agreement with the Chomskians. When it comes to explaining language, the arrows of explanation point inwards. I take it that talk of linguistic structure being projected onto linguistic externalia resonates deeply with the ECA account of words. The *head-first* direction of explanation was part of the ECA view from the start. That is one of the reasons why I take it that something like my account of words, far from being challenged by the Chomskians, is actually presupposed by them. It is only within the framework of a theory which allows mental representations to imbue externalia with significance that Collins's (2010:48-50) assertions about externalia being "invested with linguistic significance," and as having structure "projected" onto them can be fully appreciated. So I argued, in any case, in chapter two.

Nevertheless, there are two points I want to make here. First, Collins has convincingly argued that public language does not enter into generative linguistic theorising. He has *not* argued that public language does not enter into any serious science. But that is what he would have to argue in order to show that public

language phenomena fail to satisfy NRP. As suggested in the previous section, linguistic externalia do enter into such disciplines as forensic linguistics, clinical linguistics, semantics (on some construals), and so on.

My second point is that public language is at once richer and poorer than I-language. It is poorer because some of the mental operations of the language faculty are completely unconscious. These will not be taken to enter into a speaker's conception of linguistic artefacts. It is richer because the study of linguistic externalia, including public words and languages, is not exhausted by a study of humans' specialised linguistic competence. The strings outputted by FLN are taken up by other cognitive systems associated with FLB, so the properties projected on to linguistic externalia will not be fully described by a theory of FLN. For example, an utterance may be intended to refer to cats, or to a particular cat. This is a feature of the speaker's speech act, and does not reduce straightforwardly to the workings of FLN. Referring is something people do. To limit one's interest to the study of I-language, though fine for a practitioner of a specialised science, shouldn't satisfy a philosopher since it leaves out so much of interest. This point would not be denied by the Chomskians, though some would insist that theoretical progress is only possible by abstracting away from all the things the philosophers are interested in and pursuing the narrow approach of generative linguistics. That is the topic to which I now turn.

3.2 A healthy methodological scepticism

I've been trying to emphasise that my ECA theory of words is not only compatible with modern linguistic theory, but that something like it is also presupposed by many of the Chomskian critiques of public language. It is precisely because the status of linguistic externalia is projected onto them by human attitudes and intentions that the internalistic approach is pursued in generative linguistics. What human linguistic competence (FLN and FLB) enables is the ability to use an external system of signs. The performance of speech acts using such signs is a cognitive achievement involving not just a narrow computational capacity but also quite general factors including "memory,

attention and communicative intention" (see the quotation from Collins, below). These are assumptions which are shared equally by me and the Chomskians. The criticism of philosophers' tendencies to focus on public language need not, therefore, reflect contradictory views of human psychology, nor of the metaphysics of words. Instead, the disagreements reflect different theoretical interests, and perhaps different assessments of the prospects for progress regarding the chaotic domain of public language.

In the previous section I discussed arguments which aim to show that words can't be part of science in virtue of their mere status as artefacts. Those arguments were not persuasive. In this sub-section I will try to say something more reasonable about Chomskians' aversion to a science of public language.

This critique is again driven by the Chomskians' methodological naturalism. Genuine science is understood as seeking underlying principles which offer a deep explanation of certain phenomena. What is sought is not merely a detailed description of the phenomena, but an explanation. This will standardly involve an approach which seeks to idealise away from the messy detail of observable phenomena. Collins explains the point as follows:

> Our commonsense conception of language covers a host of disparate phenomena. Any endeavour to gain theoretical traction on this manifold is obliged to fractionate the phenomena and idealize away from the massive interaction effects that produce our normal linguistic behaviour… The first move in this 'divide and conquer' direction in recent times was Chomsky's (1965) distinction between competence and performance. A whole range of factors enter into performance, many of them perfectly general, such as memory, attention, and communicative intention. Chomsky's distinction, in part, was an effort to isolate the hypothesised unique linguistic system that underlies certain peculiar features of our performance, and, in the first instance, we are interested in those features simply because they are the ones that submit to theoretical understanding (the drunk looks for his keys under the street light because that is where the light is). (Collins, 2010:46)

So the idea is that linguistic behaviour is fantastically complicated, and in order to make any theoretical progress we need to ignore some of this complexity. The distinction between competence and performance, as well as that between FLN and FLB, is an attempt to gain insight into a specific aspect of human mental architecture.

