Teaching learners to process morphosyntactic cues: The passive voice in second language English

Sophie Thompson-Lee

PhD

University of York

Education and Language and Linguistic Science

September 2020

Abstract

The extent to which second language (L2) learners make use of morphosyntactic cues when parsing L2 sentences remains unclear, especially in relation to explicit instruction. The present study investigated the effect of repeated explicit practice in using morphosyntactic cues during sentence processing on online and offline language use.

The processing problem was the passive voice in English for first language (L1) Chinese learners of English. One hour of computerised training forced learners' attention to the function of verb morphology *-ed* versus *-ing* (and, separately, *by*) for assigning subject / object roles (e.g. *The boy is called* [*by the man*], *The boy is calling* [*the man*]*).*

73 upper-intermediate Chinese learners of English were assigned to one of three groups: 1) Explicit information (EI) + cue focused practice, 2) EI + noun focused, or 3) test only. 29 English L1 speakers were also tested. The cue focused training required learners to use morphosyntactic cues in order to complete activities, whereas the noun focused training did not require morphosyntactic cue use. Pre, immediate, and delayed post-tests were administered to assess the impact of training on comprehension and production. These were; visual world eye-tracking, written production, oral production, and written grammaticality judgement.

The results suggested that learners benefitted from EI + practice in using morphosyntactic cues for assigning roles, as the cue focused practice group made production and grammaticality judgement gains sooner than the other groups. Training had mixed effects on online processing (as evidenced by eye-tracking). The findings suggested that both conscious and unconscious behaviours can be influenced by instruction, and that teaching grammatical cues may be of some benefit in the language classroom.

Contents

List of tables

List of figures

Acknowledgements

I would like to express my sincerest thanks to my supervisor, Professor Emma Marsden, for your support, encouragement, and guidance. I would also like to thank you for the many opportunities you have given me outside of the PhD. I have gained so much valuable experience during my time working with you.

I am also deeply grateful to my Thesis Advisory Panel supervisor, Professor Leah Roberts, for your invaluable advice on the design of my study and the analysis of the data.

I would like to thank the many people who helped me with various aspects of my study and thesis, I am eternally grateful. Special thanks to Dr Alice Cruickshank for the many hours spent helping me learn how to use eye-tracking, Justyna Mackiewicz for your help with scoring, Ashleigh Hawkins for your advice and help with Chinese translation, and to Xiaoran Niu for your help with piloting my study.

Thank you to all of the participants who took part in all stages of my research. Without your time and effort this study would not have been possible.

Thank you to all of the friends and family who supported me during the last four years. I will be forever grateful for your encouragement, kind words, and belief in me. To my parents, Ralph and Marie Thompson, thank you for instilling in me the self-belief I needed to finish this thesis, and for your love and support from afar. Thank you to my brother Dr James Thompson for getting there first and imparting your wisdom throughout my PhD experience.

Finally, to my husband, Andrew Lee, thank you for your constant love and understanding both in my PhD journey and in life. Parts of the last year have been the hardest of my life and you have helped me use those experiences to become stronger and more determined.

Author’s declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. All data collection and analysis were carried out by the author, Sophie Thompson-Lee.

Chapter 1: Introduction

1.1 The research context

Explicit grammar instruction and its role in second language (L2) learning is an area of sustained focus in applied linguistics research (for reviews, see Ellis, 2008; Norris & Ortega, 2000; Spada & Tomita, 2010). The issues are complex, and the findings are mixed. One line of research has suggested that instruction which focuses learners' attention on grammatical form is beneficial, and that instruction that connects form and meaning is best for learning (Benati, 2001, 2005; Cadierno, 1995; Cheng, 2003; Comer & deBenedette, 2011; VanPatten & Cadierno, 1993; Wong & Ito, 2018). Furthermore, research has investigated the role of grammatical rules, examples and feedback. Numerous studies have found that grammatical rules (also known as explicit information [EI]) plus practice is more beneficial than practice alone (e.g. Alanen, 1995; Fernández, 2008; Robinson, 1995), and a number of studies have also found that the addition of feedback increases learning outcomes (e.g. Dracos, 2012; Henry, 2015).

Another question of great interest in L2 learning research is the role that prediction plays in learning. Online processing research has found that learners make real-time decisions about role assignment whilst processing input (e.g. Jackson, 2008; Jackson & Roberts, 2010; Juffs & Harrington, 1995, 1996; Roberts & Felser, 2011). The brain has been described as a prediction machine (Clark, 2013) and when prediction fails, due to unexpected input, learning can occur (Ellis, 2016). This is thought to be because new representations are formed in the brain as the result of surprising language input (Zarcone et al., 2016).

Prediction in input processing is thought to be driven by sensitivity to lexical and morphosyntactic cues (Ellis, 2006). Research has shown that first language (L1) users utilise morphosyntactic cues to help assign both meaning and roles, and that cue sensitivity plays a faciliatory role in this process (DeLong, Urbach, & Kutas, 2005; Kamide, Scheepers, & Altmann, 2003). Native speakers (NSs) have also been found to adapt their processing mechanisms as a result of unexpected input and changes in the reliability of cues (e.g. DeLong, Quante, & Kutas, 2014; Fine et al., 2010; Fine & Jaeger, 2013; Hopp, 2016, 2017; Lew-Williams & Fernald, 2010). L2 learners have also been found to use cues to facilitate learning, but that the extent to which this is the case is dependent on proficiency (Dussias et al., 2013; Jackson, 2008; Jackson & van Hell, 2011), task used (Osterhout et al., 2006; Roberts, 2013), and L1-L2 similarities (Roberts & Liszka, 2019; Sabourin & Stowe, 2008; Tokowicz & Warren, 2010).

Building on evidence that L2 learners can use cues in a facilitatory way, research has begun to investigate whether prediction can be used as a tool for learning, by teaching learners to predict and as a result facilitate processing and / or aid learning (Andringa & Curcic, 2015; Hopp, 2016). Andringa and Curcic (2015) found that providing information and training about predictive cues did not result in prediction evidenced by an eye-tracking task, but that information about predictive cues aided offline judgements of grammaticality. Andringa and Curcic (2015) suggested this may be because the training was the first exposure the learners had of the target feature (direct object marking in an artificial language). Hopp (2016) found that mastery of a feature was necessary to result in cues being used predictively. These studies suggest that prior exposure, or proficiency, may have an effect on learners’ ability to use cues predictively. The current study builds on this research to investigate whether learners who have already been exposed to a feature, but make comprehension and production errors, benefit from training focusing on morphosyntactic cues. The training aimed to increase sensitivity to morphosyntactic cues in order to assign roles in the English passive voice and to investigate its effect on online processing, as well as offline comprehension and production. The current study sought to answer the following broad research questions:

RQ1) To what extent do L1 Chinese learners of L2 English show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

RQ2) To what extent do L1 Chinese learners of L2 English show knowledge of cues in a) offline grammaticality judgements and b) production of the passive voice in English?

RQ3) To what extent do native English speakers show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

1.2 The educational context: China and UK

In addition to aiming to improve our understanding of learning theory, online processing, and instructional effectiveness (as summarised above), there was an ambition to do so for a particular group of learners within a specific educational context. This context, in part, determined the choice of linguistic focus (passive voice), and so I describe the context very briefly here, although it was not a key motivation for the design or research questions. According to the UK Council for International Affairs (2019) in the academic year 2016/17 Chinese students exceeded any other non-UK nationality studying at a UK university (roughly 95,090 Chinese students) and this number had risen by 14% since 2012/13. By definition, these Chinese students must study in English, and this presents specific challenges. Although English is often the language used in Chinese academic work, the style of academic writing differs from that used in UK universities. This is as a result of differences in ideas synthesis and argumentation between English and Chinese academia (Liao & Chen, 2009). In Chinese academic writing analogy, proverbs, and phrasing are important, whereas presenting a balanced argument and refuting opposing opinions is less so. This means that Chinese learners often struggle with rebuttal and addressing others' arguments, which is a key part of argumentation in English academic writing (Liu, 2005). Additionally, the passive voice (among other features, such as reporting verbs and signposting) is a key characteristic of English academic writing (Bailey, 2015) which can prove problematic for Chinese learners of English. It is, therefore of some practical importance to determine whether grammatical instruction can improve both comprehension and production of this structure in upper-intermediate learners in the UK university context.

1.3 Outline of the thesis

Chapter 2 details and evaluates relevant literature in L1 and L2 processing and L2 learning leading on to the processing problem posed by the English passive voice. This will be divided into five subsections: research into L1 processing; research into L2 processing; research into explicit grammar instruction; approaches to grammar instruction; and the processing problem. The rationale for the current study and the research questions will be presented following the review of the literature. Chapter 3 outlines the methodology and procedure for the main study including findings from an exploratory study carried out prior to the final study. Chapter 4 presents the results in three main sections: performance during the interventions; results of the eye-tracking; and results of the offline tests. Chapter 5 discusses these results in relation to the research questions: sensitivity to morphosyntactic cues during online processing; the effects of training on online processing and offline production and grammaticality judgement; and, sensitivity of NSs to morphosyntactic cues during online processing (this includes a discussion of the effects of animacy). Finally, chapter 6 discusses the limitations of the study, its contributions to research into processing and instruction, and its potential relevance for foreign language teaching.

Chapter 2: Literature review

The literature review provides the context and rationale for the current study. The current study aimed to teach morphosyntactic cue use to L2 English learners. NSs have been found to use these cues during processing and it is these findings that inform research into L2 learners' cue use. In the following sections (section 2.1), research investigating how NSs process input, and to what extent they use cues to do so, is outlined. This serves two purposes, to provide context for the NS results in the current study, and to provide rationale for investigating cue use by L2 learners. In this chapter, key research into NSs morphosyntactic cue use, followed by research investigating L2 learners cue use, is detailed. After discussing the way in which NSs and L2 speakers process language, the following sections detail the theories and mechanisms that underpin L2 learning, specifically relating to the learning of morphosyntax. Keeping in mind the way in which learners learn language, and in relation to the types of knowledge learned, grammar instruction and teaching are then discussed in detail. The areas mentioned above provide the theoretical background and rationale for investigating L2 processing and learning of morphosyntactic cues. Finally, the processing problem that was investigated by the current study is described in detail; the passive voice in English.

2.1. First language processing

Linguistic and semantic cues in the input provide information about upcoming language and facilitate language processing. Processing is the interpretation of input, such as the assignation of thematic roles and relations between elements in the sentence. Cues include real-world knowledge, word order and sentence structure, noun animacy, and morphosyntactic cues. The current study does not set out to test a specific theory related to cue use, such as MacWhinney's unified competition model (MacWhinney, 2005). However, a brief discussion of cue use during sentence processing is useful in order to describe the linguistic focus of the study. The next sections outline briefly how native language users utilise cues such as word order and animacy during language processing and discusses research into grammatical cues in language processing in some detail.

2.1.1 Assigning thematic roles using cues in sentence processing: Word order and animacy

In sentences with a subject and direct object, the roles denoted are the agent and patient. The agent is defined as "identifying an actor or actors performing some action" and the patient "identifies an individual undergoing some process or targeted by some action" (Concise Oxford Dictionary of Linguistics, 2014, https://www.oxfordreference.com). Word order and animacy can both be cues to agent-patient roles. However, the strength with which they are relied on varies across languages. In English, the most reliable cue for agency is preverbal positioning. For example, in "*the dog chases the cat*", *the dog*'s preverbal position makes it likely to be the agent. Indeed, various studies have found that word order is the strongest (though not only) cue to agency for English speakers, whereas other languages use markers more so than word order to determine agency (MacWhinney, Bates, & Kliegl, 1984). For example, in Spanish, word order is flexible and so an unreliable cue to agency, but the preposition *a* (direct object marking) is reliable.

Direct object marking is an example of a morphosyntactic cue and is discussed again in section 2.1.2. Since in some languages, such as Spanish, the direct object can change position in a sentence, it is mentioned here in relation to the effect of word order on role assignment. In Spanish, *a* precedes the object e.g. *el perro persigue al gato* (the dog chases (to) the cat), and so this sentence could also be rearranged to *Persigue al gato el perro* or *persigue el perro al gato*, where *al* tells us which noun is the patient (*el gato*) and which is the agent (*el perro*). In German, agency is marked on the masculine determiner (m), and not the feminine (f) or neuter (n) – *der* (m), *die* (f) and *das* (n) are nominative and *den* (m) *die* (f) *das* (n) are accusative. For example, in *Der Hund jagt die Katze* (the dog chases the cat), *Der* is a reliable cue to *the dog* being the agent. The feminine and neuter determiners are less reliable cues since they do not change form in the nominative and accusative cases. Cues not only vary across languages but also within a language. For example, in Spanish, word order, verb inflections, world knowledge, discourse context, and animacy are all used to varying degrees as cues to interpret role assignment. The determiner *a* is not used for inanimate objects, such as *el perro persigue el coche* (the dog chases the car), and direct object marking is not used with animate patients of ditransitive verbs (Primus, 2010). In contrast, some languages do use direct object marking to clarify agency in ditransitive sentences with two agents e.g. Punjabi. Thus, processing language requires the interpretation of cues which are not ‘reliable’ (i.e. they are used only in specific linguistic contexts).

Animacy often serves as a cue to agency. Animate nouns share characteristics that are usually applied to an agent. According Barker and Dowty (1993) and Primus (2010) the prototypical agent role has the following characteristics:

The agent:	*acts by their own volition*
	is sentient of, or perceives another participant or entity
	causes an event or change in another participant or entity

moves autonomously

is a possessor of another entity

These characteristics of prototypical agents lend themselves most to having an animate participant involved. This is not exclusively the case, of course. Role overlap (agent – patient) can occur "whenever a noun phrase subcategorized for patient properties acquires potential agent properties due to its intrinsic meaning" (Primus, 2010, p. 73). However, typically agents are capable of producing an action or state through their own volition and therefore typically are animate. Hence, regardless of whether a language has or doesn't have direct object marking, animacy is also often a cue to agency. In English, animacy has been found to be an important cue to agency, second only to word order (Liu, Bates, & Li, 1992).

However, the reliability of the animacy cue (in languages such as English and Spanish) is poor when non-prototypical word order is used, for instance, in the passive voice in English. The sentence *the dog is chased by the car* is ambiguous until the verb inflection is reached if one relies on word order and animacy, as *the dog* would be assumed to be the agent. Research has found that when given implausible passive sentences, NSs of English tend to rely on animacy and word order and misinterpret agent roles (Ferreira, 2003). In reduced relative clauses containing a passive, processing was found to be quicker when the initial noun phrase was inanimate e.g. "*The evidence examined by the lawyer…"* versus "*The witness examined by the lawyer*…" (as found by Ferreira & Clifton, 1986, p.366; Frazier & Rayner, 1982). This suggests that the inanimate first noun (in the first sentence) was interpreted as being the patient, whereas the animate first noun (in the second sentence) was interpreted as the agent, therefore causing processing difficulty upon encountering the past participle of the verb and the post-verbal preposition *by*.

To summarise, word order and animacy are cues to agency in many languages, in some cases in combination with direct object marking. These cues compete or are strengthened when multiple cues agree or disagree (Li, & MacWhinney, 2012; MacWhinney 2005). In circumstances where word order, direct object marking, or animacy are not reliable for a parser, such as in the following, *the cat is chased by the dog*, morphosyntactic cues must be used to determine agency. In this case, the past participle verb ending *-ed* and the preposition *by* are the most reliable cues to agency. This is a critical feature of the current study.

2.1.2 Morphosyntactic and lexical cue use, or Good Enough processing?

Research has shown that NSs use morphosyntactic cues to help assign both meaning and roles, and that cue sensitivity plays a faciliatory role in this process. A theory of input processing commonly held is that input is processed on a word-by-word basis with the application of relevant constraints during analysis of input (Kamide et al., 2003). Research has sought to investigate word-by-word analysis of a sentence and whether such analysis can lead to thematic role assignment before the point in the linguistic input at which roles are disambiguated. In other words, this research has investigated whether L1 parsers use morphosyntactic cues to predict roles prior to their resolution in input. The following studies provide evidence for this being the case.

2.1.2.1 Anticipation facilitates L1 processing

One of the first studies to investigate prediction in language processing was by Altmann and Kamide (1999). They showed English NSs visual scenes depicting, for example, a boy, a cake, and some toys, while they listened to sentences such as "*The boy will move the cake*" or "*The boy will eat the cake*" (ibid, p.250). They found that eye-movements to the cake (the only edible object in the scene) started significantly earlier upon hearing *eat* than when they heard *move*. This was seen as evidence that semantic information conveyed by a verb can be used to anticipate an upcoming theme. In a similar experiment Altmann and Kamide (2007) found that when presented with sentences such as "*the man will drink all of…"*or *"The man has drunk all of…",* and images of full or empty glasses, native English speakers looked more to the full glass on hearing "*will drink"* and more to the empty glass on hearing "*has drunk"*(p.505). This study provided further evidence that activation of representations during language processing does not rely only on lexical activation of individual items, but also on information about upcoming referents encoded in preceding lexical items, such as, verbs and their tense.

Much research into the use of cues predictively has investigated gender marking in languages in which nouns are classified by gender e.g. Spanish – masculine and feminine, German – masculine, feminine and neuter. Gender is of interest in research into the role of predictive cue use because it is often marked by determiners preceding nouns, such as *el* and *la* in Spanish. Native adults, as well as children as young as 28 months, have been found to use gender on determiners to pre-activate noun labels for visible objects (Lew-Williams & Fernald, 2007). In an eye-tracking study, native Spanish speaking children (three-year-olds) and adults were shown two images and asked, "*Encuentra el / la..."* (*Find the….*). The target image was gender matched to the article e.g. "*la…..pelota" (the ball)*, the second image in this case would be of a masculine noun that would be preceded by *el e.g. el zapato (the shoe)* (ibid, p.10). On some trials the gender was matched so that this cue could not be used i.e. both nouns were masculine, or both were feminine. Both the adults and children were faster at looking to the target when gender information was reliable (i.e. on differing gender trials). This study demonstrated that children as young as three use cues predictively. These results were replicated in a follow-up study using the same procedure and stimuli (Lew-Williams & Fernald, 2010).

In another eye-tracking study of gender, in French, NSs were found to look at the target image upon hearing the determiner *le* or *la*, prior to hearing the noun (Dahan et al., 2000). They were found to rely on phonological cues when the gender cue was absent. The researchers gave participants four images and asked them to "*cliquez sur…*" (click on…) (ibid, p.467). In some trials, *the* had gender marking i.e. *la* or *le*, in these trials the participants looked to the correct image upon hearing *la* or *le*. On some trials, *the* was neutrally marked using the plural *les* so gender information was not available. On these trials, upon hearing the start of the noun, the participants looked to the objects that had phonologically similar first sounds e.g. ***bout**ons* and ***bout**eilles* (buttons and bottles). That is, when possible, gender was the preferred cue over phonological cues. These findings were replicated by Van Heugten and Johnson (2011) in another eye-tracking study using a similar task to that used by Dahan et al. (2000).