Mind-external linguistic phenomena – construed as ECAs or something similar – involve relations between linguistic entities and human agents, their actions, beliefs, intentions, interests, etc. From a specifically psychological point of view, the mental processes which characterise individual speech acts involve interactions between FLN and other cognitive systems. The products of human action reflect a complex interaction between different cognitive systems, different individuals, communities, etc. The products of human action can therefore be regarded as *interaction effects*. But, the argument goes, science tries to discover deeper principles which abstract away from the noise associated with interaction effects to uncover aspects of a more systematic reality beneath the phenomena. Chomsky often uses the following example: it is the business of no science to theorise about the precise causes of the trajectory of a leaf on a windy day. For one thing, determining those causes would not lead to the discovery of any deeper principles about the nature of reality: it would just be a vast list of uninteresting details. Second, it may not be *possible*, from an epistemic view, to make any headway with such a complex interaction effect. Relatedly, Chomsky often makes remarks to the effect that coming up with a comprehensive theory of how people use words to communicate and talk about the world would require a 'science of everything', something which no discipline ever purports to provide.[30]

I agree with the general picture. I do think that the social sciences and philosophy are (some of the time) inquiring into complex domains where the

---

[30] In response to the point I made earlier, that disciplines such as forensic linguistics do in fact target linguistic externalia, it might be alleged that forensic linguistics is not a serious explanatory science in the same way as generative linguistics. Instead it seeks merely to describe and taxonomise phenomena.

prospects for immediate progress are somewhat dim. But this is no reason to stop trying! One way of thinking about the philosophy of language is as a kind of quasi-scientific speculation, an enterprise which hasn't quite achieved the empirical success and explanatory power associated with mature scientific theories, but which nonetheless shares many of the same goals. Alternatively, philosophy of language could be construed as doing something quite different from science. Chomsky sometimes seems to imply that attending to matters of public language is best done in novels or art:

> Plainly, a naturalistic approach does not exclude other ways of trying to comprehend the world. Someone committed to it (as I am) can consistently believe (as I do) that we learn much more of human interest about how people think and feel and act by reading novels or studying history or the activities of ordinary life than from all of naturalistic psychology, and perhaps always will; similarly, the arts may offer appreciation of the heavens to which astrophysics cannot aspire." (Chomsky, 2000:77)

I'm not sure how many philosophers would be ready to convert to writing novels, but I think that there is hope for a more theoretical and systematic approach to theorising about public language. Theorising about public language does not entail renouncing all ambition of being systematic, rigorous, empirically corrigible, etc.

In short, the Chomskian critique of public language need not be construed as opposing public language views, in their conceptions of human psychology or the metaphysics of words. A major aspect of the critique boils down to a pessimistic assessment of likely progress. What's new?

Finally, it is a longstanding feature of the scientific philosophy associated with generative linguistics (methodological naturalism) that scientific theories should aim for unification with other disciplines. This does not require that we be able to literally derive higher-order theories from lower-order ones, just that we aim to uncover links between theories at different levels, and understand broadly how phenomena at different levels relate to each other. This attitude explains the

223

efforts of linguists to integrate the discipline with theories in neurobiology. To my mind, this willingness to seek unification downwards with neurobiology would sit uneasily with a rejection of any attempt to unify social scientific approaches to language with theories in generative linguistics. Part of the point of this thesis has been to explore the extent to which a theory of public language can be illuminated by pursuing connections with generative linguistics. The prospects for integration between the disciplines may be remote, but the goal is not unworthy.

4. CONCLUSION

I started this chapter by considering some radical arguments which involved taking either an eliminativist or an idealist attitude towards public language phenomena. These arguments failed to demonstrate that any feature *specific* to linguistic artefacts rendered them metaphysically dubious. Instead, the arguments turned on features of linguistic artefacts which are common to many kinds of ordinary objects. Thus, the arguments considered in the first section are no more plausible than wholesale eliminativism of ordinary objects, or wholesale idealism. If we are happy to keep cars, vegetables, schools, contracts and coins in our ontology then it would be discriminatory to exclude words and languages.

I then considered a secondary line of retrenchment for the public language sceptic which consists in allowing the existence of words and languages but denying that they are the kinds of things which feature in scientific inquiry. Again, these arguments turn not on the specifics of linguistic artefacts, but on general features of artefacts and social entities. I responded that no clear reason had been provided to explain why artefacts and social entities can't enter into scientific theorising, and also pointed out that they do.