Predictive cue use is not only evidenced by gender marking. Evidence has been found for the indefinite article in English (DeLong et al., 2005), case marking in German (Hopp, 2017) and voice marking in English (Kamide, Altmann, & Haywood, 2003b). In an event-related potential (ERP) experiment with English NSs, DeLong et al. (2005) found that when participants heard sentences ending in less expected article and noun combinations (as in example 1b below), a large N400 was recorded on both the article and the noun. N400s are a component of an electroencephalography (EEG) signal that in linguistic research indicates sensitivity to a lexical-semantic anomaly. Thus, the increased N400 on the article was argued to demonstrate that the expected article, *a* or *an*, was pre-activated based on the context of the sentence prior to hearing the article or the final noun. The N400 on the article showed that the participants were responding to the unusual article – noun combination upon hearing the article and before hearing the noun. So, they were anticipating article + noun based on hearing the context (e.g. *The day was breezy*) and the verb (e.g. *fly*).

1a. *“The day was breezy so the boy went outside to fly...a kite”* (expected)
1b. *“The day was breezy so the boy went outside to fly...an aeroplane”* (unexpected)
(p.1117, ibid.)

Less conclusive evidence of cue use compared to the studies described above (i.e. DeLong et al., 2005; Hopp, 2016) was found in another visual world eye-tracking study investigating the integration of semantic and syntactic constraints in German (Kamide et al., 2003a). Case marking in German was manipulated to see if the information carried by the first noun in each sentence would be enough to result in the prediction of the second noun in the sentence (as in example 2 below).

2a. Der Wolf tötet gleich den Hirsch.

TheNOM wolf kills soon theACC deer

'The wolf will soon kill the deer.'

2b. Den Wolf tötet gleich der Jäger.

TheACC wolf kills soon theNOM hunter

'The hunter will soon kill the wolf.' (Hopp, 2017, p.12).

The results were mixed and suggested that case marking was used as a predictive cue, but that real-world semantics may also have played a role. The same study was carried out with English NSs, and again results were mixed. The authors suggested that this may have been due to issues with their experimental design and stimuli. Specifically, that the images may have provided extra cues to agency since some of the referents depicted in the images were oriented towards images of logical antecedents e.g. the image of the wolf was oriented to face the deer and the hunter was facing away from the deer.

In order to address some of the methodological issues acknowledged in the previous study, a partial replication of Kamide et al. (2003a) was carried out by Hopp (2017). The same aural stimuli were used with German NSs (see 2 above for example stimuli). The results were more conclusive than those of the original study. The findings showed that NSs did look towards the agent in the OVS (object – verb – subject) (as in 2b) condition and the patient in the SVO (subject – verb – object) (as in 2a) condition. It appeared that the NS use both word order and case marking (grammatical role) to make predictions.

Huang et al. (2013) furthered the previous case marking research carried out by Kamide et al. (2003) in Chinese children and adults. In a visual world eye-tracking test, children and adults heard sentences in which the passive marker *bèi* appeared either prior to an expressed noun or its pronoun (e.g. *seal* or *it* as in example 3a and 3b below). The participants were also presented with three objects which the noun or pronoun referred to; the expressed item (seal), a probable agent (shark), a probable patient (fish). In 3a, a preference for first noun agency would mean the seal would be assumed to be the agent prior to hearing *bèi* which signifies that the following noun is in fact the patient in this sentence. In other words, the *seal* would be assumed to be *eating*, rather than being *eaten*, so upon hearing *bèi* role assignment would need to be reviewed. In 3b the role of the pronoun is not so easily assigned since the noun it refers to is ambiguous so the participants would reserve role judgement till after hearing *bèi*, so would be less likely to need to reassess role assignment upon hearing *bèi*.

3a. "Seal B*ÈI* it quickly eat

海 豹 (被) 它 很 快 就 吃 掉 了

The seal is quickly eaten by it"

3b. "It B*ÈI* seal quickly eat

它 被 海 豹 很 快 就 吃 掉 了

It is quickly eaten by the seal" (Huang et al., 2013, p. 589)

Both children and adults were found to mis-parse sentences in which *bèi* appeared after the referential noun (as in 3a above) as they assumed the first noun to be the agent. They did not have such issues interpreting sentences like 3b. These findings suggested that passives are easier to process in Chinese when the initial role assignment does not need to be reviewed. This study also suggested that Chinese speakers tend to rely on first noun agency when interpreting roles. The findings also provided evidence that children and adults process, and reassess input, incrementally.

2.1.2.2 Anticipation and the Given-New approach to language processing

A contrary view to the notion that L1 users employ cues predictively to aid processing is that of the Good Enough Processing approach (Ferreira, Bailey & Ferraro, 2002). This approach is based on the tendency for L1 users to produce "superficial and even inaccurate interpretations" of language input (Ferreira et al., 2002, p.217). Evidence for this approach comes from NSs' comprehension of garden path sentences. For example, the following sentence was given to NSs of English, "*While Anna bathed the baby played in the crib"* (ibid, p.220). They were then asked two questions; "*did Anna bathe the baby?"* and "*did the baby play in the crib*?". For the second question, accuracy (correct answer 'yes') was almost 100%, but for the first question (correct answer 'no') it was only 50-60%. It was suggested that this is because the internal parsing system must reassess the parse upon hearing the word *played,* since *the baby* was assumed to be the patient of the first clause i.e. *Anna bathed*… when *the baby* was in fact the agent in a subordinate clause i.e. *the baby played*…. The inability to answer the question referring to the subordinate clause was explained as the language processing system failing to review its original interpretation and reassess the original mis-parse. These results were replicated using the same garden path sentences by Christianson et al. (2001) and Patson et al. (2009).

More evidence for "Good Enough" processing has come from the processing of implausible sentences. Ferreira and Stacey (2000) found that native English speakers were more likely to misinterpret simple passive sentences when syntactic structure was inconsistent with real-world knowledge, as in *The cat was chased by the mouse*. Participants were asked to judge the sentences they heard as plausible or implausible. The participants were highly accurate with active sentences (whether plausible or implausible), but for passives, they were less accurate for the implausible sentences, reporting that they were plausible 25% of the time. This suggests that semantics were used to process these types of sentence rather than syntax, and that the morphosyntactic cues present in the passive voice were ignored.

The "Good Enough" approach has been developed in recent years to incorporate Haviland and Clark's (1974) Given-New Strategy of language processing (Ferreira & Lowder, 2016). The Given-New Strategy holds that the primary purpose of language is to comprehend *new* information. The internal parser addresses the *given* (already known) information in any sentence first so it can be used to access previous representations, and secondly, to identify the *new* information and integrate it with the previously stored information (given). Difficulty arises when the *given* information is harder to access and when *new* information is harder to integrate.

Ferreira & Lowder (2016) combine the Given-New approach with that of good enough processing and define prediction as a facilitator of the integration of new information. This approach suggests that information that is already known by the parser (given information) is processed in a shallow, good-enough way, and that for new information, prediction is used as a mechanism for integration. Prediction, therefore, is the assessment of new information based on "prior linguistic context, world knowledge, and other forms of stored knowledge" to generate the probability of various possible candidates for its correct interpretation (ibid, p. 239). This, more recent, conceptualisation of "good enough" processing does not dismiss the role of prediction and the cues used in its employment but conceptualises "good enough" processing as the superficial processing of *given* information (see figure 1 below). So, in the examples provided above, such as "*The cat was chased by the mouse*" the given information was provided by previous experience and real-world knowledge, resulting in shallow processing and the ignoring of the passive morphosyntactic cues.

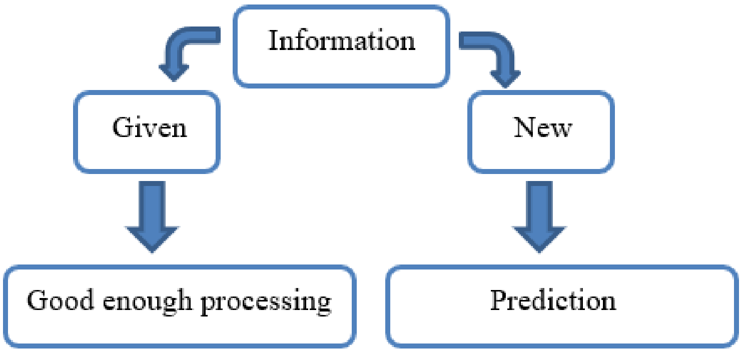


Figure 1. Given-New approach to language processing (Ferreira & Lowder, 2016)

The research reviewed above provides evidence for the predictive use of cues by NSs. The exact role that prediction plays is debated, for example how it facilitates processing and whether it plays a part in learning (as is discussed in section 2.3.2 with regard to L2 learning). The following section outlines research comparing language processing by NSs and non-native speakers, and the extent to which non-native speakers use cues predictively.

2.2 Second language processing

A question of great debate and research effort is whether L2 learners, or non-native speakers, are able to process language input in the same way as NSs. Evidence from online processing studies has shown that L2 learners incrementally process language input, making real-time processing commitments during comprehension (e.g., Jackson, 2008; Jackson & Roberts, 2010; Juffs & Harrington, 1995; 1996; Roberts & Felser, 2011). The information used to facilitate this may be that of the lexical and morphosyntactic cues used by NSs (see section 2.1), or it may be that learners use different processing strategies to incrementally parse input. One theory that posits that learners may not use lexical and morphosyntactic cues to facilitate processing is that of the First Noun Principle (VanPatten et al., 2013).

The First Noun Principle presupposes that learners "tend to interpret the first (pro)noun they encounter in an utterance as the subject / agent" (ibid, p. 508). This processing strategy may lead to mis-parsing of sentences in which the first noun is not the agent (e.g. English passives) and may cause learners to ignore other cues (such as case marking). This is one of the key components of VanPatten's input processing theory and the pedagogical approach associated with it, processing instruction, which is discussed in the context of L2 instruction in section 2.4. Previous studies have found evidence for the First Noun Principle (e.g. Isabelli, 2008; Lee, 2015; Lee & Malovrh, 2009; VanPatten et al., 2013; VanPatten & Houston, 1998), some research suggests that the First Noun Principle might play a role mostly in early L2 learning (Ervin-Tripp, 1974). A study into L2 learners of Spanish, found that First Noun Principle only applied to sentences when two nouns were present, in sentences with only one, learners assumed the first noun to be the patient (object) rather than the agent (Tight, 2012). This suggests that the First Noun Principle may not be as universal as it was originally claimed.

More recent research has employed online techniques to investigate how L2 learners and L1 users process language input, such as event related potentials (ERP), eye-tracking, self-paced reading (SPR), and EEG. Much of this research has focused on if and how learners use the morphosyntactic and lexical cues described above. A number of studies have found that learners can come to process L2 structures in nativelike ways (e.g., Dussias et al., 2013; Osterhout et al., 2006, 2008; Sabourin & Stowe, 2008; Tokowicz & Warren, 2010). Whereas other studies have found mixed results or have found that learners do not use cues at all. These are now reviewed.

2.2.1 Studies investigating syntactic cue use

Hopp (2017) partially adapted Kamide et al.'s (2003a) experiment one (see section 2.1.2). This study investigated whether L2 learners of German (English L1) integrated morphosyntactic and lexical-semantic information in predictive processing, specifically case marking on determiners signifying nominative or accusative nouns, for example:

4a. *Der Hase frisst gleich den Kohl.*

TheNOM hare eats soon theACC cabbage

'The hare will soon eat the cabbage.

4b. *Den Hasen frisst gleich der Fuchs*.

TheACC hare eats soon theNOM fox

‘The fox will soon eat the hare.’ (Kamide et al., 2003, p.41).

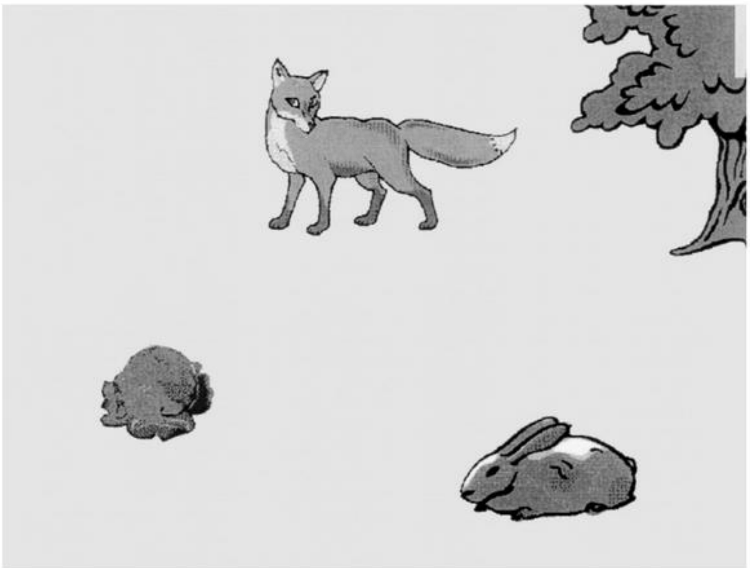


Figure 2. Example visual stimuli from Kamide et al. (2003a) and Hopp (2017)

The stimuli were taken from Kamide at al. (2003a), but some changes were made in response to possible limitations suggested by Kamide et al. For example, the positioning of some images was changed to prevent the subject of a sentence from facing towards the appropriate target on all trials. NSs of German and learners of German were tested. The participants saw an image containing various objects and heard sentences like those in example 4 above. Their eye-movements were tracked to see which object they looked at most upon hearing the case marked determiners.

The findings showed differences between the NSs and learners. The NSs looked towards the agent (S or O depending on the case marking) in the OVS condition (example 4b above) and the patient (S or O depending on the case marking) in the SVO condition (example 4a above) before hearing the second noun. In other words, they anticipated the second noun based on the case-marked determiner. The learners looked to the patient in both conditions before hearing the second noun. In other words, upon hearing "soon" the learners looked to the item which was semantically most likely to be the patient (the cabbage in figure 2). This was the case even for items in which the word order was OVS and the case marking of the first noun should have resulted in anticipatory looks to a subject (the hare in figure 2). This shows that the learners were making anticipatory looks, but that they did not use morphosyntactic information, i.e. case marking, as the NSs did. The findings for NSs were different to that of the original study, Kamide et al. (2003), which only found anticipatory looks towards the patient in the SVO conditions. However, Kamide et al. suggested that this finding may have been due to issues with the images (resolved by Hopp, 2017) (agent facing the patient etc.) which biased the NSs' interpretation.

Another study which found differences between NSs' and non-native speakers' cue use used ERPs to investigate lexical prediction in English NSs and Spanish-English bilinguals reading in English (Martin et al., 2013). They presented the participants with nouns which would be expected or unexpected according to the sentence context, these were either preceded by *a* or *an*. The unexpected condition being that the article given did not match a predicted noun. So, in the following example ''*She has a nice voice and always wanted to be an artist*'' (Martin et al., 2013, p.576), the *an* would be unexpected since the participants would probably be expecting *a singer*. They found that adult L2 learners did not exhibit prediction effects based on the sentence context. The NSs did, consistent with previous research into the use of *a* / *an* as a predictor in English language processing (e.g. DeLong et al., 2005). This difference between the L1 and L2 participants may have been due to the lack of this feature (allomorph of the indefinite article) in Spanish, i.e., L1-L2 cross-linguistic differences. When an L1 and L2 share certain features, research has found that L2 users show prediction effects, albeit to a lesser extent than their NS counterparts.

As noted for L1 processing research, gender marking is a language feature than lends itself to studies investigating anticipation and morphosyntactic cues. Since gender marking is a feature of some languages, but not all, examining learners usage of gender in an L2, in particular when their L1 does not have this feature, allows researchers to compare L2 cue use with L1 cue use. One such study to address this issue of cross-linguistic difference did so by manipulating the expectancy of nouns preceded by a feature that exists in both the L1 and L2, in this case gender marking in French and Spanish (Foucart et al., 2014). Spanish NSs and non-native speakers (L1 French) were tested whilst reading Spanish sentences (see example 5 below). The sentences differed in the expectancy of the noun phrases.

Expected:

5a. *El pirata tenía el mapa secreto, pero nunca encontró el tesoro [masc] que buscaba.*

[The pirate had the secret map, but he never found the treasure he was looking for.]

Unexpected:

5b. *El pirata tenía el mapa secreto, pero nunca encontró la gruta [fem] que buscaba.*

[The pirate had the secret map, but never found the cave he was looking for.]

(Foucart et al., 2014, p.1464)

Both groups were found to show significantly greater anticipation effects when given sentences with expected noun phrases (as in 5a), as evidenced by the N400 effects during ERP recording. The ERPs were analysed on the preceding article and the critical noun.

Foucart et al. (2014) concluded that the bilinguals relied on both anticipation and integration to process the input, as also found for the NSs. In follow-up studies, Foucart, Ruiz-Tada, & Costa (2015, 2016) investigated whether the anticipatory effects found in Foucart et al. (2014) would be replicated with auditory input. They also sought to test whether the effects found could be attributed to the context prior to the article preceding the noun phrase, or whether the article itself resulted in the predictive effects observed. In order to test this, the same stimuli were used as in the 2014 study, but the noun was blanked out, as in example 6 below. In the 2015 study, native Spanish speakers heard the Spanish sentences and ERP recordings were taken. In the 2016 study the same procedure was followed using French-Spanish late bilinguals. In both studies it was found that expected articles had a facilitatory effect of processing, confirming the findings of the 2014 study.

6. *El pirata tenía el mapa secreto, pero nunca encontró la XXX [fem] que buscaba*

Foucart et al. (2015, 2016) also investigated the effects of anticipation on recall by using a lexical recognition task after listening. Participants were asked to recall which nouns they had heard during the listening phase. Participants in both studies incorrectly believed they had heard expected words more than unexpected words. In other words, the expected articles created a memory trace which made participants feel that they had heard expected nouns even when this was not the case. This finding suggested that lexical pre-activation is taking place in order to speed up processing. The similar findings from NSs of Spanish (Foucart et al., 2015) and in French speaking learners of Spanish (Foucart et al., 2016) suggested that similar processing was taking place in NSs and non-native speakers.

The experiments by Foucart et al. found that L1 and L2 processing can be similar with regard to the use of prediction and integration. Not all research has found this to be the case. Lew-Williams and Fernald (2010) used three eye-tracking experiments to investigate gendered articles in Spanish. In their first experiment, English-Spanish bilinguals were shown images corresponding to audible sentences. On hearing the article (*el* or *la*) in each sentence, whether or not their gaze moved to the correct image as a result of hearing the gendered article was observed. Half of the trials contained the same gender (i.e. two images with the same gendered nouns) and half different gendered nouns. The learners' reaction times (RT) were also recorded. RTs, from hearing the article to the gaze moving to the image, showed that the native adults and children used the articles' gender to look at the correct image, whereas L2 Spanish learners did not.

The second experiment investigated the possible effect of the preceding article on prediction. Trials contained unknown objects and their nouns preceded by definite articles. Learners were first exposed to the novel objects and their labels. As in experiment one, participants then listened to test trials in which they heard sentences containing the newly trained words and were asked to choose the correct images from two possibilities. Eye movements and RTs were recorded to determine if the article was used to predict the correct images. It was found that both the NSs and learners used gender marking to predict the correct image on different gender trials more so than on same gender trials.