Finally, I argued that the best versions of the Chomskian challenge, far from rejecting the artefactual account I've been elaborating, actually presuppose that something like it is correct. They insist that naturalistic inquiry into language is best pursued by ignoring and abstracting away from concerns with public

language, and they express a healthy methodological scepticism about the prospects for a fruitful, explanatory science of public language. I agree with them on both points. None of this, I have suggested, impugns the practice of philosophers and social scientists pursuing explorations of public language, though it does perhaps impose an obligation on public language theorists to explain just what they take the status of their discipline to be.

CONCLUSION

In the first chapter I used Thomasson's theory of artefacts to develop a notion of essentially communicative artefacts. In particular, I drew on her (2014) and pursued a clearer understanding of the role of intended recognisability in determining the character of ECAs. This work provided the resources to bring insights from social ontology to bear on debates about words in philosophy of language. The result is a novel theory of the nature of words, which puts speaker intention at the forefront of the account. Like Frisbees, chopsticks, chairs, pens, knives, houses, footballs and desk lamps, words are essentially intentional artefacts: their instances are partially constituted by their creators' intentions. Moreover, like some other essentially intentional artefacts, such as wedding rings, chess pawns, traffic lights, military uniforms, foot-high fences, police tape and crucifixes, words are essentially communicative artefacts: that is, at least some of the features of a word utterance serve only to signal their creators' intentions regarding that utterance, and it is only in virtue of the recognisability of these intentions that the utterance has the linguistic profile it has. The signalling of linguistic intentions is possible in virtue of an arbitrary association between word forms and clusters of linguistic intentions.

I have argued that this conception of words captures an ordinary, pre-theoretic conception of words, one which is on active duty in ordinary cases of language use, including inter-personal communication. This hypothesis draws support from explaining widespread intuitions about such phenomena as swamp words, minimal creation, and exaptation. In addition, the ECA theory was originally motivated on more general grounds, not related specifically to linguistic artefacts. That the ECA theory is independently plausible as an account of (non-linguistic) artefacts means I have not been introducing specialist machinery just to deal with problems in the theory of words. The theory also enables a convincing story about such phenomena as malapropisms, spelling errors and typos.

Cappelen's approach to words has two features in common with my own: it treats the theory of words as part of social ontology, and it assumes that words are to be characterised in terms of their synchronic, specifically linguistic properties. However, Cappelen defends a form-theoretic approach to word individuation. I have rehearsed the well-known arguments against this approach, adding a further criticism pertaining to the form-theorist's assumption that intrinsic, acoustic form is what matters to the theory. And I spent time articulating and debunking the often ignored motivation which underlies form-theoretic assumptions. A further significant contribution in this area has been to question the suitability of Searle's XYC schema as a way of theorising about words: I have argued that this approach would require the possibility of providing two type descriptions of word types, one to fill the X slot, the other to fill the Y slot; the demise of the form-theoretic approach is a serious obstacle to fulfilling this requirement. Happily, my ECA account has the resources to tell a more convincing story about the role of utterance forms in the metaphysics and epistemology of words.

Another rival account considered here was Kaplan's common currency account. Like my own, this view accords a constitutive role to speaker intentions, but it has shortcomings which can be appreciated from the social ontological perspective developed here. For example, Kaplan's mistaken idea that having the right intention could be sufficient for uttering a given word is remedied by appeal to a general feature of the theory of artefacts: our artefact-regarding intentions are not self-fulfilling; they impose satisfaction conditions; our intentions are successfully realised to the extent that the world conforms to those conditions. To take another example, Kaplan assumes that our word-producing intentions can be transparent, involving an intention merely to *make another one of those*, without requiring that the speaker have a substantive conception of the linguistic profile associated with the word. The discussion of the Frisbee and the plate in chapter one (§3.2) reveals the shortcomings of this view. Thomasson (2003b) made a similar point about artefacts, but the parallel objection to Kaplan has not been made, to my knowledge. The discussion of artefacts yielded a further

objection to Kaplan's view: I have argued that words (like many other artefact kinds) could be multiply invented, something ruled out by Kaplan's commitment to individuating words in terms of common ancestry.

From the outset, one of my goals has been to provide an account which is at least compatible with insights in generative linguistics, but which also aspires to a kind of unification with current linguistic theory. Unlike other attempts to marry an intentionalist account of words with an interpretation of generative linguistic theory, I have refused to depart from the current, orthodox interpretation of that theory. That is, I do not, like Barber, insist upon an interpretation which holds that the mental states posited in linguistic theory are assumed within the theory to be genuinely contentful. In light of this, there is no way of reading off an account of speakers' linguistic intentions from a linguist's grammar (a theory with phonological, syntactic and semantic components). A grammar is intended to model an aspect of linguistic competence, but it abstracts away from the messy detail of linguistic performance, and the phenomena it targets are partly unconscious.

The promised unification of the ECA account of words and generative linguistics consists in a shared approach in the philosophies of mind and science. Crucially, the ECA account shares with generative linguistics a head-first approach to explaining the significance of linguistic externalia. The linguistic properties of utterances and inscriptions are treated as mental projections, the ultimate source of linguistic properties being the internal structure of the mind/brain. Moreover, it is agreed that language use is a person-level phenomenon involving ordinary conscious knowledge and communicative intentions. The disagreement is about whether we can make any significant theoretical progress in the study of public language. Since language, broadly construed, is the product of an interaction between various cognitive systems (between FLN and the various other systems which make up FLB), the broad conception of language is thought to be too complex and chaotic to submit to naturalistic inquiry. Nonetheless, when we ask – in the most general terms – what language is, a full answer requires an

228

approach which draws on the contrasting conceptions of language provided in linguistics and philosophy, and attends to language as both a biological and a cultural phenomenon. I've tried to show how part of that story might go.

REFERENCES

Alward, Peter (2005). "Between the lines of age: reflections on the metaphysics of words." *Pacific Philosophical Quarterly*, 86 (2):172-187.

Antony, Louise M. (ed.) (2003). *Chomsky and His Critics*. Malden MA: Blackwell Publishing.

Antony, Louise M. (2003). "Rabbit-pots and supernovas: On the relevance of psychological data to linguistic theory." In Alex Barber (ed.), (2003) *Epistemology of Language*. Oxford University Press.

Baker, Lynne Rudder (2004). "The ontology of artifacts." *Philosophical Explorations*, 7 (2):99 – 111.

Barber, Alex (ed.) (2003). *Epistemology of Language*. Oxford University Press.

Barber, Alex (2006). "Testimony and Illusion." *Croatian Journal of Philosophy*, 6 (3):401-429.

Barber, Alex (2013) "Understanding as knowledge of meaning." *Philosophy Compass*, 8 (10):964-977.

Bloom, Paul (1996). "Intention, history, and artifact concepts." *Cognition*, 60 (1):1-29.

Bloomfield, Leonard (1933). *Language*. New York: Henry Holt.

Bromberger, Sylvain (2011). "What are words? Comments on Kaplan (1990), on Hawthorne and Lepore, and on the issue." *Journal of Philosophy*, 108 (9):486-503.

Bromberger, S. and Halle, M. (2000). *The ontology of phonology* in Burton-Roberts, N., Carr, P., and Docherty, G. (eds.) (2000). *Phonological Knowledge: Conceptual and Empirical Issues*. Oxford University Press UK.

Brown, K. (2006). *Encyclopedia of Language and Linguistics*. Elsevier.

Cappelen, Herman (1999). "Intentions in words." *Noûs*, 33 (1):92-102.

Chomsky, Noam (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, Noam (1993). *Language and Thought*. Moyer Bell.

Chomsky, Noam (1995) *The Minimalist Program*, The MIT Press.

Chomsky, Noam (2000) *New Horizons in the Study of Language and Mind*, Cambridge University Press.

Chomsky, Noam (2003). "Reply to Millikan." In Antony, L.M. (ed.), *Chomsky and His Critics*. Malden Ma: Blackwell, pp. 308-315.

Chomsky, Noam and Halle, Morris (1968). *The Sound Pattern of English*. New York: Harper & Row.

Collins, John M. (2005). "Faculty disputes." *Mind and Language*, 19 (5):503-33.

Collins, John (2008). "Knowledge of Language Redux." *Croatian Journal of Philosophy*, 8 (1):3-43.

Collins, John (2010). "Naturalism in the philosophy of language; or, why there is no such thing as language." In Sarah Sawyer (ed.) (2010) *New Waves in Philosophy of Language*. Palgrave Macmillan.