Lew-Williams, & Fernald (2010) pointed out that in reality learners are exposed to new nouns preceded by numerous determiners, not only definite articles. To investigate whether prediction learnt with one type of determiner is generalisable to other determiners, experiment three exposed participants to a teaching phase using indefinite articles (*un / una*) and followed this with the same test phase as experiment two using definite articles (*el / la*). In this experiment, whereas L2 learners did not use gender marking to predict the correct image, L1 users did. Furthermore, on gender different trials, when the article was informative, L2 users did not respond quicker than on gender same trials, as shown by RTs. These experiments suggest differences in the processing of gender markers between L1 and L2 users, except when exposure is very controlled, as in experiment two, in which the differences were not so great. This study's findings are in contrast to those of Foucart et al. (2014, 2015, 2016), in which NSs and non-native speakers were found to process in similar ways.

This difference in findings could perhaps be attributed to the difference in task type. Foucart et al.'s experiments all recorded the effect of unexpected words based on context, whereas Lew-Williams & Fernald (2010) investigated the effects of article morphology. Perhaps prediction based on sentential context as well as article morphology (as found in Foucart et al. 2015, 2016) is used more easily in L2 processing than that of the article in isolation (i.e. without context). In Lew-Williams & Fernald (2010) for instance, sentences such as "*find the ball / cookie*" relied only on the article for prediction (Lew-Williams & Fernald, 2010, p.462). Since Foucart et al. found that the preceding context was used by both L1 and L2 users to predict the upcoming article and noun, perhaps Lew-Williams & Fernald's findings suggest that the article alone is not used in L2 processing. Another explanation is that the difference between English and Spanish is greater than that between French and Spanish in this circumstance, as gender is not a feature of English. Research has found that if a learner's L1 encodes grammatical gender, they can come to use gender marking predictively at intermediate proficiency levels in their L2 (Dussias et al., 2013; Morales et al., 2015).

The above studies found that sensitivity to gender marking, and its use predictively, was limited by the presence of gender in the learner's L1. In contrast, Hopp (2013) found that native-like predictive gender processing is possible even for speakers whose L1 does not contain grammatical gender. Advanced English-speaking learners of German were found to use gender predictively in an eye-tracking test. It was suggested that this might be partially due to the explicit nature of the task used; the learners were told where to look as part of the task by asking questions, such as, "*where is the yellow [noun]*?" (ibid, p. 40). L2 processing appears to be more native-like in tasks involving explicit or metalinguistic knowledge rather than those that only involve comprehension (Roberts, 2013). Proficiency has also been found to affect the extent to which learners are able to use cues during processing (e.g. Jackson, 2008; Jackson & Van Hell, 2011). The learners in Hopp (2017) were advanced learners so this might explain their ability to use cues that are not present in their L1.

The effect of L1-L2 similarity has not only been found for gender marking. A number of studies into various language features have found that when the L1 and L2 are different learners are less sensitive to the associated cues. For example, grammatical inconsistencies appear to be picked up on by L2 learners only if the same morphosyntax was present in their L1 (as found by Sabourin & Stowe, 2008; Tokowicz & Warren, 2010; Osterhout et al., 2006). Differences in both online and offline performance based on L1-L2 similarity or difference have been found in several studies (e.g. Grüter, Lau, & Ling, 2020; Hawkins & Liszka, 2003; Roberts, Gullberg, & Indefrey, 2008; Roberts & Liszka, 2013, 2019; Tokowicz & MacWhinney, 2005).

The studies outlined above suggest a mixed picture for L2 processing. Non-native speakers appear to use morphosyntactic cues to process L2 input, but the extent to which they are able to do this may depend on proficiency, i.e. higher proficiency learners are better able to use cues (Jackson, 2008; Jackson & Van Hell, 2011), the facilitatory effect of L1-L2 similarity, and the type of words or grammar feature being parsed.

The research above raises the question: If learners do use cues during L2 processing, what role do these cues play in learning and, by extension, teaching? The next section focues on how cue use and prediction may play a role in language learning.

2.3 Second language (L2) learning of English morphosyntax

2.3.1 Processing and its role in learning

Language processing is thought to play an active role in L2 learning. The way in which this role is operationalised depends on whether processing and learning are thought to be separate (Pienemann, 2010) or interdependent systems (O'Grady, 2005). One way in which processing and learning are interdependent is the idea that language structure is developed by the processing system to put as little strain on working-memory as possible. In other words, language structure, morphology and syntax, develop to allow working memory to work as efficiently as possible (O'Grady, 2005). Another way in which processing and learning are thought to be linked is surprisal-driven processing which is driven by prediction (in this thesis, the terms prediction, expectation and anticipation are used interchangeably). Surprisal-driven processing is explained in the following section.

2.3.2 Expectation and Surprisal

Building on the idea that processing is integral to acquisition, recent research has focused on surprisal-driven language processing and acquisition, both in an L1 and L2 context. Surprisal-driven language acquisition theories posit that learning is driven by prediction errors. In other words, surprising language input, which does not meet expectations, is more likely to be learned than expected input (Clark, 2013; Rumelhart, Hinton, & Williams, 1986; Wagner & Rescorla, 1972; Wills, 2009). This is because surprisal results in new representations being formed in the brain (Zarcone et al., 2016). This has been demonstrated in structural priming studies, in the L1 and L2, in which speakers have been observed to use a new or rarer structure more when their interlocutor recently used it (e.g. Jaeger & Snider, 2013). Prediction errors are also thought to be a mechanism for pre-empting overgeneralisation. For instance, a learner encountering an English irregular past participle for the first time after learning the regular verb forms first, as is typical of many beginner English language classes. Upon hearing "*yesterday I went to the shop and* …." a learner unaware of irregular verb forms would likely be expecting "*buyed*". Upon hearing the unexpected irregular past form "*bought*" a new representation would be formed as a result of the prediction error (Mitchell, Myles, & Marsden, 2019). Thus, the ability to predict might be an important mechanism for learning an L2. That is, if learners can use language predictively then they may be able to learn from surprisal and / or prediction errors.

In contrast, the RAGE (reduced ability to generate expectations) hypothesis posited that L2 learners cannot use cues to predict, regardless of prediction's hypothesised role in L1 processing or learning (Grüter & Rohde, 2013). This hypothesis suggested that L2 learners do not have the cognitive resources available for prediction during real-time processing.

Word retrieval and integration use most of the mental resources available to a learner resulting in “little or no resources … for taking up non-essential cues to update expectations” (Grüter, Rohde, & Schafer 2014, p.189). Research supporting the RAGE hypothesis comes from visual world eye-tracking studies such as Lew-Williams and Fernald (2010). Adult learners were found to be unable to make predictions using semantic and morphosyntactic information, in this case gender marking on determiners in Spanish to predict an upcoming noun from a visual scene. Other studies have found similar results e.g. Grüter, Lew-Williams and Fernald, (2012), Hopp (2015, 2016) and Martin et al. (2013). Evidence contrary to RAGE hypothesis has been found in research finding evidence of semantic prediction (Dijkgraaf, Hartsuiker, & Duyck, 2017; Ito, Pickering, & Corley, 2018) and morphosyntactic prediction (Dussias et al., 2013; Hopp, 2013; Hopp & Lemmerth, 2018; Trenkic, Mirkovic, & Altmann, 2014) in L2 learners and bilinguals.

Another view of the role of prediction in learning is that anticipation is a by-product of learning, not its driver (Phillips & Ehrenhofer, 2015). For a learner to be able to make predictions about input they need to have enough prior knowledge, and representations, to generate those predictions. This may not be possible for (particularly beginner) learners (ibid). According to this view, prediction is available to both NSs and L2 learners, but it is prior experience that moderates its role (Foucart et al., 2015). For example, cross-linguistic differences have been found to limit prediction in L2 sentence comprehension (Martin et al., 2013). In other words, L1 language experience appears to moderate the ability to make predictions in the L2. However, cross-linguistic difference tends to reduce as proficiency increases and sentence processing tends to become more native-like (e.g. Hopp, 2010; Ullman, 2005) this suggests that L2 learners may be able to use prediction in the same way as NSs as their proficiency increases (Foucart et al., 2014). Prediction therefore improves with experience, rather than being a fundamental mechanism driving L2 learning.

It has also been suggested that learning can take place without prediction. Evidence for this comes from studies in which L2 learners fail to predict during online processing but are shown to have the acquired the grammatical knowledge of the target feature (Kaan, Ballantyne, & Wijnen, 2015). In Martin et al. (2013), L2 learners were not able to predict whether a noun would be preceded by *a* or *an* in online processing, but when they were tested offline, in a cloze probability test, learners knew when to use *a* or *an* correctly. It has also been shown that learners may weigh possibility differently to NSs. In an eye-tracking study, Cowles and Wijnen (2015) showed that when given options to complete sentences, NSs look at all options before choosing the correct image. Non-native speakers were not found to entertain all the possible images. This was taken as evidence that native and non-native speakers assess probability differently, and this could result in prediction errors different to those experienced by NS. Therefore, learning as a result of prediction error may not result in nativelike language use (Kaan, Ballantyne, & Wijnen, 2015).

The above studies have not provided conclusive evidence about the role of prediction and surprisal in L2 learning. Whether prediction drives learning, is a result of learning, or perhaps both in some cases is currently unclear. In fact, it has been suggested that prediction has been treated in some ways “monolithically” in L2 research (Phillips & Ehrenhofer, 2015, p.422). In reality, it may be that learners can use some predictive information, but that they will find difficulty due to, for example, linguistic context, L1, or L2 experience (proficiency).

Of some (albeit limited) relevance to the current study is that, to the best of our knowledge, this line of research has not yet investigated whether NSs or L2 learners use a morphological cue (such as a verb inflection) to predict role assignment to an upcoming noun (i.e., whether it is the agent or patient of a sentence). It is emphasised that this is not a direct core theoretical motivation behind the current study, because in the current study, the inflectional verb morphology cue was also being used to assign a role to the *previous* noun (i.e., not predictively, but retrospectively). Nevertheless, the previous section is an important body of research to acknowledge in terms of addressing the role of sentence processing in learning.

2.3.3 Explicit and implicit knowledge and their role in acquisition

An important aspect of learning research is determining the nature of the knowledge being learned. In cognitive psychology and L2 learning, knowledge has been categorised as being explicit or implicit (Godfroid & Winke, 2015). Implicit knowledge is automatic and subconscious. In other words, the language user is unaware of the knowledge they are accessing during language production or processing (Ellis, 2009; Sanz & Leow, 2011). Implicit knowledge is part of a learner's linguistic competence and is described as "intuitive and tacit, rather than conscious and explicit" (R. Ellis, 2005, p. 143). It is thought to be less prone to decay over time.

Explicit knowledge is knowledge the language user is aware that they possess and can employ when making conscious decisions (Andringa & Rebuschat, 2015; DeKeyser, 2008; Dracos, 2012; Ellis, 2004; Hulstijn, 2005; Marinis, 2010; Rebuschat & Williams, 2012). Explicit knowledge is defined by Ellis (2004) as being part metalinguistic explanation and part analysed knowledge. Metalinguistic explanation is defined as "knowledge of grammatical metalanguage and the ability to understand explanations of rules" (ibid, p. 95). Analysed knowledge is a "conscious awareness of how a structural feature works" (ibid). Both explicit knowledge and implicit knowledge may be rapidly accessed to retrieve information about language, but the key difference between the two knowledge types is that accessing explicit knowledge requires awareness, whereas accessing implicit knowledge does not (Suzuki & DeKeyser, 2017).

One of the key areas of debate in L2 learning research is whether explicit knowledge and implicit knowledge are entirely distinct (non-interface hypothesis) or whether explicit knowledge can become or may serve implicit knowledge (strong-interface hypothesis and weak interface hypothesis). Proponents of the non-interface hypothesis believe that explicit knowledge is acquired in an entirely distinct way to implicit knowledge, and that one cannot become the other. Krashen (1982) proposed that acquisition of language (e.g. child acquisition of an L1) can only occur implicitly. So, learners are not aware that they are acquiring a language and are unaware of its grammar or rules. He also distinguished *acquisition* from *learning*. Learning was defined as gaining "knowledge about a language…grammar and rules" (ibid, p.10). Whereas acquisition is the subconscious development of knowledge. Therefore, the term learning would apply to gaining explicit knowledge and acquisition to gaining implicit knowledge. The non-interface hypothesis proposes that explicit knowledge gained through learning cannot become part of the linguistic competence and cannot be accessed spontaneously i.e. explicit knowledge cannot become implicit knowledge. Krashen (1982) also proposed that both L1 and L2 learners are able to acquire language through developing implicit knowledge as a result of naturalistic practice involving meaningful and spontaneous interactions (White et al., 1991). As a consequence, rules and focus on grammar would be of no benefit to L2 learning.

The strong-interface hypothesis takes the opposing view, that explicit knowledge can change in its nature and, after extensive practice, can either become implicit knowledge itself or something that is essentially indistinguishable from it (automatised knowledge) (Bialystok, 1994; DeKeyser, 2017; DeKeyser & Criado, 2012; R. Ellis, 2005; Norris & Ortega, 2000; Suzuki & DeKeyser, 2015; 2017). That is, repeated practice, such as meaningful interaction, can result in explicit knowledge becoming proceduralised and then automatised over time. Furthermore, explicit knowledge can be developed about a feature, that may be represented as implicit knowledge, if language rules and grammar are studied and metalinguistic knowledge is developed (such as, a NS who becomes aware of a rule or pattern in their own language).

A third view of the relationship between explicit knowledge and implicit knowledge is loosely referred to as the weak interface hypothesis. In fact, there are several views that could fit within this perspective. One is that explicit knowledge can become implicit knowledge after practice but only if a learner is in the correct developmental stage for the aspect of language being learned (R. Ellis, 2005). Another, of more relevance for the current study, is that explicit knowledge and implicit knowledge are “dissociable but cooperative” (N. C., Ellis, 2005, p. 305). In this view, “explicit learning involved in the initial registration of pattern recognizers for constructions … are then tuned and integrated into the system by implicit learning during subsequent input processing” (p.305). That is, explicit knowledge aids the development of implicit knowledge through drawing attention to form in input (R. Ellis, 2005; Schmidt, 1990; VanPatten, 2002).

Although measuring these different types of knowledge and the extent to which they can interact during learning were not main aims of the current thesis, the short review above is relevant to the current study for two main reasons. First, the basic notion that different knowledge types exist partly motivated the choices of measures of learning used in the current study. That is, broadly speaking, it was important to measure both knowledge that had to be drawn upon very fast, and, likely without awareness (e.g., when hearing input and looking at a visual scene) and knowledge where more time was available to consciously access it (e.g. when reading sentences and deciding if they are grammatical or not). Second, and relatedly, it was important to understand how instruction that conveys explicit knowledge, via rule explanation and orientation of learners' attention to form and function, may influence measures that likely tap into implicit kinds of knowledge. Of additional importance was the kind of knowledge that can be accessed given more time in offline measures. A key feature of the instruction designed in the current study was attention to form and function. The roles of attention and noticing are discussed in a little more depth in the next section.

2.3.4 Noticing and attention

An influential view in second language acquisition research over the last few decades has been that noticing a new feature (such as a phoneme, morpheme or syntactic structure) is essential, or at least useful, for learning these features (Schmidt, 1990; Schmidt, 2001). The noticing hypothesis stated that noticing cues in input results in their acquisition and that other less salient (albeit complex to define) cues are often not noticed sufficiently to be acquired. Over time, certain cues can come to be relied upon (whether due to, for example, their physical salience or their existence in languages that are already known) and others can become redundant (see Ellis, 2007; Ellis, & Sagarra, 2011). For example, in the case of the English past simple, the past verb form becomes redundant when used in combination with the temporal adverb "*yesterday*" e.g. *Yesterday, he played football.* It is thought that drawing learners' attention to less salient cues (or cues to which learners do not orient their attention for whatever reason) can aid their learning (N. C., Ellis, 2005). As Schmidt noted "since many features of L2 input are likely to be infrequent, non-salient, and communicatively redundant, intentionally focused attention may be a practical (though not theoretical) necessity for successful language learning" (2001 p. 23). So, explicit instruction, or focus on form, can aid the noticing of language features, and therefore learning if given alongside sufficient examples and practice (N. C., Ellis, 2005).

There are a number of theories about the role of attention and awareness in L2 learning. For example, the concept of consciousness raising, in which learners' attention is focused onto language form, was an earlier theory proposed by Sharwood-Smith (1981). It led to the idea of input enhancement which proposed that the visual presentation of language features, e.g., in bold or colour, could help learning (Sharwood-Smith, 1993). Other examples of orienting learners' attention to the form of language include: explicit explanation of linguistic features; metalinguistic description; negative evidence through overt error correction; and input flood.

There has been a great deal of research investigating attention and awareness and their roles in learning (e.g. DeKeyser, 1995; Grey, Williams, & Rebuschat, 2014; Hama & Leow, 2010; Leow, 2015; Leung & Williams, 2011; Robinson, 1997; Rosa & Leow, 2004; Williams, 2005; for reviews see: Andringa & Rebuschat, 2015; Leow & Donatelli, 2017;

Rebuschat et al., 2015; and for early theorising see: Tomlin & Villa, 1994). Attention and awareness are key aspects of explicit grammar instruction. The current study designed explicit instruction to focus learners' attention to the morphosyntax of the passive voice. The design of the current study was influenced by a large body of previous research into explicit instruction, the next sections detail the nature of explicit instruction in L2 learning (section 2.4) and, finally, demonstrate where the current study fits in this body of research (section 2.5).

2.4 Explicit grammar instruction

Explicit instruction is instruction in which "rule explanation comprise[s] part of the instruction" or "learners [are] directly asked to attend to particular forms and to try to arrive at metalinguistic generalizations on their own" (Norris & Ortega, 2000, p. 437). The following sections briefly describe various explicit instructional methods and a few key studies that have investigated their efficacy. The focus is on input-based instruction because, as reviewed above in section 2.3.4, noticing and attention to form and meaning are considered by many researchers to play an important role in language learning. Explicit instruction can, according to the research discussed below, promote and manipulate noticing, and the orientation of attention, to result in learning.

2.4.1 Different types of input-based explicit instruction

2.4.1.a) Input enhancement

Input enhancement is a method of instruction that is based upon the importance of noticing (Sharwood-Smith, 1993). Input enhancement involves drawing learners' attention to form through textual enhancement such as underlining, **emboldening**, different fonts and different colours of print (Balcom & Bouffard, 2015). Eye-tracking studies have examined how input enhancement might affect attention and L2 development when learning rule-based morphosyntactic forms (e.g. Indrarathne & Kormos, 2017; Issa & Morgan-Short, 2019; Loewen & Inceoglu, 2016; Simard & Foucambert, 2013; Winke, 2013). The results of studies investigating input enhancement and its effect on morphosyntactic forms have varied. On one hand, input enhancement appears to result in greater attention to target forms in some studies, as measured by increased total fixation time (Winke, 2013) and second-pass time (the summed fixation duration made when eye-movements orient to an area for the second time) (Simard & Foucambert, 2013; Winke, 2013), on the other hand some studies have not found a difference in attentional processing as a result of input enhancement compared to non-enhanced input (Indrarathne & Kormos, 2016; Loewen & Inceoglu, 2016). Enhancing text alone is not always enough to result in learning, even though attention is likely oriented to the target form (as found by Lee & Huang, 2008).

2.4.1.b) Input processing theory and processing instruction

The role of noticing in language learning was further developed in ‘input processing theory’ proposed by VanPatten (2004). Input processing theory posits that noticing alone is not enough for learning, but that noticing form and meaning, and making connections between the two, is essential for successful L2 acquisition (Benati, 2005). VanPatten’s notion of input processing focuses on the form-meaning connection and the mechanisms that promote it (VanPatten & Cadierno, 1993). According to VanPatten, these form-meaning connections can be made when there is “referential meaning” (but is less likely, or not possible, where a form does not have referential meaning, such as abstract grammatical gender or syntax). VanPatten and colleagues proposed that "explanations about how language works" are not considered *input* (VanPatten & Cadierno, 1993, p.46), though there is evidence that providing rules and grammatical explanation prior to hearing input can speed up the process of making form-meaning connections (see Culman, Henry & VanPatten, 2009). So, according to this view of input processing, noticing in itself is not enough to promote learning, but it is the establishment of form-meaning connections that results in acquisition.

VanPatten and Cadierno (1993, p.46) proposed the following schematic (figure 3 below) of L2 learning which illustrates the role of input and the resulting processes that aid acquisition.

Figure 3. Stages of acquisition (VanPatten & Cadierno, 1993, p.46).

Input is first converted into intake (1). Intake refers to the part of the input which is processed and held in working memory in order to restructure and develop the linguistic system (2). Following this, according to this model, restructuring and accommodation are employed (3), these are further processes concerned with language production and include "access, monitoring and control" (ibid, p.46).

Based on input processing theory and the model of L2 learning in figure 3, VanPatten developed a package of three types of activity which together was called 'processing instruction' (PI). This instruction started with a brief explanation of grammar followed by a) referential tasks which focus learners' attention on a form-meaning connection and b) affective tasks which contain multiple examples of the target form but do not require a form-meaning connection to complete.

According to VanPatten and Cadierno (p. 46), processing instruction aims to manipulate the first stage of acquisition (1 in the schematic above).VanPatten argued that, in contrast to processing instruction, so-called ‘traditional grammar instruction’ focuses on intervening in stage (3), typically by teaching grammatical rules to learners and then having them practice the taught form through ‘output’ e.g., communicative and mechanical production drills (VanPatten, 2000). For example, asking learners to transform an infinitive verb into the past tense in English (mechanical practice), or to complete the following sentence using the past tense – “*Last year my school*…..” (labelled ‘meaningful practice’) (Benati, 2005, p.89). VanPatten and Cadierno’s (1993) ‘traditional instruction’ included a structured role play and sentence transformations.

The three components of processing instruction (PI)

The key aim of PI is to “push learners to abandon their inefficient processing strategies for more optimal ones so that better form meaning connections are made” (Wong, 2004, p.35). This is achieved, according to VanPatten, using three stages. First, grammatical rules are given in the form of explicit information (EI) about a target grammatical structure, its meaning, and potential processing problems, as follows from VanPatten, Collopy & Qualin, 2012 (p. 271):

[screen 1]

In Russian, nominative and accusative case designate "who does what to whom". That is, case tells us which noun in the sentence is the "verb-er" and which is the "verb-ed".

Take the following sentence for example: "The girl sees the boy".
In this case, the girl is the one who sees (verb-er) and the boy is seen (verb-ed). The girl is the subject of the verb and the boy is the direct object.

To illustrate it's meaning, the grammatical feature is contrasted with another grammatical feature, for example:

[screen 2]

In Russian, one way to tell what is the subject and what is the direct object is by looking at the ending of the nouns. When masculine and feminine nouns are the subject, the noun stays the same. When a masculine animate noun is the direct object, the ending is -a (as in the example below) or – Я

Example: **Мальчик** идёт домой. "The **boy** goes home"

Девочка видит **мальчика**. "The girl sees the **boy**"

BUT

When a feminine noun is the direct object, the ending changes to -y (as in the example below) or -ю.

Example: **Девочка** идёт домой. "The **girl** goes home"

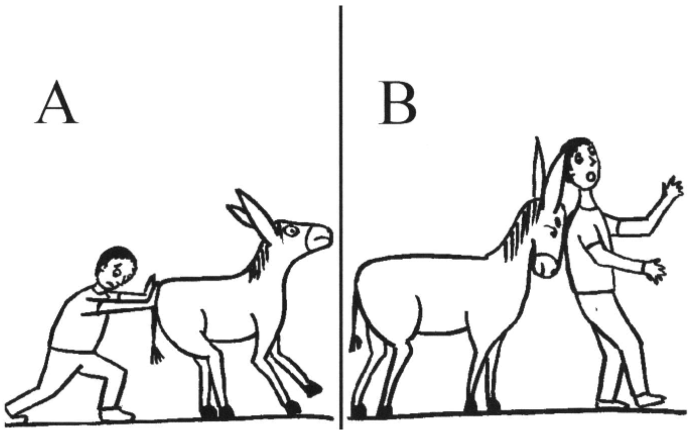
Мальчик видит **девочку**. "The boy sees the **girl**"

After providing information about the grammar feature's use and meaning learners are warned about possible processing problems they might encounter and how to avoid them, e.g. (ibid, p. 272):

[screen 3]

Thus, case markings on nouns become important so that you don't misinterpret who does what to whom. Learners of Russian often rely on word order to determine who did what to whom, thinking the first noun is always the subject. But it may not be! If you see or hear -a /-ä at the end of a masculine noun (remember that a masculine noun in the nominative has no ending), or -y /-io at the end of a feminine noun, that noun is not the "verb-er" and can't be the subject!

In PI there are two types of structured input activities; referential and affective. First, referential activities force the learner to attend to the target grammatical form and its meaning (or function) to complete the task (VanPatten, 2002; Wong, 2004). This may be aural or written and may involve sentence transformation, closed gap-fill, or sentence matching, for example (VanPatten et al., 2012, p. 275):



The man pushes the donkey

Figure 4. Example referential activity from VanPatten et al., 2012.

Affective activities get learners to "express an opinion, belief or some other affective response as they are engaged in processing information about the real world" (Wong, 2004, p. 42). These activities aim to make language comprehension more authentic than that required in the referential activities. Unlike the referential activities the affective tasks do not require focus on the target form for their completion but provide increased exposure to it. In contrast to the referential activities, it would be possible to complete an affective task without using the target form at all (Marsden & Chen, 2011; Wong, 2004). For example, the use of temporal adverbs *usually* and *last night*, mean that the learner does not have to attend to whether the auxiliary *do* was in the present or past e.g.

Step 1. *Answer the following questions for yourself*.

	Yes	No
1. Do you usually clean up the dishes right away after eating?	☐	☐
2. Did you clean up the dishes right away after eating last night?	☐	☐
3. Do you usually make your bed in the morning?	☐	☐

(VanPatten, 2018, p. 6)

2.4.2 The effectiveness of processing instruction

The first study to investigate the effect of PI compared to other instruction was VanPatten and Cadierno's 1993 study. In this study, PI was compared to the effects of traditional instruction. A group of L1 English university students who were enrolled in a Spanish course received either PI, traditional instruction or no instruction on subject and object pronouns in Spanish.

The effect of the two types of instruction were measured using pre and post-tests. An interpretation task was given to all participants in which learners listened to sentences and chose the matching image. A written production task was also given to the learners. This was a sentence completion task in which learners were given a sentence stem and an image as a prompt and asked to write the sentence. The interpretation task was similar to the referential task used in the PI and the production task was similar to the production task given in the traditional instruction. This was intended to counterbalance the advantage each group might have had due to task differences.

This study found that the PI group improved statistically significantly on both tasks, whereas the traditional instruction group improved statistically significantly only on the production task. This suggested, the researchers argued, that PI changed the way the learners processed the input, therefore affecting the learners' developing system and the knowledge accessible during interpretation (VanPatten, 2002; VanPatten & Cadierno, 1993). The researchers argued that the traditional instruction resulted in changes in "learned linguistic knowledge" which is knowledge of "prescriptive (metalinguistic) rules of grammar that do not actually reflect their "real" internal grammar of language" (Schwartz, 1993, p. 151).

Following VanPatten and Cadierno's seminal study, many studies have further investigated PI and its effects on learning. A number of these have produced the same findings as VanPatten and Cadierno's 1993 study (e.g. Benati, 2001, 2005; Benati & Lee, 2012; Cadierno, 1995; Cheng, 2002; Comer & deBenedette, 2011; VanPatten & Wong, 2003; Wong, 2004). These studies found that PI resulted in gains in both interpretation tasks and production tasks whereas traditional instruction only resulted in improvements in production. This was found for various grammatical features, for example, English past tense '*-ed*' (Benati, 2005), French causative *faire* (VanPatten & Wong, 2004), Italian future tense (Benati, 2001), Spanish *ser* versus *estar* (Cheng, 2002), Spanish past tense (Cadierno, 1995) and Russian case-marked nouns (VanPatten et al., 2013).

One limitation with some of these studies was that by operationalising traditional instruction as 'EI plus mechanical drills', the meaningless nature of the practice may have been the reason that the traditional instruction groups performed worse than the PI groups, rather than an effect of the PI itself (see DeKeyser et al, 2002). Studies that have compared PI with more meaningful practice activities have produced mixed results. One study comparing PI, traditional instruction, and meaning-based output instruction for teaching the English past simple tense, found that all three types of instruction had a positive effect on production, but only the PI improved performance on an interpretation task (Benati, 2005). The meaning-based output instruction was very similar to the PI except that structured *output* activities were used as practice tasks as opposed to the structured *input* activities used in PI. The structured output activities involved the learners engaging with, and producing the target form to communicate meaning (there were no mechanical drills). The tasks had two key features; “1) They involve[d] the exchange of previously unknown information and 2) they require[d] learners to access a form or a structure with the intent to express meaning” (Farley & Aslan, 2011, p. 123). Benati’s findings supported those of the studies mentioned above which compared PI only with traditional instruction and claimed that only PI resulted in changes to the language system that affected both interpretation and production.

In contrast to the findings of Benati (2005), at least three other studies (Farley & Aslan, 2011, Keating & Farley, 2008 and Morgan-Short & Bowden, 2006) found that learners given PI or meaning-based output instruction improved an equivalent amount on an interpretation task. In fact, Farley and Aslan (2011) found that not only did PI and meaning-based output instruction have positive effects on interpretation, but meaning-based output instruction had a greater positive effect on production than PI. It must be noted though, that the control group, who received no instruction, also improved on the interpretation task, suggesting a test effect could explain the meaning-based output instruction group’s gains on the interpretation task. All three studies cited above found that the meaning-based output instruction group outperformed the PI and control group in a production task.

Further evidence of the benefits of PI has come from comparisons with enriched input. In these studies, PI has been found to be more beneficial than providing learners with exemplars and tasks that do not focus attention on the target form. Marsden (2006) compared PI and enriched input (EI + tasks which did not require focus on form) and found that learners who received PI outperformed those who received enriched input. This study showed that focus on form that was not task essential (i.e. enriched input) did not result in learning. Engaging with the target form as an essential part of a task appears to be one of the key aspects of PI which results in greater learning gains than training which does not require task essential attention to form. Kasprowicz and Marsden (2018) compared two types of training in which focus on form was task essential, EI + focus on form (form meaning connections) and EI + form spotting (word noticing) and found that both resulted in similar learning gains. This study suggested that attending to form (with or without meaning) results in learning. So, when compared with exposure alone, training on form-meaning connections is more effective for learning, and form spotting (without meaning connections) can be equally effective. Therefore, the more explicit, and task essential, the focus on form perhaps the greater the learning gains e.g. by using EI + tasks focusing on form and meaning + metalinguistic feedback.

2.4.3 Structured input and online processing

The majority of studies investigating the effects of PI have used offline methods as pre and post-tests after training, for example, sentence interpretation tasks, picture choice tasks and production tasks. A few, more recent studies, have used online methods to investigate PI's effects to collect "fine-grained information about moment-by-moment sentence comprehension" and additionally, online methods are thought to tap into more implicit types of knowledge (Keating & Jegerski, 2015, p.2).

One such study used online and offline measures to investigate the effect of feedback and structured input activities on the use of adverb-verb tense and subject-verb agreement in Spanish (Dracos, 2012). Three groups all received the same structured input followed by three different types of feedback (none, correct / incorrect, metalinguistic), there was also a test-only control group. All the learners who received training improved on a number of offline tasks (regardless of feedback type). This showed that structured input alone seemed to be enough to facilitate learning. The group that received metalinguistic feedback performed slightly better than the other two groups. This was not the case in the online task (an SPR) which assessed sensitivity to agreement violations during reading. Longer RTs upon encountering a violation indicated sensitivity to, and implicit knowledge of, the correct grammar. All three groups did not perform differently at post-test compared to pre-test. So, focus on form as part of structured input activities aided offline comprehension and interpretation of the target forms, but did not appear to have an effect on online, implicit processing.

Mixed results were found in a comparison of PI and traditional instruction on the processing of German accusative case markers (Henry, 2015). This study involved two experiments. Experiment one found that while the group that received PI outperformed the traditional instruction group on an offline comprehension task, results of an SPR task did not show significant differences between the two instruction types during online processing of the target structure. The SPR recorded RTs when reading SVO and OVS sentences. Since OVS sentences are less common, increased RTs were expected on the disambiguating segments of these sentences (case-marking). The lack of change seen in RTs on the SPR was suggested to be due to the learners' inability to identify and process relevant cues or integrate this information into their representation of the input. In the second experiment, a third group was added, PI + prosodic cues (P). In this experiment the PI and PI + P groups again outperformed the traditional instruction group in the offline task. The SPR showed that although none of the learners processed the input in a native-like way, the PI and PI + P group showed some changes in processing after training. The two PI groups showed longer RTs for OVS sentences on the definite article of noun phrases (morphosyntactic cue) at post-test compared to the other regions of the sentence and compared to the traditional instruction group. This was taken to be evidence that the learners were sensitive to case-marking on the first noun. So, in the second study, PI did appear to increase sensitivity to the definite article during online processing.

Similar results were found using eye-tracking to investigate the effects of PI on online processing of Spanish active and passive sentences by NSs and non-native speakers (Lee & Doherty, 2019). The non-native speakers were tested before and after receiving PI on the Spanish passive. Their eye-movements were compared to those of the NSs (who did not receive instruction). In terms of accuracy, the learners significantly improved on both active and passive sentences after training, performing almost at the level of the NS group. The learners were faster and more accurate in selecting a picture matching each previously read sentence. The eye-tracking also showed that after training the learners first fixation, and first and second pass times, were reduced for passives and actives. This indicated that less attention to form was needed to accurately assign roles after PI. So, whilst this study did not compare PI to another type of instruction, as in the previous studies, PI was found to result in increased accuracy and more native-like processing behaviour. These findings were in line with previous research using offline measures showing that PI aided the learning of the Spanish passive (Lee, 2014, 2015).

As discussed in section 2.3.4, attention and noticing are key tenants underpinning explicit instruction and PI. In an eye-tracking study, Issa and Morgan-Short (2019) investigated the role of attention using structured input practice, in comparison to textual input enhancement, on the processing of Spanish accusative case pronouns. In the input enhancement condition, pronouns appeared in red font. Both types of training resulted in gains, but the structured input group improved more in terms of accuracy than the input enhancement group. Although the structured input group performed the best in terms of accuracy, they did not show any increase in attention to form (as indicated by eye fixation duration). This suggested that attention played less of a role than expected for the structured input group. Whilst both structured input and input enhancement resulted in attention to the target form, for the input enhancement group, in cases when attention to the target was increased (longer fixations compared to pre-test), greater gains were seen on the post-test than for the structured input group. So, this study provided more evidence for the benefits of structured input for improving accuracy but raises questions about the role of attention.

2.4.4 The role of explicit information (EI) in explicit instruction

One of the constituents of PI is EI, the role of EI in L2 learning has also received attention in recent research. EI is “providing learners with information on a target structure” (Henry et al., 2009, p. 560). The effect of EI prior to meaningful practice was investigated by VanPatten and Oikkenon (1996). Three groups of learners were assigned to one of the following groups, EI plus structured input (i.e. PI), structured input only, and EI only. The PI and structured input only groups outperformed the EI only group. The researchers concluded that the structured input activities, which forced the learners' attention to the target form-meaning connection within meaningful input, resulted in the improvements seen, not the EI. Various studies have replicated VanPatten and Oikkenon's (1996) research and have also found evidence that the presence or absence of EI does not affect learning and does not mediate the positive effects of structured input activities (e.g. Benati, 2004; Sanz, 2003; Sanz & Morgan-Short, 2005; Stafford, Bowden, & Sanz, 2012). Marsden and Chen (2011) found that when EI was paired with two types of training (EI + referential activities and EI + enriched input) the EI + enriched input training did not have a positive effect on learning, whilst the EI + referential activities did. So, the referential activities aided learning, rather than the EI.

Following these studies into the effect of EI on performance on post-training tests, Fernández (2008) investigated the effect of EI *during* training on the learning of three structures in Spanish. EI was found to have a beneficial effect on one structure (the subjunctive in Spanish) but not the other structures (object pronouns and OVS word order in Spanish). For the subjunctive, the learners who received EI processed sentences correctly sooner than the learners who received structured input activities without EI (an index of learning known as 'trials to criterion'). This research suggested that EI might have differing effects depending on the grammar structures being taught.

Henry et al. (2009) conceptually replicated Fernández (2008) to further investigate the effects of EI but with German accusative case markings on articles with both SVO and OVS word orders. In Spanish, agency is indicated in SVO sentences by word order and in

OVS sentences by object pronouns (e.g. *lo*, *la*). In German, agent-patient roles are indicated by case marking. A masculine first noun in an SVO sentence is marked with nominative case marking (*der*) and in OVS sentences by accusative (*den*). Contrary to Fernández, Henry et al. found that the group with EI reached trials to criterion sooner than the group without EI for both SVO and OVS sentences. They suggest that the difference between their findings and those of Fernández may be due to “the intersection of the processing problem and the particular structure” (Henry et al., 2009, p.571). They suggest that object pronouns in Spanish OVS sentences are more complicated than case marking in German since each pronoun in Spanish differs in morphological inflection. In German, there are only two, *der* (nominative) and *den* (accusative).

These studies (Fernández, 2008; Henry et al., 2009) indicate that EI helps learners to process accurately sooner (after fewer practice items) than practice without EI. This may explain why studies using offline measures have not always found an advantage for EI, because when tested after training, the group without the EI has managed to accumulate sufficient practice items to remove any observed advantage. Documenting whether grammar is processed accurately during training may provide a window into whether the benefits of EI are in terms of learners needing ‘less’ exposure (fewer practice items).