Davidson, Donald (1984). *Inquiries into Truth and Interpretation*, Oxford University Press.

Dennett, Daniel C. (1990). "The interpretation of texts, people and other artifacts." *Philosophy and Phenomenological Research*, 50:177-194.

De Saussure, Ferdinand (1916/1983). *Course in general linguistics*. La Salle, IL: Open Court.

Devitt, Michael & Sterelny, Kim (1999). *Language and Reality*, 2nd edition. John Wiley & Sons.

Devitt, Michael (2006). *Ignorance of Language*. Oxford: Clarendon Press.

Dipert, Randall R. (1993). *Artifacts, Art Works, and Agency*. Temple University Press.

Dretske, Fred (1981). *Knowledge and the Flow of Information*. MIT Press.

Elder, Crawford (2007). "On the place of artifacts in ontology." In Eric Margolis & Stephen Laurence (eds.), *Creations of the Mind: Theories of Artifacts and Their Representation*. Oxford University Press, pp. 33-51.

Egan, Frances (2003). "Naturalistic inquiry: Where does mental representation fit in?" In Louise M. Antony (ed.), *Chomsky and His Critics*. Malden MA: Blackwell Publishing 89-104.

Egan, Frances (2012). "Representationalism." In Eric Margolis, Richard Samuels & Stephen Stich (eds.), *The Oxford Handbook of Philosophy and Cognitive Science*. OUP.

Egan, Frances (2013). "How to think about mental content." *Philosophical Studies*, (1):1-21.

Epstein, Brian (2008). "The internal and the external in linguistic explanation." *Croatian Journal of Philosophy*, 8 (22):77-111.

Epstein, Brian (2009). "Grounds, convention, and the metaphysics of linguistic tokens." *Croatian Journal of Philosophy*, 9 (1):45-67.

Epstein, Brian (2013). "Social Objects Without Intentions." In Anita Konzelmann Ziv & Hans Bernhard Schmid (eds.), *Institutions, Emotions, and Group Agents: Contributions to Social Ontology*. pp. 53-68.

Epstein, Brian (2015). *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford University Press.

Evans, Gareth (1981). "Reply: Semantic theory and tacit knowledge." In S. Holtzman & Christopher M. Leich (eds.), *Wittgenstein: To Follow a Rule*. Routledge

Fodor, Jerry; A., Bever ; T. G., Garrett & M., F. (1974). *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. Mcgraw-Hill.

Fodor, J. A. (1980). "Methodological solipsism considered as a research strategy in cognitive psychology." *Behavioral and Brain Sciences*, 3 (1):63.

Fodor, Jerry (1981). "Some Notes on What Linguistics Is About". In N. Block (ed.) *Readings in Philosophy of Psychology*, Vol. 2. Cambridge: Harvard University Press, pp. 197–207.

Gasparri, Luca (2016). "Originalism about Word Types." *Thought: A Journal of Philosophy*, 5 (2):126-133.

Grice, H. P. (1957). "Meaning." *Philosophical Review*, 66 (3):377-388.

Guala, Francesco (2007). "The Philosophy of Social Science: Metaphysical and Empirical." *Philosophy Compass*, 2 (6):954-980.

Hardie, C. D. (1936). "The formal mode of speech." *Analysis*, 4 (2/3):46 - 48.

Hauser, Marc D., Chomsky, Noam, and Fitch, Tecumseh W. (2002). "The faculty of language: What is it, who has it, and how did it evolve?" *Science*, 298 (22):1569-1579.

Heck, Richard (2006). "Reason and Language." In C. Macdonald & G. Macdonald (eds.), *McDowell and His Critics*. Blackwell. pp22-45.

Hawthorne, John & Lepore, Ernest (2011). "On words." *Journal of Philosophy*, 108 (9):447-485.

Hilpinen, Risto (1993). "Authors and Artifacts." *Proceedings of the Aristotelian Society*, 93:155-178.

Hilpinen, Risto, (2011). "Artifact." *The Stanford Encyclopedia of Philosophy* (Winter Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2011/entries/artifact/>.

Isac, Daniela & Reiss, Charles (2008). *I-Language: An Introduction to Linguistics as Cognitive Science*, Oxford University Press.

Jackendoff, Ray S. (1983). *Semantics and Cognition*. Cambridge: MIT Press.

Jackendoff, Ray S. (1990). *Semantic Structures*. Cambridge: MIT Press.

Jackendoff, Ray (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press.

Jackendoff, Ray S. (2006). "Locating meaning in the mind (where it belongs)." In Robert J. Stainton (ed.), *Contemporary Debates in Cognitive Science*. Malden MA: Blackwell Publishing.

Jackendoff, Ray, S. (2007) *Language, Consciousness, Culture: Essays on Mental Structure*. Cambridge: MIT Press.

Julien, M. (2006). "Word." In Brown, Keith, (ed.) *Encyclopedia of Language and Linguistics*, Elsevier, pp.617-624.

Kaplan, David (1990). "Words." *Aristotelian Society Supplementary Volume*, 64 (1):93-119.

Kaplan, David (2011). "Words on words." *Journal of Philosophy*, 108 (9):504-529.

Katz, Jerrold J. (ed.) (1985). *The Philosophy of Linguistics*. Oxford University Press.

Khalidi, Muhammad Ali (2015). "Three Kinds of Social Kinds." *Philosophy and Phenomenological Research*, 90 (1):96-112.

Kornblith, Hilary (1980). "Referring to artifacts." *Philosophical Review*, 89 (1):109-114.

Laver, J. (1994). *Principles of Phonetics* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Lewis, David (1983). "New work for a theory of universals." *Australasian Journal of Philosophy*, 61 (December):343-377.

Liberman, A.M. (1996). *Speech: A Special Code.* Cambridge, MA: MIT Press.

Lowe, E.J. (2014). "How real are artefacts and artefact kinds?" In Franssen, Maarten ; Kroes, Peter ; Vermaas, Pieter & Reydon, Thomas A. C. (eds.) (2014). *Artefact Kinds: Ontology and the Human-made World.* Synthese Library.

Ludlow, Peter (2011). *The Philosophy of Generative Linguistics.* Oxford University Press.

Margolis, Eric & Laurence, Stephen (eds.) (2007). *Creations of the Mind: Theories of Artifacts and Their Representation.* Oxford: University Press.

Marslen-Wilson, W.D. (1987). "Functional parallelism in spoken word-recognition." *Cognition*, 25(1-2):71-102.

Matthews, P.H. (1974; 1991). *Morphology 2nd edition.* Cambridge Textbooks in Linguistics, Cambridge: Cambridge University Press.

McArthur, Tom (1992). *The Oxford Companion to the English Language.* Oxford: Oxford University Press.

McGilvray, James A. (1999). *Chomsky: Language, Mind, and Politics.* Polity Press.

McGurk, H. & MacDonald, J. (1976). "Hearing lips and seeing voices." *Nature*, (264): 746-748.

Miller, G. & Isard, S. (1962). "Some perceptual consequences of linguistic rules." *Journal of Verbal Learning and Verbal Behaviour*, Volume 2; Issue 3., 217-228.

Millikan, Ruth G. (1984). *Language, Thought and Other Biological Categories.* MIT Press.

Millikan, Ruth G. (2003). "In defense of public language." in Louise M. Antony & H. Hornstein (eds.) *Chomsky and His Critics*, Blackwell.

McCulloch, Gregory (1991). "Making Sense of Words." *Analysis*, 51 (2):73 - 79.

Modrak, Deborah K. W. (2001). *Aristotle's Theory of Language and Meaning.* Cambridge University Press.

Predelli, Stefano (2010). "Malapropisms and the simple picture of communication." *Mind and Language*, 25 (3):329-345.

Putnam, Hilary (1981). *Reason, Truth, and History*. Cambridge University Press.

Pylyshyn, Zenon W., (1988). "Rules and representations: Chomsky and representational realism."

Quine, W. V. (1987). *Quiddities: An Intermittently Philosophical Dictionary*. Belknap Press of Harvard University Press.

Rey, G. (1996). "Resisting Primitive Compulsions." *Philosophy and Phenomenological Research*, 56(2), 419.

Rey, Georges (2003a). "Intentional content and a Chomskian linguistics." In Alex Barber (ed.), *Epistemology of Language*. Oxford University Press 140-186.

Rey, Georges (2003b). "Chomsky, Intentionality, and a CRTT." In Louise M. Antony (ed.), *Chomsky and His Critics*. Malden Ma: Blackwell Publishing 105-139.

Rey, Georges (2005). "Mind, Intentionality and Inexistence." *Croatian Journal of Philosophy*, 5 (3):389-415.