Building on evidence from offline methods, the following studies investigated the effects of EI on online processing. In an eye-tracking study on the acquisition of French causative *faire* (Wong & Ito, 2018), structured input activities vs traditional instruction, and the presence or absence of EI, were examined. Two experiments were run. In the first, traditional instruction and structured input activities were compared (both without EI) and in the second, traditional instruction + EI and full PI (EI + structured input activity) were compared. A pre and post-test were administered consisting of a dichotomous picture decision task with eye-tracking. The PI group (without EI) was significantly more accurate than the traditional instruction group (in terms of picture choice). Eye-movements to the correct image also showed that in the experiment without EI, the structured input group performed better than the traditional instruction group (in terms of looks to target). In the second experiment, where both groups received EI, neither group was more accurate than the structured input group in the first experiment. In other words, EI + structured input or EI + traditional instruction, did not have a greater effect than structured input alone, in terms of correct picture choice. Interestingly, eye-movements from experiment two showed that both the EI + structured input and EI + traditional instruction groups looked to the target images at around the same point as the structured input group (without EI) did in experiment one. This suggested that the EI had assisted the traditional instruction group with online processing. The EI + structured input group did not choose the correct image sooner than the structured input group (- EI). So, EI did not appear to be faciliatory when paired with structured input, the structured input activities themselves were sufficient to result in faster looks to target and more accurate picture choice.

2.4.5 Studies teaching cue use for prediction

The studies outlined in the previous sections investigated whether instruction involving focus on form-meaning connection affects not only offline production and comprehension but also online processing. Related to this, is the question of whether instruction that focuses on grammatical form might help learners to use morphosyntactic cues to predict or anticipate upcoming language and so aid them in the speed of their interpretations. This is of particular importance for understanding whether prediction may play a role in learning (as discussed in section 2.3.2), in that unexpected input, which by definition requires that some input be 'expected', may result in the establishment of new representations and therefore aid learning (Zarcone et al., 2016). Expected input is likely to reinforce representations and reduce processing load.

One of the first studies to attempt to train learners to predict was conducted by Andringa and Curcic (2015). This study used an artificial language, based on Esperanto, to investigate whether learners were able to predict an object based on direct object marking and noun animacy. In the Esperanto based language, the direct object marking was the preposition *al* for animate objects (such as "*Ese edzo forigas al.$_{DOM}$ ese kuzo" [The man is feeding the dog]),* but there was no direct object marking for inanimate objects (such as "*Ese edzo visas ese zono" [The man is washing the car])*(ibid, p.41).

The learners in this study were L1 Dutch speakers. Half were assigned to an implicit training condition and half to an explicit condition. Both groups saw animate and inanimate objects and heard sentences describing them. The explicit group's instruction also included an explanation of the direct object marking rules plus examples. In order to assess the effects of the training an eye-tracking test was used. Participants listened to sentences and saw images of two objects on a screen. Half the sentences had inanimate direct objects and half animate. The learners chose which image correctly matched the final noun in the sentence. Half of the trials contained images depicting two nouns with the same animacy and half were different animacy. In the same animacy trials the gender of the article could not be relied upon to determine the correct response. If learners had learned the gender marking rule to such an extent that they were able to use it to predict upcoming language, they would be expected to look to the animate direct object upon hearing *al*. The participants also completed an offline aural grammaticality judgement test to assess explicit knowledge.

The results of the grammaticality judgement test (GJT) suggested that the implicit group did not have explicit knowledge of direct object marking (they performed around chance i.e. 50%) whereas the explicit group scored much higher (83%). As for the eye-tracking and RT results, both groups overall accuracy rates were high suggesting both groups learned the nouns equally well. However, no clear differences were found with regards to the use of direct object marking to predict nouns. The implicit group were expected to show more tendency toward prediction due to their implicit instruction (as the authors considered that the implicit group might develop the kind of knowledge that could be accessed during online processing), but this was not the case. Interestingly, the explicit group showed a different pattern of looks for same image animacy trials compared to the different image animacy trials, and the implicit group. They looked to the target image *less* after hearing the determiner *ese* compared to the implicit group whose looks to target started to increase after this point (as they heard the second noun). The authors suggested that this could be because the same animacy trials were ambiguous for longer and the explicit group were experiencing an "extended period of indecision …. triggered by the presence of metalinguistic information" (ibid, p.262). In other words, the explicit nature of the training may have resulted in the learners thinking about their choice for longer and using conscious effort to try to understand the newly learned language. The researchers concede that this explanation is tentative since the observed difference was "small and unexpected" (ibid, p.263).

Andringa and Curcic's (2015) findings were contrary to those of the studies their hypothesis was based on i.e. that learners are sensitive to morphosyntactic cues during processing (e.g., Dussias et al., 2013; Osterhout et al., 2006; Osterhout et al., 2008; Sabourin & Stowe, 2008; Tokowicz & Warren, 2010). This may be because, in Andringa & Curcic's study, the learners were exposed to the target language for the first time and the instruction was brief, the grammatical explanation given to the explicit group comprised of only a rule statement, and there was no practice phase. As the researchers suggest, reinforcing the rule during the explicit groups' instruction, for instance by repeating the rule or providing feedback, might lead to instruction effects being observed in predictive eye movements indicating sensitivity to the morphosyntactic cues. Furthermore, more experience of the L2 may be needed as the learners may not have been ready to learn a new language structure (R. Ellis, 2005) and apply their knowledge predictively due to low proficiency (Foucart et al., 2014).

A study into predictive cue use in learners with some experience of their L2 investigated whether gender marking could be taught and result in predictive processing (Hopp, 2016). This study consisted of two experiments. In experiment one, instruction on German gender was given to intermediate L1 English learners. The training consisted of drills presenting a picture along with the written form of the article and noun that matched the image. The learners were asked to say the article and noun out loud. After exposure to a set of images and corresponding written forms the learners were shown the images only and ask to recall the name of the item. The training was preceded and followed by a picture naming task and an eye-tracking task (as pre- and post-tests). The eye-tracking test provided the learners with images of four objects whilst they heard a sentence containing one of the objects. If the learners were using gender marking on the article predictively it would be expected that they would look towards the correct image before hearing the noun i.e. upon hearing the article. The pre-test showed that L2 participants did not use grammatical gender as a predictive cue for agreement processing. After training, learners who were consistent in their ability to use the correct article and noun in the picture naming task (training) showed evidence of predictive use of gender marking in the eye-tracking post-test. On the other hand, learners who varied in accuracy in the picture naming task did not show predictive use of the gender marking in comprehension. So, in order to use gender predictively the learners needed to master gender marking offline. This study, and Andringa and Curcic's study, suggested that learners need some experience of an L2 to be able to use cues predictively.

Experiment two (Hopp, 2016) provided evidence of the 'adaptiveness' of predictive processing. Participants were exposed to gender marking errors in German (i.e. incorrectly gender marked determiner–noun combinations). The participants were found to use gender prior to encountering an error, but upon encountering errors, their processing strategy adapted, and they stopped using gender as a predictive cue. This is in line with the theory that prediction is probabilistic and based on a hierarchy of information determined by internal representations of language (Hopp, 2016). Once these representations become unstable, prediction is adjusted to make it more reliable. As argued by Hopp, the parser adapts, and new representations are formed, and that prediction is adjusted, weakened or abandoned (as also found by DeLong et al., 2014; Fine & Florian Jaeger, 2013; Fine et al., 2010).

2.5 Identifying a gap in the research: Simulating cue-use for role assignment in interpreting the passive voice

Both the studies reviewed above (Andringa & Curcic, 2015; Hopp, 2016) investigated whether learners can be trained to use cues predictively. These investigations were based on previous research demonstrating that L1 users do use cues predictively (see section 2.1). They were also, arguably, conducted with a view to finding evidence for whether L2 learners are able to use cues predictively and, if so, contributing to our understanding of whether or not prediction and prediction 'error' may be a learning mechanism available to L2 learners.

As discussed in section 2.1.2.1, NSs use prediction to aid processing, but it is also thought to be an adaptive mechanism. The probabilistic nature of prediction (Kuperberg & Jaeger, 2016) results in prediction being adjusted due to unexpected or novel features in input that appear contrary to internal representations of language (e.g. lexical and syntactic representations). In other words, after encountering a new feature, the existing internal representation becomes less stable and so prediction based on it becomes less reliable; in turn, new representations are formed, and the prediction processes change to accommodate this new feature. For this to occur, a certain amount of experience of the new feature is likely needed to actually change the predictive processes. It is, therefore, thought to be the repeated encountering of previously unexpected information that results in new mental representations and learning (Ellis, 2016).

However, in research to date, L2 research studies have not investigated whether direct training in morphosyntactic cue use including, critically, practice with corrective feedback, may 'simulate' a kind of error-based prediction and, as such, serve as a mechanism for learning. That is, the two studies described above aimed to investigate whether prediction was observable during online processing, but they did not aim to train sensitivity to morphosyntactic cues. The current study aimed to teach leaners to become more sensitive to morphosyntactic cues, through repeated practice and corrective feedback, and then to use them to facilitate online processing and offline production and comprehension.

The language feature that was used to train morphosyntactic cue use was the English passive voice, and the cues which help distinguish it from various types of active voice. These constructions can be disambiguated from each other by a morphosyntactic cue, allowing a listener to assign a role to both the first noun (retrospectively) and the subsequent noun, for example:

*7a. The boy is chas**ed by** the girl*

*7b. The boy is chas**ing** the girl*

Sentence meaning and agent / patient roles are disambiguated by the verb inflection. A passive interpretation is also further supported, in many contexts, by the presence or lack of *by* (although *by* can sometimes serve as a preposition of location, as in '*is chased by (alongside) the river'*, this phenomenon was not examined by the current study). Since first noun animacy is a reliable cue to agency in English, the animacy of the first noun in these sentences would also support the interpretation of that noun as the agent (Bates & MacWhinney, 1989), reducing the chances of expecting a passive construction until a point of clear disambiguation (encountering the '*ed*' and then, in some passives, also the '*by*'). By receiving training that focuses on the morphosyntactic cue ('*ed*' versus '*ing*') at the point of disambiguation, new representations could be formed resulting in increased expectations of encountering the passive voice. Thus, increased sensitivity to the morphosyntactic cue may facilitate processing and subsequent interpretation of future sentences encountered.

The present study investigated the extent to which it is possible to increase sensitivity to a morphological cue to help agent / patient assignment via a type of instruction that draws on the input-based processing training research reviewed above. The study extends that body of research to training expectations of upcoming linguistic input *during* processing itself (whilst learners are interpreting sentences), by including a practice element, with corrective feedback, to help interpretation of the (recently encountered) first noun and the (upcoming) second noun in passive sentences.

The next section reviews the processing problem for L1 Chinese learners of L2 English when interpreting the passive voice in English. First, acquisition of the English passive and issues specifically associated with its processing and acquisition for Chinese NSs are outlined. Second, differences between the passive voice in English and Chinese are described. Finally, research investigating instruction on the passive voice in English is discussed to situate the current study within that body of research.

2.6. The processing problem: The English passive voice

2.6.1 Learning the English passive

The English passive is a feature which takes a long time to master (Hinkel, 2002; Izumi & Lakshmanan, 1998; Larsen-Freeman, 1997; Quinn, 2014; Williams & Evans, 1998), and is acquired late in English NSs as shown by child language acquisition studies (see Armon-Lotem et al., 2016; Marinis & Saddy, 2013; Stromswold et al., 2002). It is also believed to be a complex grammar feature in terms of the quantity of transformations (from a ‘canonical order’) involved in its construction, thus, the passive increases processing load and makes integration more difficult for young children (Stromswold et al., 2002) and for L2 learners (Quinn, 2014). It has also been noted than “learning when to use the English passive

….presents the greatest long-term challenge to ESL / EFL students" (Celce-Murcia & Larsen-Freeman, 2015, p.352).

Research has found that bilingual children process the passive voice slower than NS children (Marinis & Saddy, 2013; Stromswold et al., 2002). This finding has been explained by the competition model (Bates & MacWhinney, 1989). The competition model details how learners use word order, verb agreement, and noun animacy cues to identify agency in a (possible) agent-patient relationship. English as a second language (ESL) research in both adults and children has shown that competing cues in input contribute to slower processing. Cue strength has been found to vary between languages. That is, two languages may have the same cues (i.e. word order), but the strength of one may differ across the languages (see section 2.2). Where an L1 agent-patient cue is different or of different strength to that of an L2 cue, reliance on the L1 cue may result in mis-parsing of agency roles, for example in passive sentences.

Another reason that learning the passive may be difficult is its lack of frequency in input. NS do not use the passive as frequently in spoken English compared to written English, particularly in written academic English. More than double the instances of passive voice have been found in the British National Corpus of written English as compared to the British National Corpus of Spoken English (as found by Roland, Dick, & Elman, 2007). As there is a tendency for instruction to focus on the most frequently used linguistic structures (and often, in more communicatively leaning contexts, on spoken English), the passive is a structure unlikely to receive much attention in the ESL classroom. In most courses that associate themselves with the Common European Framework of Reference (CEFR), British Council, or Cambridge English levels, the passive voice is first introduced in level B1 or intermediate in the International English Language Test (IELTS 5). According to the CEFR the passive could first be introduced in the context of instructions e.g. "*the kit is assembled by*…" (Van EK & Trim, 2001, p. 40). The passive construction as a factitive is introduced later in B1, and in B2 or upper-intermediate (IELTS 6.5) e.g. "*This cathedral was built in the thirteenth century*" (Van Ek & Trim, 2001, p. 55). Given that many universities require IELTS 6 - 6.5 to enter a Master's program, it seems likely that a large number of international students are exposed to academic writing and reading on their Master's programme having only recently been exposed to complex structures like the passive voice in their English language instruction.

This lack of exposure to the passive voice is perhaps reflected in an underuse of the passive voice, compared to NSs, found in advanced ESL learners studying in a university context. In a comparison of advanced English as a foreign language (EFL) and English NSs' writing, EFL learners were found to use around 30% less passive structures compared to the NS writing samples (Granger, 1998). This was also found in ESL and English NSs' essays (i.e., academic essays). Two corpora of 30 essays written by NSs and ESL learners, studying at undergraduate level in a US university, were analysed to compare the frequency of several linguistic features, including the English passive voice. The ESL learners used fewer passive voice structures (when it would have been appropriate for information ordering) and ESL learners often made mistakes when using passives (Russell, 2014).

The research outlined above shows that there is evidence that passives are likely to cause difficulties in processing and acquisition for L2 learners due to cue competition in differing L1s and L2s, and due to the complexity of the passive structure. Another possible reason for L2 learners struggling to parse passive voice sentences is the frequency with which they are exposed to them in language education and in language input.

2.6.2 Why Chinese learners of English?

The present study aimed to train Chinese learners of English because this group was likely to benefit from the training due to both L1-L2 differences and their learning context. Over the last decade, Chinese university students have increased in numbers and now make up the vast majority of non-EU students in the UK (Adams, 2020), suggesting that the amount of academic writing produced by Chinese learners is increasing. Furthermore, there is little research into the processing and learning of the English passive by non-native speakers in general, and in particular by Chinese NSs, as reviewed in the following sections (2.6.3). It has been suggested that the English passive is a potentially difficult construction for Chinese learners to master due to the differences between the construction in English and Chinese and also due to more general differences found between Chinese and English (see Hinkel, 2002; Quinn, 2014). The differences between the Chinese and English passive that are likely to affect Chinese learners of English are outlined in the next section.

2.6.2.1 The passive construction in Chinese and English

The passive in English exists in a long form (8a) and a short (8b) in which the agent is elided, often because the agent is unknown, obvious or unimportant (Collins & Hollo, 2017).

8a. "*Jim was confronted by the inspector"* (ibid, p.136)

(*Patient + Be + Past Participle + By + Agent*).

8b. “*The cathedral was built in 1458*” (ibid, p.137)

(*Patient + Be + Past Participle)*

Another passive form exists in English, which is the *get* passive (as in 9). The *get* passive is used 30 times less per 100,000 words than *by* (ibid).

9. *I'll get the car fixed by the garage tomorrow.*

Chinese has five syntactic passives and three lexical passives (Xiao, McEnery, & Qian, 2006). The lexical passives are words which are semantically passive i.e. *ai* (suffer, endure), *shou* (suffer, be subjected to) and *zao* (suffer, meet with) (see appendix 1 for examples). As outlined in 2.6.2.2 overall the passive voice is less common in Chinese than it is in English.

However, when it is used, the most common syntactic passive in Chinese is the passive formed with *bèi*: *Patient + Bèi + Agent + Verb* (as in 10) (ibid). *Bèi* is considered to be similar to *by* as it marks the agent in the sentence and always appears before the agent (Li et al., 2016a).

10. 热狗 被 男孩 吃 了

Règǒu bèi nánhái chī le

Hotdog by the boy eat [ASP]

The hot dog was eaten by the boy.

In sum, the basic difference in syntax between the English and Chinese passive is that the agent appears before the verb in the Chinese passive, whereas in English the agent appears after the past participle of the verb or is omitted.

However, although the passive in Chinese can be as in (10), as the next section shows there are differences in the amount and type of usage of the passive in Chinese compared to English, and in fact, despite being the most common passive form, in reality the Chinese passive using *bèi* is very rarely used.

2.6.2.2 Frequency of the English and Chinese passive

Chinese is a topic prominent language, so the topic and comment are the canonical parts of a sentence. English is subject prominent, so the subject and predicate make up the principal part of a sentence (Li & Thompson, 1976). To emphasise information in Chinese a different structure is used instead of the passive (Cowen & Reed, 1988). A topicalised object construction would be used to express what in English would normally be a passive, as in (11), and the Chinese passive marker *bèi* is not used.

11. *"Douzi xiaohai reng le*

    (Beans throw boy)

    The beans were thrown by the boy" (Li et al., 1993, p.175)

The *bèi* passive is rarely used in Chinese, and, in fact, all passive-like forms are rare when compared to the frequency of the passive voice in English. In a comprehensive corpus study, the English *by* with a passive function (rather than a preposition about location) was found to be used 955/100,000 words and the *get* passive 31/100,000 words across a variety of genres (Xiao et al., 2006). In academic genres the number of *by* passives was higher, between 1,200-1,400/100,000 words. In Chinese, the various passive constructions in total appear 110/100,000 words (91% of these were *bèi* passives). This shows that in English, passives occur 10 times more than in Chinese. Similarly, reduced numbers of passives were found in in academic genres in Chinese, 50-100/100,000 words.

As mentioned above, unlike in English, Chinese academic writing tends not to use the passive to indicate the importance of information (e.g., an English academic article might read '*The participants were tested*…' so as to foreground '*the participants'*; in Chinese this would read …*The researcher tested the participants'*). That is, Chinese writers tend to use the author of a study as the subject of the sentence (Wang & Wang, 2012). Relatedly, citations are usually given at the beginning of a sentence and are followed by the point, reported findings, or research conducted. In contrast, in English the passive is often used to avoid sentences in which the author of the cited research is positioned in an active subject role as in (12a). This emphasises the importance of the ideas or findings above the individual who stated them. In Chinese the author takes greater prominence in sentence structure, as illustrated in (12b).

12a. *The college was founded in 1925 by Walter Trimble* (Bailey, 2015, p.121)

12b. *Huang (1996) 努力 证明 副词 "都" 是*

Huang (1996) nuli zhengming fuci 'dou' shi

Huang (1996) tries to prove the adverb "dou" is….. (Yulin, 2012).

The findings of Xiao et al. (2006) also showed that in English the passive voice tends to be used to affect an impersonal, objective and formal style, but in Chinese it is most often used as an "inflictive voice" indicating unfortunate events (Chu, 1973; Li et al., 1993). Cowen & Reed (1990) tested these negative context constraints by showing Chinese L1-English L2 learners and an English L1 control group situations which were semantically negative or positive combined with a passive and active sentence (they also manipulated animacy). Chinese and English sentences were used for the L2 group. Participants were asked to select which sentence they preferred given the situation. The native English speakers were not significantly sensitive to the type of situation. In other words, their choice of passive or active did not correlate with the situation being negative or positive. The Chinese group chose passives significantly more in negative situations when written in Chinese, and this tendency was also seen when they were presented with English stimuli, suggesting some transfer of their Chinese L1 preferences about when the passive is appropriate.

The differences outlined above in terms of both the structure and usage of the passive in Chinese compared to in English suggest that it may be a construction that could cause some difficulty for Chinese learners of English, particularly during input processing. The next section outlines previous research into processing of the English passive by Chinese learners of English.

2.6.3 Input processing of the English passive by Chinese native speakers

The Chinese language uses lexical and syntactic cues differently to English. Various studies have found that Chinese speakers are more sensitive to the animacy of nouns than word order when determining the subject in NVN, NNV, VNN[1] sentences (Li et al., 1992, 1993;

Miao et al., 1986). It has also been observed that "speakers of Chinese may have a particular disadvantage when dealing with English passive constructions because their L1 does not have a syntactically-derived passive voice." (Hinkel, 2002, p.15).

Li et al. (1993) investigated the strength of various cues (animacy, lexical markers, word order) in a RT study using auditory input and a picture-decision task. They found that Chinese speakers relied almost exclusively on the marker *bèi* to determine the agent in *bèi* constructions. Animacy was found to have the strongest influence as a cue in *bèi* sentences with animate first nouns. Word order was found to have little effect. This is in contrast to English speakers who tend to rely on word order to determine roles (Liu et al., 1992) This study investigated transfer from L1 - L2 and L2 - L1, between Chinese and English. Chinese-English early and late bilinguals, English-Chinese early and late bilinguals, and English and Chinese monolinguals completed the experiment. Participants heard sentences and were asked to identify the agent. Sentences varied in word order and animacy and were in Chinese and English. Late bilinguals showed evidence of transfer of L1 strategies to the L2, whereas early bilinguals responded more like monolinguals, demonstrating differentiation between L1 and L2 strategies. This study showed that NSs in English and Chinese use different cues to process input, including the passive, and furthermore, that these differences have some effect on L2 processing, especially at lower proficiency levels.

In another study investigating differences in Chinese and English cues, Su (2001) tested Chinese EFL and English Chinese as a foreign language (CFL) learners', as well as English and Chinese monolinguals', on their use of cues in sentences of various word orders (NVN, NNV, VNN, NVN, NNV and VNN[1]). Animacy of the nouns also varied on three

[1] Noun (N), Verb (V)

levels (A-A, A-I, I-A[2]). The same sentences were translated into English and Chinese. Participants listened and indicated the agent in each sentence. The monolingual participants performed as expected, with the L1 English speakers relying on word order (the first noun tended to be interpreted as the agent, regardless of animacy), and L1 Chinese speakers on animacy (animate nouns tended to be interpreted as the agent more than inanimate nouns, regardless of position) (as also found by Bates et al. 1982, Li et al. 1993). In the EFL group, animacy (as per their L1 Chinese) was the strongest cue for beginners, animacy and word order were approximately equal strengths for intermediate learners, and advanced learners used word order more (as English NSs) but still relied on animacy for more unusual sentence orders (NNV and VNN). This showed that L1 strategies were carried over to the L2, but less so at higher proficiencies. CFL learners were found to rely more on word order than animacy, as expected, also indicating L1 influence. Proficiency appeared to be a moderating factor in learners' ability to use cues in a native-like way.

In another study, animacy of grammatical subjects was found to influence grammaticality judgements of active and passive constructions made by Chinese learners of English. English NSs' and L2 English speakers' perceptions of lexical animacy of English nouns and noun phrases were investigated by Hinkel (2002) to explore the influence of these perceptions on the learners' grammaticality judgements of active and passive constructions. Chinese, Korean and Spanish non-native speakers of English participated in the study along with English NSs. In a GJT, non-native speakers' judgements of passive and active sentences differed compared to those of NSs in sentences where the subject (as the agent or patient) was inanimate. In the sentences containing animate subjects (as agents and patients), the

[2] Animate (A), Inanimate (I)

learners and NSs' judgements were much more similar. Of most relevance to the current study, Chinese non-native speakers appeared to make significantly more incorrect judgements compared to NSs when sentences did not contain animate subjects. Overall, all the non-native speakers, even at high proficiency levels, made erroneous judgements based on the animacy of subjects (as agents or patients).

In an online processing study, Chang & Wang (2016) investigated the difference in ERPs during processing of sentences that could be 'literally translated' and 'appropriately translated' from English into Chinese by Chinese-English bilinguals, with differing proficiencies in English. To construct the sentences the CET-4[3] word list was used. Literal English translated sentences were those that could be translated into *bèi* sentences in Chinese ("*The apple was eaten by my uncle*." ibid, p.88), these are syntactically similar in both languages. The non-literal sentences were those in which a *bèi* construction could not naturally be used in Chinese (e.g. "*The violin was made by my father*." ibid, p. 88), so the English and Chinese translations were syntactically different. There were four conditions: "non-violation, semantic violation, syntactic violation, and double violation. Semantic violation referred to the inappropriate meaning of verbs (e.g. 'The violin was **cooked** by my father.'). Syntactic violation involved a misuse of the past participle as the base form of a verb (e.g. 'The violin was **make** by my father.'). Double violation referred to a misuse of the past participle as a base form of a verb along with an inappropriate meaning (e.g. 'The violin was **cook** by my father.')." (ibid, p.89).

[3] The College English Test (CET-4) is a national English as a foreign language test in the People's Republic of China. CET-4 is based on 4500 words.

The learners read sentences and indicated whether they were correct or incorrect by pressing a button. The sentences were delivered using a word-by-word display and ERPs were recorded. The time participants took to indicate if sentences were correct or not was also recorded (RT). Reduced RTs and accuracy rates when interpreting syntactic violations showed that the high proficiency group was more accurate across all conditions compared to the low proficiency group. The higher proficiency group also showed greater sensitivity to syntactic and double violations, whereas the lower level group showed no difference across conditions. The ERP results indicated that the low proficiency learners were not as sensitive to the verb ending cue (the past participle) as the high proficiency group and so were unaware of the syntactic violation. This was demonstrated by smaller P600s recorded in the lower proficiency group compared to the higher group. P600s indicate sensitivity to syntactic violations. Further evidence for intermediate learners' lack of sensitivity to syntactic cues was provided by analysis of N400s. N400s indicate sensitivity to semantic violations. Only double violations resulted in N400s for high proficiency learners whereas all types of violation (semantic or syntactic) resulted in N400s for low proficiency learners. For both high and low proficiency learners the non-literal translated sentences caused processing problems (as evidenced by longer RTs). This differences in interpretation and processing shown by low and high proficiency learners suggested that it may take a lot of practice and / or experience for Chinese learners of English to become sensitive to the morphosyntax of the English passive. A limited number of studies have investigated training and practice of the passive voice, in particular for Chinese learners of English. The next section details research into training and practice of the passive voice.

2.6.4 Instruction and the passive voice in English

In the first study to my knowledge into the effects of instruction on Chinese learners of the English passive, Qin (2008) investigated the effects of dictogloss versus PI on the learning of the English passive by Chinese beginner learners. Dictogloss involves listening to a text, note taking and reconstructing the text, often in pairs. The experiment used a pre-test post-test design with outcome measures principally focusing on measuring explicit knowledge. The PI group received metalinguistic information about the passive and five structured input activities. The dictogloss group received the same metalinguistic information, and dictogloss activities using the same texts as those used in some of the structured input activities included in the PI. There were two dictogloss activities presented in stages in order to provide similar length instruction to both groups.

There were seven outcome measures. The first two tests asked participants to identify the agent or patient in active and passive sentences. The third and fourth tests asked learners to choose the appropriate passive or active response to a question (from a pair of options). The fifth test asked the participants to translate English sentences (passives and actives) into Chinese. The sixth test required the learners to complete sentence stems with infinitive verbs provided. Finally, the seventh test asked the participants to recreate a short story in English after being given an outline in Chinese. Both groups showed significant gains at post-test and these improvements endured to delayed post-test, but the PI group performed significantly better than the dictogloss group on the comprehension tests and the production test. So, both types of instruction were beneficial, but PI more so.

In order to further explore how PI would affect the interpretation and production of the passive, VanPatten and Uludag (2011) examined the effectiveness of PI on pre-intermediate Turkish learners of English. The participants were assigned to a control (test-only) or PI group. The treatment packet for the PI group included EI about the passive and first noun strategy (see section 2.2). It also included nine structured input activities. The control group did not receive any instruction and did tasks from a textbook not involving the passive. There were three assessment tasks given as a pre-, post- and delayed post-test; these were an interpretation task (aural input followed by a choice of two sentences, one of which correctly described the aural input), a sentence level production task, and a passage reconstruction task. There were no significant differences between both groups at pre-test. Compared to the test-only group, the PI group made significantly greater improvements in all test scores between the pre-test and first post-test and maintained the gains until delayed post-test. This showed that PI changed the outcome of not only interpretation tasks (as often used with PI) but also on two kinds of production tasks.

In a replication of the above (Uludag & Vanpatten, 2012) a third group was added, so that the groups were PI, dictogloss and control. The PI training was the same as used in the 2011 study. The dictogloss condition received the same EI on the passive in English as the PI group, but the information given did not include first noun strategy information. The dictogloss group heard nine texts with one-three tokens of the passive. The number of tokens the PI and dictogloss groups were exposed to during treatment were "roughly the same" (ibid, p. 198). As in VanPatten and Uludag’s (2011) study, the assessments used were an interpretation task, a sentence level production task, and a passage reconstruction task, these were given as pre-, post-, and delayed post-tests.

The results for the interpretation task found that both the PI and dictogloss group improved between pre- and first post-test. The control group did not show significant differences between pre- and post-tests. The PI group's results improved significantly more than the dictogloss group’s results. Furthermore, the PI group outperformed the dictogloss group at the delayed post-test. For the sentence level production test, PI did not result in gains significantly different to the dictogloss group, with both showing similar improvements compared to the control group. The same pattern was found for the sentence reconstruction task. In sum, the PI group outperformed the dictogloss and control groups on the interpretation task, but not on the two production tasks. The PI and dictogloss resulted in similar improvements at each time on the other two tasks as compared to the control group who showed no improvements at each time for any of the three tests. These findings are in line with previous research into PI and dictogloss (VanPatten et al., 2009) but not with Qin’s findings (2008). Uludag and VanPatten (2012) explain this as a result of issues with Qin's methods, such as the test instruments, lack of control group, and uncontrolled variables (e.g. lexical semantics of agents and patients used in the test stimuli).

The above studies investigated the effects of instruction for learners who have not previously been exposed to the English passive voice. They do not investigate whether processing can be changed after a structure has been previously learned in order to correct erroneous parsing. A study that investigated the effects of instruction on the use of the English passive by Chinese learners who have previously learned the construction but who made errors interpreting and producing it was carried out by Li et al. (2016a).

This study investigated the effect that timing of feedback has on learning. Learners were assigned to one of four groups: immediate feedback, delayed feedback, no feedback, and control (test-only). The three experimental groups performed two dictogloss tasks in groups. The two groups with feedback received either immediate or delayed corrective feedback in the form of a prompt, followed by recasts of utterances that had contained errors using the target structure (the past passive voice). The effects of feedback were tested using a JT (intended by the authors to measure explicit knowledge) and an elicited imitation test (intended by the authors to measure implicit knowledge).

Both types of feedback were found to improve scores on the JT, with immediate feedback having the greatest effect. Feedback therefore appeared to be beneficial for the development of explicit knowledge, as measured by the JT. Feedback did not have an effect on the scores for the elicited imitation test, claimed to measure implicit knowledge. However, in a similar study into task-based and task-supported instruction immediate feedback *was* found to have an effect on elicited imitation test scores when learners were divided into those with prior knowledge of the passive and those without (Li et al., 2016b). In this study, four groups were given different types of instruction with or without immediate feedback; 1) task only, 2) EI + task, 3) task + immediate feedback and 4) EI + task + immediate feedback. The 'task' was two dictogloss tasks.

The general findings were that the more explicit the treatment, the greater the effect. In other words, the EI + task + feedback group made the greatest gains, and the two groups who received EI performed better than the groups that did not. This was found to be particularly true for the development of explicit knowledge as demonstrated by JT scores. General gains, over all learners, on the elicited imitation test were not seen. However, when the learners were divided into those with prior knowledge of the passive and those without, some improvement was seen for the elicited imitation test scores for those with prior knowledge. This was especially true for those learners with prior knowledge who were given the most explicit treatment (EI + task + feedback). This suggests that EI in combination with a task and feedback may result in (possibly implicit knowledge) gains if learners have some prior knowledge of the passive structure.

The current study built on the previously outlined research in some ways. The learners had some prior knowledge of the passive voice. The training in the current study explicitly focused learners' attention on the relevant morphosyntactic cue necessary for interpreting passive constructions (*be* + the past participle of the verb and, in some sentences, *by*) during training to help interpretation of passive sentences. Along with EI and feedback, this training aimed to increase cue sensitivity and result in online and offline processing changes.

2.7 The current study and research questions

2.7.1 Rationale for the current study

Most previous research into the effectiveness of instruction on processing and usage of morphosyntax has used EI plus tasks designed to focus on form and meaning e.g., PI. EI plus tasks (e.g. focus on form, structured input, meaning-based output instruction, dictogloss). EI plus tasks and feedback have all been shown to be beneficial for learning as demonstrated using various offline and online tests. This suggests that the more explicit the instruction the better for learning, at least on the measures used in studies to date. However, it is not clear whether explicit knowledge and / or implicit knowledge is / are generated as a result of increasingly explicit instruction and practice, as the measures have tended to tap more into explicit knowledge, though with some claims being made about elicited imitation tasks demonstrating implicit knowledge. Furthermore, a small number of studies have built on studies of L1 and L2 grammar processing (see sections 2.1 and 2.2) and have attempted to train learners in cue use in order to facilitate (predictive) processing (e.g. Andringa & Curcic, 2015; Hopp, 2016). However, these studies did not combine the teaching of morphosyntactic cues along with EI, practice, and feedback. Moreover, no study has attempted to do this with the English passive voice which is rich in competing cues for L2 learners – word order, animacy, and morphosyntax.

The current study aimed to build on research into EI, input-based practice with feedback, morphosyntactic cue use, and L2 processing, to design training which forced attention to morphosyntactic cues in order to complete the training tasks correctly. When designing the study, it was thought that surprisal may result from exposure to cues that were not in line with expectations or predictions that the learners had made on starting to parse a sentence as an active, which in fact turned out to be passive. In the current study the terms 'expectation' or 'prediction' are used to refer to the pre-processing of input and generation of expectations incrementally during processing (DeLong et al., 2014).

The current study explored the effects of such training in a more comprehensive way than many previous studies by testing learning using both online and offline tests, and comprehension and production measures. The passive voice was chosen as the target structure because it is a structure that is disambiguated by morphosyntactic cues (*be* + past participle, and sometimes *by*) and is likely to be misinterpreted by learners due to reliance on the first noun principle (VanPatten et al., 2013). On encountering the (less likely) morphosyntactic cue 'past participle' in a passive (encountering '*he is helped*' rather than the -*ing* in '*He is helping'*), a listener is forced to assign the patient role to the first noun (retrospectively) and the role 'agent' to the subsequent noun. So, the current study investigated training on morphosyntactic cues that could serve role assignment, if learners were indeed (mis-parsing) passive sentences as active sentences during online processing.

Noun animacy is thought to play a significant part in how languages assign grammatical and thematic roles (Carroll & Schlesewsky, 2006; Primus, 1998). Noun animacy has been found to affect online processing of agent-patient roles (Andringa & Curcic, 2015; Jackson & Roberts, 2010; Mak, Vonk, & Schriefers, 2002; Traxler, Morris, & Seeley, 2002; Traxler et al., 2005; Weckerly & Kutas, 1999), to add to this research the current study also investigated animacy effects. Animacy was manipulated in all of the outcome measures but was mainly expected to affect online processing. Its role in offline interpretation, in this study in a written acceptability judgement task, was also investigated since research has found that animacy affects learners online processing of verb phrases (e.g. Mak et al., 2002; Traxler et al., 2002; Traxler et al., 2005) and so may also affect offline grammaticality judgements in the same region (i.e. the verb). Animacy was not expected to play a significant role in production since in the current study the tasks were controlled, and interpretation of language input was not required to complete the tasks.

The current study also investigated the learners' ability to generalise knowledge learned in the interventions to constructions not included in the training. Previous research has found that awareness, and access to explicit knowledge, is a predictor of the ability to generalise in offline measures (e.g. in elicited oral production as found by Brooks et al., 2006; Brooks & Kempe, 2013; Kempe et al., 2010; Kempe & Brooks, 2008). Therefore, training that is likely to increase awareness would be expected to result in a greater ability to generalise. This was investigated for both online and offline language use in the current study.

The learners selected for the current study were Chinese L1 upper-intermediate learners of English studying pre-master's English courses at the University of York. Learners with prior knowledge of the passive were chosen to investigate whether erroneous language processing and / or use could be improved by training. The passive is also of special interest to this particular group of learners because of its prevalence in academic English.

To set our findings about leaners of L2 English in context, we also investigated English NSs' online processing of passive and active constructions. This adds to research into the nature of NS processing by investigating if NSs of English use morphosyntactic cues during online processing of the passive voice or if NSs do not rely on these cues and instead process passive constructions in a more shallow fashion, as posited by the good enough processing approach (Ferreira et al., 2002).

2.7.2 Research questions

In line with the rationale set out above, the current study aimed to answer the following research questions:

RQ1) To what extent do L1 Chinese learners of L2 English show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

Is cue sensitivity affected by:

a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?
b) First and second noun animacy?
c) Trained (present simple) and untrained (present perfect) constructions?

RQ2) To what extent do L1 Chinese learners of L2 English show knowledge of cues in a) offline grammaticality judgements and b) production of the passive voice in English?