Rey, Georges (2006a). "The intentional inexistence of language—but not cars." In Robert J. Stainton (ed.) *Contemporary Debates in Cognitive Science*, Malden Ma: Blackwell Publishing 237-55.

Rey, Georges (2006b). "Conventions, intuitions and linguistic inexistents." *Croatian Journal of Philosophy*, 6 (3):549-569.

Rey, Georges (2008). "In defense of folieism." *Croatian Journal of Philosophy*, 8 (2):177-202.

Rey, Georges (2012). "Externalism and inexistence in early content." In R. Schantz (ed.), *Prospects for Meaning*. De Gruyter.

Sapir, E. (1933). "The psychological reality of phonemes." In Katz, Jerrold J. (ed.) (1985). *The Philosophy of Linguistics*. Oxford University Press, pp.65-79.

Sawyer, Sarah (ed.) (2010). *New Waves in Philosophy of Language*. Palgrave Macmillan.

Searle, John (1995). *The Construction of Social Reality*. Free Press.

Searle, J. R. (2007). "Social ontology and the philosophy of science." In Eric Margolis & Stephen Laurence (eds.), *Creations of the Mind: Theories of Artifacts and Their Representation*. Oxford University Press 3-17.

Sellars, Wilfred (1963). *Empiricism and the Philosophy of Mind*. London: Routledge & Kegan Paul Ltd.

Smith, Neil (1999) *Chomsky: Ideas and Ideals*, Cambridge University Press.

Sperber, Dan (2007). "Seedless grapes: Nature and culture." In Eric Margolis & Stephen Laurence (eds.), *Creations of the Mind: Theories of Artifacts and Their Representation*. Oxford University Press, pp.124-137.

Sperber, D. & Wilson, D. (1995). *Relevance*. Blackwell.

Stainton, Robert J. (2006). "Meaning and reference: Some Chomskian themes." In Ernest Lepore & Barry C. Smith (eds.), *The Oxford Handbook of Philosophy of Language*. Oxford University Press, 913-940.

Stainton, Robert J. (2012). "In defense of public languages." *Linguistics and Philosophy*, 34 (5):479-488.

Stainton, Robert J. (2014). "Philosophy of linguistics." Oxford Handbooks Online.

Stebbing, L. S. (1935). "Sounds, shapes, and words." *Aristotelian Society Supplementary Volume*, 14 (1):1-21.

Stoljar, D. (2015). "Chomsky, London and Lewis." *Analysis*, 75 (1):16-22.

Szabó, Zoltán Gendler (1999). "Expressions and their representations." *Philosophical Quarterly*, 49 (195):145-163.

Thomasson, Amie L. (1999). *Fiction and Metaphysics*. Cambridge University Press.

Thomasson, Amie (2003a). "Foundation for a Social Ontology." *ProtoSociology*, 18.

Thomasson, Amie L. (2003b). "Realism and human kinds." *Philosophy and Phenomenological Research*, 67 (3):580–609.

Thomasson, Amie (2007). "Artifacts and human concepts." In Eric Margolis & Stephen Laurence (eds.), *Creations of the Mind: Theories of Artifacts and Their Representation*. Oxford University Press, pp.52-73.

Thomasson, Amie L. (2009). "Social entities." In Robin Le Poidevin (ed.), *The Routledge Companion to Metaphysics*. Routledge

Thomasson, Amie L. (2010). "The controversy over the existence of ordinary objects." *Philosophy Compass*, 5 (7):591-601.

Thomasson, Amie (2014). "Public artifacts, intentions and norms." In Pieter Vermaas et. al., (eds.) *Artefact Kinds: Ontology and the Human-Made World*, Springer: Synthese Library, Switzerland, 45-62.

Van Fraassen, Bas & Sigman, Jill (1993). "Interpretation in Science and in the Arts." In George Levine (ed.), *Realism and Representation*. University of Wisconsin Press.

Warren, R. & Warren R. (1970). "Auditory illusions and confusions." *Scientific American*, 223 (6), 30-36.

Wetzel, Linda (2008). *Types and Tokens: On Abstract Objects*, MIT Press.

Wetzel, Linda (2009). "Linguistic Reality." In Barber, Alex & Stainton, Robert (eds.) *Concise Encyclopedia of Philosophy of Language and Linguistics*. Elsevier, pp.368-371.