Is production affected by:

a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?
b) Trained and untrained constructions (present simple and present perfect)?

Are grammaticality judgements affected by:

a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?
b) Trained and untrained constructions (present simple and present perfect)?
c) First and second noun animacy?

RQ3) To what extent do native English speakers show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

To what extent is native speaker processing of the passive different to learners?

Chapter 3: Methodology and Methods

This intervention study used a pre, immediate post, and delayed post-test design. The study was conducted in a laboratory setting in order to best control and manipulate the experimental variables. Although laboratory settings result in lower ecological validity than a classroom study it allowed for the training that each participant received to be identical (Gass, Mackey, & Ross-Feldman, 2005; Hulstijn, 1997). Since the aim of the current study was to manipulate training for a specific grammar feature and then assess its comprehension and production as a result of the training tasks alone, a classroom study could have diminished the validity of any claims i.e. “if rival causes or explanations can be eliminated from a study then clear causality can be established; the model can explain outcomes” (Cohen, Manion, & Morrison, 2017, p.272). It is noted though, that for research to inform practice, instruction needs to be tested in the classroom (Hulstijn & de Graff, 1994; Nunan, 1991). This is discussed in the implications for future research (section 6.3).

3.1 Methodology

The first half of this chapter discusses the methods that were selected for the current study, highlighting their strengths and weaknesses and so justifying their use in the current study. Many methods have been developed to investigate L2 learning and knowledge. In the current study, for RQ1 and RQ3 it was necessary to use a method that measured online processing behaviour. For RQ2 it was necessary to use methods that measured offline comprehension, and production. This methodology section covers only those techniques that were deemed relevant to the current study i.e. for evidence about offline comprehension, JTs; for evidence about online processing, visual world eye-tracking; for evidence about production, oral production tests and written sentence completion.

3.1.1 Offline measures of comprehension: the grammaticality judgement test (JT)

Offline techniques, such as picture-matching tasks, truth-value tasks and act-out tasks, usually allow, promote, or even require awareness of the language, and sometimes metalinguistic knowledge to complete the task (Marinis, 2010). One of the most common offline measures of comprehension and grammatical knowledge are JTs. They are often used alongside online measures, and also in studies investigating the effectiveness of instruction. Other offline comprehension tests are not discussed in detail as this is beyond the scope of this thesis (for more about offline tests see Doughty, 2008; Ellis, 2009; Norris & Ortega, 2000).

3.1.1.1 Judgement tests (JTs)

JTs have been widely used in L2 learning research to assess knowledge of various aspects of language, including syntactic structure, morphological features and plausibility (see Plonsky et al., 2019 for a systematic review). They serve to investigate the extent to which learners and / or NSs find grammatical and ungrammatical language (written or oral) acceptable or not. The notion of grammatical and acceptable is not clear cut as, for example, NSs may deem a grammatically incorrect sentence acceptable in certain contexts. Grammaticality, or acceptability, is usually indicated using a dichotomous choice (correct / incorrect, acceptable / unacceptable) or a Likert scale (1 = ungrammatical – 5 – grammatical) (Plonsky et al., 2019). The majority of L2 learning research (53%) has used a dichotomous choice for judgements (as found by Plonsky et al., 2019).

In some cases, participants are also asked to identify the error, and / or correct the error, and this has been thought to tap into explicit knowledge (as shown by Bialystok, 1979, 1982). By identifying which part of the sentence the participant believed to be wrong, and by correcting it, the researcher can have some insight into why the sentence was judged incorrect or unacceptable. A small number of studies also asked participants to rate the certainty of their responses or to explain the basis for their decision (Plonsky et al., 2019).

Most JTs used in L2 research are untimed (Plonsky et al. 2019). Untimed JTs are thought to tap (more) into explicit knowledge because the lack of time pressure allows participants to use metalinguistic knowledge to actively think about their judgements. Timed JTs are often used with the intention of avoiding or reducing the extent to which explicit knowledge can be tapped into. A time limit gives participants less opportunity to consider each stimulus carefully, reducing the likelihood of being able to access metalinguistic knowledge (Godfroid & Winke, 2015; R. Ellis, 2005; Loewen, 2009). Research has found that time pressure is a key influencing factor in L2 learners' performance on JTs (e.g. Bowles, 2011; R. Ellis, 2005; Ellis & Loewen, 2007; Godfroid et al., 2018; Han & Ellis, 1998; Zhang, 2015). In a meta-analysis, scores on untimed JTs were found to be higher than when timed (d = 1.35; interquartile range = 1.74) (Plonsky et al. 2019).

However, recent research indicates that timed tests may tap into automatised explicit knowledge rather than implicit knowledge (Suzuki & DeKeyser, 2015), or that at least, it is difficult to distinguish automatised explicit knowledge from implicit knowledge. Automatised explicit knowledge is explicit knowledge which is accessed rapidly and with awareness (ibid). What distinguishes automatised explicit knowledge from implicit knowledge is thought to be the presence of awareness. So, a timed JT may result in rapid access to knowledge that the learner is aware they are accessing. Some evidence for this comes from a confirmatory factor analysis study on the results from a timed auditory GJT and a timed visual GJT (Suzuki & DeKeyser, 2015). Confirmatory factor analysis showed that both types of JT accessed automatised explicit knowledge rather than implicit knowledge. This was determined by high factor loadings for automatised explicit knowledge and low factor loadings for implicit knowledge for both JTs. Furthermore, online tests included in the study (SPR, visual world eye-tracking and a word monitoring task) loaded onto separate factors to the JTs. This suggested that the online tests and the JTs (whether timed or untimed) measured different constructs. This study suggests that timing alone may not be enough to ensure implicit knowledge is being accessed during a JT. Furthermore, Gutiérrez (2013) found that the grammaticality of items had the greatest effect of the type of knowledge accessed, rather than time pressure. Grammatical items were more likely to measure implicit knowledge, she argued, whereas ungrammatical items measured something different, argued to be explicit knowledge.

In addition, modality may have an effect on the type of knowledge tested, auditory JTs have been found to tap into implicit knowledge, whereas written JTs into explicit knowledge (Bialystok, 1979, 1982; Kim & Lee, 2019). Spada et al. (2015) tested learners with auditory and written timed JTs and found that they loaded onto different factors in a confirmatory factor analysis. The auditory JT was found to be associated with implicit knowledge and the written JT with both knowledge types. Ungrammatical items are thought to measure explicit knowledge, rather than implicit knowledge, so if modality *is* linked to knowledge type its manipulation could result in changes in judgement accuracy of ungrammatical items. So, auditory input may result in implicit knowledge being accessed and an impaired ability to assess ungrammatical items. Evidence for this is provided by research that has found that learners are more accurate in grammaticality judgments that are presented visually (written) rather than aurally, and that modality effects are more significant for ungrammatical items (Johnson, 1992; Murphy, 1997). However, in a meta-analysis of 17 JT studies using both modalities, Plonsky et al. (2019) found little reliable difference between auditory and written JTs in terms of learners' performance (median d = .14; interquartile range = 1.04). Therefore, evidence as to the effects of modality is somewhat mixed but for the purposes of measuring explicit knowledge, as in the current study, a written JT appeared to be most appropriate.

A weakness of JTs is that it is not possible to record the precise moment a learner becomes aware of problematic language. So, the parsing strategies used to interpret grammaticality, or acceptability, of a sentence are not observable through JT data alone. For example, the reader may read and re-read a sentence before making a judgement, they may guess when making a judgement, or they may make a judgement based on only one segment of the sentence which is not the intended grammatical manipulation. Online measures are able to record these differences as they observe incremental processing during the parsing of a sentence.

3.1.2 Online measures of comprehension: Visual world eye-tracking

Online measures have an advantage over offline measures because they are thought to be not affected (or less affected) by metalinguistic knowledge (Marinis, 2010). These methods allow researchers to tap into (more) automatic processes during input processing, such as; the application of grammar, discourse rules, lexical connotations etc. (Roberts, 2012). For the current study, one of the aims was to investigate how learners process real-time language input and assign meaning to it whilst listening. Various online methods are available, such as SPR and self-paced listening or EEG. These aim to record behaviour or brain activity whilst learners are trying to get meaning from language they are exposed to, often with the aim that learners do not become explicitly aware of the purpose of the study so that processes that occur without awareness can be recorded. In the present study, visual world eye-tracking was deemed more suitable than self-paced reading or listening because in self-paced reading or listening, input is presented segmentally (and usually non-cumulatively in the case of SPR) and this can impede normal parsing strategies. Whilst the RTs collected are demonstrative of processing load, they cannot tell us about actual interpretation of the meaning of a sentence (unlike visual world eye tracking which can indicate meaning [role] assignment, as was desirable in the current study).

3.1.2.1 Visual world eye-tracking

Visual-world eye-tracking paradigms in language research involve providing auditory input along with visual input, and tracking eye-movements whilst participants listen to the input. Visual world eye-tracking is based on the eye-mind hypothesis or eye-mind link (Reichle & Reingold, 2013). This hypothesis states that there is, as Lee and Doherty put it, “a close relationship between what the eyes are fixating upon and what is being processed by the brain” (2019, p. 857). Eye-movements are thought to be linked to attention, specifically attention orienting (Godfroid, 2019). As previously discussed in 2.3.4, attention is an important aspect of processing and subsequent learning. Underpinning the use of visual world eye-tracking is pre-motor theory which argues that the overt eye-movement, recorded by eye-movement tracking, reflects the covert orienting of attention to the sound stream (Rizzolatti et al., 1987; Sheliga et al., 1994; Sheliga et al., 1997). Research has shown that it is possible to measure covert orienting of attention by tracking overt attention (Wright & Ward, 2008). Pre-motor theory posits that in normal input processing, covert attention orienting (e.g., to a particular sound in the input, or in the case of reading research, to letters or words) occurs prior to overt attention orienting. In other words, covert attention will be directed at a target stimulus before an eye-movement can happen. It is thought that the time taken to plan an eye-movement (i.e., the time between covert attention and overt attention), is around 220ms (Wright & Ward, 2008). Eye-tracking researchers study overt orienting, by observing eye movements, to learn more about covert orienting.

The visual world paradigm has been used in language research to observe incremental online syntactic processing and to show that listeners appear to integrate grammatical roles into sentence interpretation as soon as they hear enough information (as shown by Altmann & Kamide, 1999; Kamide et al., 2003b). In the visual world paradigm, eye-movements are believed to indicate which linguistic representations are activated during the processing of audio input. For example, in the case of gender marked pronouns, eye-movements upon hearing a pronoun indicate how they are interpreted in relation to images depicting their possible antecedents. To give an example, in Spanish, gender of a noun is marked on the article, *el* (masculine) and *la* (feminine). In the example below it would be expected that upon hearing *la* in *'encuentra la…'*. eye-movements would orient to the image of the only feminine noun represented in four images i.e. *la pelota* (the ball) (see figure 5 below).

Figure 5. Example adapted from Lew-Williams & Fernald (2007, p.10).

Eye-movements in visual-world are less likely to be able to provide insight into the precise moment at which processing load occurs, relative to text-based eye-tracking (reading studies) because eye-movements are prompted by the incoming auditory input so, regression and skipping are not possible, as they are in reading (Godfroid, 2019). Visual world eye-tracking does, however, give insight into the nature of interpretations made upon hearing oral linguistic cues and into how sensitive participants are to linguistic features such as cues for interpreting meaning. Text-based and visual world eye-tracking paradigms therefore investigate different types of input processing (reading and listening, respectively) and different aspects of that processing (processing difficulty and mental representations / interpretation, respectively). For the current study, we were interested in how participants interpreted oral morphosyntactic cues, thus visual-world eye-tracking was deemed the most appropriate.

3.1.3 Measures of production: Oral and written production, from free to constrained

Including measures of production provides more comprehensive insight into the efficacy of instruction and the nature and generalisability of learning gains. Various production measures have been used in L2 research, measuring more or less explicit knowledge or implicit knowledge depending on their design. The below figure (6) (Ellis, 2009, p.40) shows how certain task features result in a task measuring more implicit knowledge or explicit knowledge. The next section outlines tests used in L2 research to measure oral and written production.

Criterion	*Implicit knowledge*	*Explicit (analyzed) knowledge*
Degree of awareness	The task requires the learner to respond according to 'feel'	The task encourages the learner to respond using 'rules'
Time available	The task is time-pressured	The task is performed without any time pressure
Focus of attention	The task calls for a primary focus on meaning	The task calls for a primary focus on form
Systematicity	The task results in consistent responses	The task results in variable responses
Certainty	The task results in responses that the learner is certain are correct/incorrect	The task results in responses the correctness/incorrectness of which the learner is uncertain about
Utility of knowledge of metalanguage	The task does not require the learner to use meta-linguistic knowledge	The task invites the learner to use metalinguistic knowledge
Learnability	The task favors learners who began learning as children	The task favors learners who have received form-focused instruction

Figure 6. Task features and corresponding knowledge type measured (Ellis, 2009, p.40)

Whilst there are many types of oral production test the most common types appear to be narration / story retelling or picture description. After conducting a search on IRIS[4], which at the time of searching held 233 oral production tests used in L2 research, 228 of these were either narration / story retelling or picture description. These two tasks may measure more or less explicit knowledge or implicit knowledge depending on whether or not time limits are imposed, or whether metalinguistic knowledge is required to complete them, and the awareness of the individual participants. For example, the Marsden project (Ellis, 2009) used an oral narrative task designed to measure implicit knowledge. A time limit was used, and the learners needed to focus on meaning and intuition to complete the task. Metalinguistic knowledge was not required or encouraged by the task. The oral production test was used alongside an oral elicited imitation task, a timed JT and an untimed JT, and a metalinguistic knowledge test. Confirmatory factor analysis found that the oral elicited imitation task, oral production test and timed JT measured a separate type of knowledge compared to the untimed JT and the metalinguistic knowledge test, defined as implicit and explicit respectively (Ellis, 2009). The tasks designed to measure production or comprehension did not appear to measure different types of knowledge.

Using an untimed narrative task, Spada et al. (2015) found that story re-telling with no time pressure appeared to measure explicit knowledge, not implicit knowledge, based on confirmatory factor analysis. The task was originally designed to measure implicit knowledge, since there was no focus on rules or form, and although not timed, the learners were encouraged to tell the story without much pause. Therefore, the finding that the oral production test measured explicit knowledge was surprising (Spada et al., 2015). However,

[4] www.iris-database.org

since the task was not timed, and prompts to the correct verb were provided on each slide, its design may have resulted in an unintentional focus on form. Similar was found by Erlam (2005) in a study using a narration task to elicit direct object pronoun forms. There was no time pressure, so the learners had time to prepare and they were given verb prompts. This task design was therefore more likely to measure explicit knowledge than implicit knowledge. Learners received rule-based training (deductive) or inductive training prior to the task. Those who received rule-based training performed best. This might be because the explicit nature of deductive instruction lends itself to a measure tapping into explicit knowledge.

Another common group of oral production tests are sentence completion tasks and picture cue tests. These involve the completion of sentences, possibly describing an image, using prompts and cues. For example, in VanPatten and Cadierno's study (1993), learners were shown two images depicting a scene and asked to complete a sentence e.g. *The boy is thinking about the girl and then* …………………… The images were of a boy sitting at home thinking about a girl, and the second showed the boy calling the girl on the phone. Other studies provide the target verb, for instance, the following images were provided to elicit the passive, preceded by these instructions:

"*Complete the sentences with a present passive. Use the followings verbs: ship pick take dry sort*" (Seyednejad & Gholami, 2017, p.742).

Figure 7. Example sentence completion task from Seyednejad & Gholami, 2017.

Tasks like these are likely to measure explicit knowledge since they focus on form and are untimed. The task used by VanPatten and Cadierno (1993) was designed to be similar to those used in traditional instruction, i.e. tasks focusing on form rather than meaning. Free oral production measures which focus on meaning and are unstructured, such as, asking learners to talk for a minute on a topic without preparation, might be more likely to tap into at least some implicit knowledge. However, it is not always easy to design free production tasks to investigate the effects of instruction on specific grammatical features since specific features may not be elicited by the task, as learners can avoid using the structure. Hence the need for restricted production tasks, such as sentence completions or picture descriptions.

Written sentence completion tasks that are similar to the oral production tests described above are common in L2 research. Due to the increased time usually given for written tests, they are more likely allow for reflection and correction and therefore to measure explicit knowledge more than oral sentence completion tasks (R. Ellis, 2005; Godfroid et al., 2015). An example of a written task is the following narrative task from Hsu (2017) (see figure 8 below). In a study investigating the effects of planning on production, learners were asked to write a story corresponding to the sequence of images shown in figure 8. Tasks such as this one, which allow for freer writing than a sentence completion task, are not often used in studies of language processing in which specific grammatical features are investigated, because, as noted above, it is harder to ensure production of specific features when the task is less restricted. There is a trade-off therefore in intervention studies between evidencing the learning of a specific grammar feature and testing natural language use.

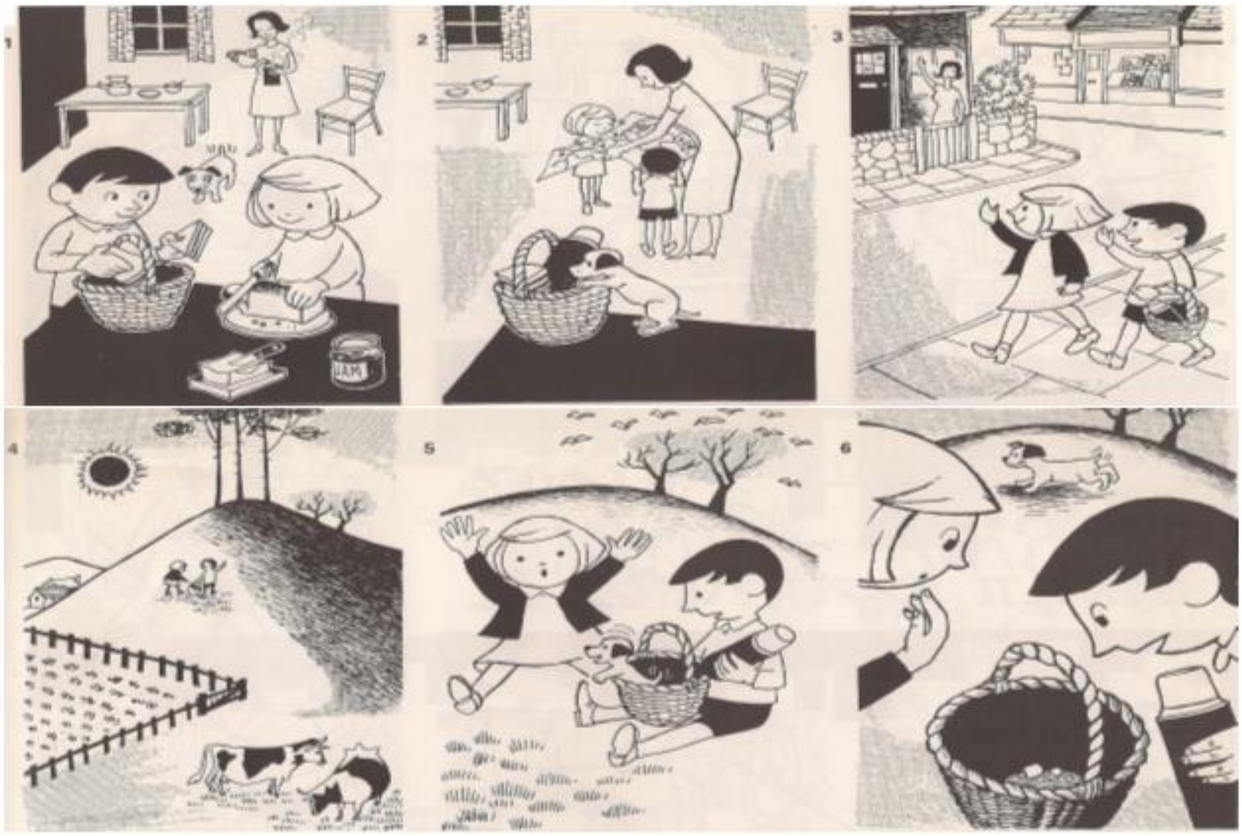


Figure 8. Narrative task from Hsu (2017)

3.1.4 Test modality

As mentioned above, written, oral, and aural tests are all used to measure comprehension and production. The effect that modality may have on language processing and production must be considered during study design. It has been observed in psychology research that aural and visual input are processed in different parts of the memory system (Penney, 1989). L2 learning research has also demonstrated that modality influences how learners process input (Wong, 2001) and that modality has an influence over the knowledge type being accessed by a task (Spada et al., 2015). The modality of input has been investigated in a number of studies. For example, Ito and Wong (2019) investigated whether modality effected training and subsequent language processing. They replicated a previous study into the effectiveness of written PI (Wong & Ito, 2018) by using the previously tested written PI input in an auditory form. The results from the two studies were then compared. Findings suggested that written training improved learner sensitivity to grammatical cues more so than auditory training (based on eye-movement data after training).

The brief summary of methods used for measuring language knowledge given above suggests that tests choice needs to consider modality, time, and whether the knowledge is being accessed 'online' or 'offline' or in comprehension versus production. All these considerations are needed as the current study aimed both to inform our theoretical understanding of processing and cue use, as well as our understanding of pedagogical effectiveness. The following sections now describe the actual methods adopted: the materials and procedures utilised in the current study.

3.2 Ethical considerations

Ethical approval was sought from The University of York Education department ethics committee prior to data collection. All the participants were adults and no part of the study was expected to cause harm or distress. Anonymity was ensured by assigning a participant identifier (a number) to each participant which was used in all parts of the experiment. Their name and contact details were kept in a password protected spreadsheet to which only the researcher had access. This was necessary given the longitudinal nature of the study. The participants' contact details were deleted after data collection. Furthermore, all steps were taken to minimise the time required to complete the tasks, and all participants were given a small cash reward for their participation.

Prior to taking part, all participants were asked to provide informed consent after reading an information sheet (see appendix 2 for forms). The information sheet informed participants about the tasks they would complete and explained that they would need to come for three or four sessions (depending on the group they were randomly assigned to). The consent form explained that their anonymised data would be kept for an indefinite period and would be used in future publications. The participants were asked to indicate that they agreed to this and were informed that they could withdraw at any point during data collection.

Since some of the participants would be in the control group and not receiving training, they were offered the chance to return after the end of the experiment for the training sessions.

3.3 Participants

Participants were 73 Chinese learners of English enrolled on the Education Department pre-sessional at The University of York. They had been in the UK for less than 2 weeks at the start of the study and were intending to study at postgraduate level at The University of York for a period of 1-3 years after completion of the pre-sessional programme. The pre-sessional programme is intended for students whose IELTs score is slightly below that of the requirements for their future post-graduate course (usually an overall score of 6.0 or 6.5). The pre-sessional course at the University of York focused mainly on academic writing and speaking and the requisite language and organisational skills.

In the current study, the learners' proficiency was determined using previously attained IELTs scores, performance on a C-test, and self-rated proficiency. Three measures were used in order to better ensure that the learners were of a similar proficiency. The learners completed a language background questionnaire which determined their experience learning English and asked them to self-rate their proficiency level in the four core skills (reading, writing, speaking and listening) (see appendix 3). Self-rating has been found to correlate with proficiency measured by independent tests (Delgado et al., 1999; Flege et al., 2002; Jia, Aaronson, & Wu, 2002). The learners had attained a mixture of IETLS 6-7 (mean = 6, SD = .42), a quarter had completed post-graduate study, and they had been learning English for between 7 and 20 years (mean = 12 years, median = 12 years). The c-test used was adapted by Spada et al. (2015) who adapted it from Ishihara, Hiser, & Okada (2003) (see appendix 4). This test was a combination of a C-test and a cloze test in that the first half of each target word remained and the second half of the word was completed by the learners (as in a c-test) but the words were not removed strictly systematically (for example one in every three words) but were removed at a rate of between the fifth and eighth word (as in a cloze test). C-tests and cloze tests have been found to provide an accurate indication of proficiency and correlate well with proficiency scores found on standardised tests (e.g. IELTS) (Chapelle, 1994; Dörnyei & Katona, 1992; Eckes & Grotjahn, 2006; Jafarpur, 1995; Tremblay, 2011). The questionnaire and C-test were administered using Qualtrics (Qualtrics, 2005).

The mean score on the c-test was 63%, SD = 12. The c-test results show that the learners were not near-native, as NSs would score 92% or higher (as found by Ishihara, Hiser, & Okada, 2003). The c-test was able to reveal more variation in the learners' proficiency than that suggested by the IELTs scores, as the scoring was more fine-grained.

Learners were also asked to self-rate their ability in the four communicative skills (speaking, listening, writing and reading). There was a great deal of variety in the scoring, as indicated by the mean scores given to each skill (see table 1 below).

Taken as a whole the three proficiency measures indicated that the participants were around CEFR level B2 or upper-intermediate. At this level the learners would be expected to be familiar with the passive voice but would not be using it without error, or at a near-native level (as discussed in 2.6).

Table 1. Language background and proficiency of Chinese participants.

	Chinese learners (M, SD, range)
Years studying English	12 (3.17) (7-20)
Years residency in the UK	.37 (.50) (< .5-4)
Self-rating in communicative skills ((1 = your best skill, 4 = your worst skill)	
Speaking	2.71 (.90) (1-4)
Writing	2.68 (.89) (1-4)
Listening	2.21 (.82) (1-4)
Reading	2.06 (.90) (1-4)
Experience teaching English	Yes = 42, No = 31
Overall IELTS	6.00 (.42) (5.5-8.5)
C-test score	63% (12) (32%-81%)

The learners were randomly assigned to: 1) EI + cue focused practice (*n* = 25), 2) EI + noun focused (*n* = 25) and 3) test-only (*n* = 23). Randomisation was important to ensure effects were attributable to training, and not other uncontrolled variables (Torgerson & Torgerson, 2001).

29 NSs were also tested. 29 completed the eye-tracking to provide a comparison with the learner data, 10 of these also completed the other outcome measures to give some indication of test reliability. The NSs were recruited on a voluntary basis using convenience sampling. The NSs were not bilinguals (they had not spoken any language other than English at home before the age of five). They would have had experience of a basic level of foreign

language education (likely French, Spanish or German) in their secondary schools. They were all speakers of British English. 18 were university students studying psychology undergraduate degrees. 14 were members of the public (not currently students), all had studied to at least undergraduate degree level in the past.

3.4 Small-scale exploratory study

Prior to the main study, a small exploratory study was carried out to investigate Chinese learners' knowledge of the passive and active voices in English in the target population (Chinese L1 studying in English at UK universities). The aims were to ascertain whether instruction on this construction would be useful and to inform training and test design for the main study. The study and its findings, and how they informed the materials design for the main study are summarised here.

Participants were 24 Chinese-English bilinguals aged 21-30 studying on MA courses at The University of York (n = 19) and at The London School of Economics (LSE) (n = 5). Participants had studied English for a mean of 14.48 years (SD = 3.62) and had been in the UK for mean of .63 years (SD = .52). Their proficiency was determined by IELTs score (mean = 6.90, SD = .38). None of these participants took part in the main study.

Participants completed a vocabulary test, a picture-interpretation task and a JT. All tests were administered remotely using Qualtrics (Qualtrics, 2005), and participants were sent the link via email. Participants had to complete the tests in one sitting and could not repeat items.

The test items in both the picture-interpretation task and a JT manipulated agent-patient roles and, for the JT, inserted errors. Since animacy has been found to affect online processing of L2 input, in particular for Chinese learners of English (see section 2.6.3), animacy was manipulated to explore whether sentences with animate agents and patients were interpreted differently to those with inanimate agents and patients.

3.4.1 Vocabulary test

A vocabulary test was given to participants a week prior to the picture-interpretation task and JT to ensure that none of the vocabulary items in the tasks would cause a problem for interpretation. The nouns that would be used in the tests were tested in a picture choice task, in which the participants were presented with a word and the choice of two images and asked to choose the image matching the word. The verbs were tested by giving the learners the verb in Chinese and four options in English, the learners had to select the matching English translation. Four options were given to reduce the role of chance. The Chinese translations were provided by a NS of English with C2 proficiency in Chinese (level 6 HSK[5]) (see appendix 5 for English word list).

3.4.2 Picture decision task

The picture decision task used images and lexical items tested in the vocabulary test. The items were all checked for plausibility by five NSs. 10 critical items were created with animate subjects and animate objects (see example 13a) and 10 critical items with inanimate subjects and inanimate objects (see example 13b).

13a. The boy is pushed by the girl

[5] Hanyu Shuiping Kaoshi - International standardized test of Chinese language proficiency

13b. The car is hit by the bike

These 10 animate-animate and 10 inanimate-inanimate items were used to create 10 passive versions and 10 active versions (see examples 14a and b) – with five of each type. The active versions served as distractors.

14a. The horse is frightened by the dog

14b. The dog frightens the horse.

From this set of 20 items, 10 different passive sentences (five inanimate-inanimate and five animate-animate) and 10 active sentences (five inanimate-inanimate and five animate-animate) were selected for the experimental list. These sentences were recorded and were presented alongside two cartoon images constructed to show either a passive or active interpretation of the stimulus (see 15 below for an example).

15. Audio stimuli: *The child is carried by the dog* [or in another list: *The child carries the dog*]

a) 

b) 

These were pseudo-randomised and presented alongside 20 filler sentence which were all in the active voice and constructed using a variety of grammatical structures (past simple, present perfect, present simple and present progressive). Unlike the critical items, the filler images presented did not all focus on the subject-object relationship in the stimuli but focused on the lexical items used. They were also not semantically reversible (see examples 16a and b). The ratio of critical items to fillers was 1:1 and they were presented alternately (see appendix 6 for stimuli).

16a. *The boy does the homework* (images distinguish the subject)

a) 

b) 

16b. The woman is driving the car (images distinguish the object)

a) 

b) 

3.4.3 Written untimed acceptability judgement test (JT)

The JT contained 18 critical items (passive) and 18 distractors (active). Participants were asked to judge the sentences as grammatical or ungrammatical and to correct those judged ungrammatical. Half the sentences had animate subjects and objects, and half inanimate subjects and objects. Varied tenses were included – present perfect, present progressive, and past simple. In order to investigate complexity, half the stimuli contained only direct objects (*k*=18) and half contained a direct and indirect object (*k*=18). To find out whether type of error would affect judgements, three error types were included, following Spada et al. (2015) and Li et al. (2016). These were missing auxiliary verb (*k* = 12) (see 17), base form of the main verb (*k* = 12) (see 18) and gerund form of main verb (*k* = 12) (see 19) (see appendix 7 for stimuli).

*17) The child * being helped by the teacher*

18) The dog is being walk by the boy*

19) The person is being following by the police*

12 present perfect active sentences (see example 20 below) were also included to investigate whether judgements were affected by a more complex structure featuring an auxiliary and past participle in the active voice. The same error types were included. There were equal numbers of animate and inanimate sentences (*k* = 6) and direct and indirect object sentences (*k* = 6). This would indicate whether the learners experienced problems with passive structures specifically, or whether the auxiliary and past participle in other structures would also cause problems.

20) The dog has eaten the rabbit

3.4.4 Results

All the learners scored 99% on the vocabulary test. This indicated that the items would not cause interpretation issues, and that the images used were likely to depict sufficiently well the nouns they were intended to.

3.4.4.1 Picture decision test

The data from the picture decision task and JT data were not normally distributed, so the results were analysed using Wilcoxon's signed ranks test.

On the picture decision task, the mean score on the active items (80%, SD = 9) was significantly higher than that of the passive items (62%, SD = 11, z = -3.47 p = .00). Furthermore, the size of the effect of voice on score was large (d = 1.79, CIs =1.15, 2.42) This suggests that the agent and patient roles were more difficult to decipher for the passive items (21a) than for the active items (21b).

21a. The horse is frightened by the dog

21b. The dog frightens the horse.

Score also was affected by animacy. Mean scores on animate-animate stimuli were significantly higher than on inanimate-inanimate stimuli regardless of voice (active or passive) (z = -4.21, p = .00, d = 3.98, CIs = 3.06, 4.90).

The strongest effect was that of animacy in the passive voice with inanimate-inanimate scoring significantly lower than animate-animate (z = -4.20, p = .00, d = 2.84, CIs = 2.09, 3.60). Voice also had an effect with regard to inanimate stimuli with passive inanimate-inanimate sentences interpreted significantly less accurately than active inanimate-inanimate (z = -4.12, p = .00, d = 1.55, CIs = .95, 2.16). There was no statistically significant effect of voice for animate stimuli (z = -2.31, p = .021, d = 1.41, CIs = .82, 2.01) or of animacy for active stimuli (z = -2.15, p = .32, d = 3.18, CIs = 2.38, 3.99), although the effect sizes for both suggest effects were evident.

In sum, the data showed that passive structures were indeed more difficult for these learners when determining agency, and particularly when agents and patients were inanimate.

3.4.4.2 Written untimed acceptability judgement test (JT)

The mean score for the JT was high (86%, SD = 9.45). The total score was determined by calculating a global average of correct judgements (i.e. hits and rejections). Hits were defined as correctly identified grammatical items. Correct rejections were ungrammatical items both correctly identified and subsequently corrected. The JT included passive items (k = 36) and present perfect active items (k = 12). Passives items were judged correctly more than the present perfect items; passive items had a mean percentage score of 88.57% (SD = 9.46), present perfect items had a mean percentage score of 73.64% (SD = 16.19) and this difference was found to be significant in a Wilcoxon's signed ranks test (z = - 3.04, p = .00).

Overall, the results of this exploratory study suggested that animacy should be manipulated in the main study. However, the picture-interpretation task only investigated animate-animate and inanimate-inanimate agent patient roles. The current study intended to expand on these findings by including animate agents combined with inanimate patients and the reverse; inanimate agents with animate patients. Furthermore, tense appeared to play a role in interpretation. Therefore, the main study would include items in the present perfect passive and active voice.

The mean proficiency of the participants in the main study (mean IELTS = 6.00, SD = .42) was lower than the proficiency of participants in the exploratory study who were already enrolled on MA courses (mean IELTS = 6.90, SD = .38). Despite a higher proficiency level, since the exploratory study participants found the passive more difficult to interpret than the active sentences, and were affected by animacy, it was expected that training would be of benefit to the participants in the main study who were of lower proficiency.

3.5 Design of the main study

The study was a pre, post-, delayed post-test intervention with Chinese learners of L2 English. The NSs completed only one test session to provide a baseline for comparison. The learner participants were randomly assigned to one of three groups: EI + cue focused practice, EI + noun focused practice, and a control (test-only) group.

3.5.1 Pilot study

A small-scale pilot was run to test the intervention and test materials and to allow some preliminary analysis. Three participants took all of the tests and the interventions and gave feedback on the tasks. Subsequently the tasks were improved based on any issues observed by the researcher or the participants. Following these changes, four participants carried out another small pilot. Two participants took the cue focused intervention and two the noun focused. All took the pre- and post-tests. The following sections provide a summary of the pilot.

3.5.1.1 Participants and design for the pilot study

The participants were all Chinese learners of English enrolled on a pre-sessional course at the University of York. The participants had been living in the UK for less than six months and had a mean IELTs score of 6.5. This makes this sample comparable to the population that was used for the main study. The pilot study followed a pre-test, intervention, immediate post-test design:

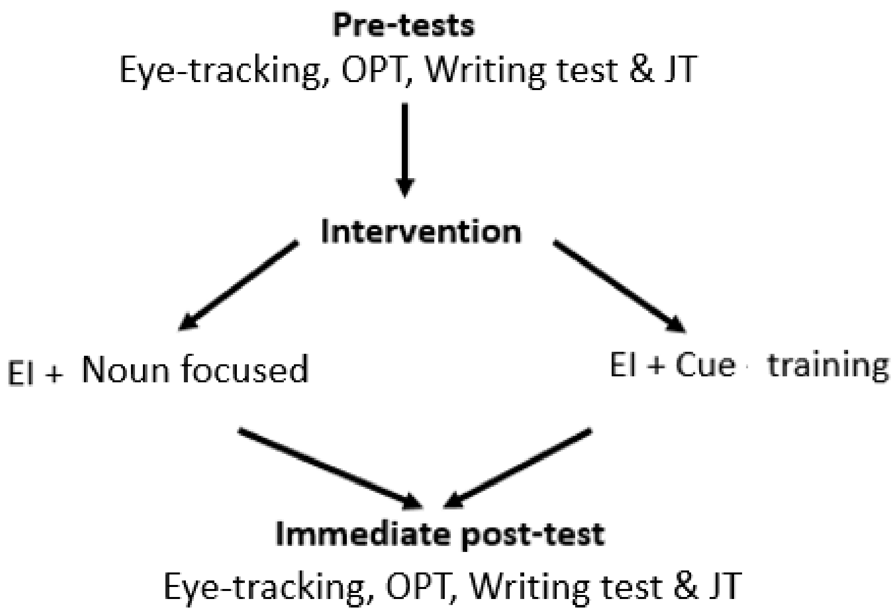


Figure 9. Pilot test design

3.5.1.2 Implications from the pilot for the design of the main study

3.5.1.2 a) The interventions

Both interventions were piloted. Participants completed one session, either cue focused, or noun focused (the main study would include two sessions). The aim of piloting a partial intervention was to check the stimuli and the tasks, and to check for any issues with the format, instructions, and items. Some minor changes were made to the wording of the instructions and to some images that appeared ambiguous. For example, the addition of arrows to show direction of action. The major change made to the intervention was the addition of more stimuli. Both intervention sessions (cue focused and noun focused) were completed quickly (10-15 minutes). A lack of exemplars and practice in the training might have reduced the ecological validity and reduced the chances of observing any learning. The number of items was doubled, from 144 items to 288. This was done by reversing the agent and patient roles in the original items and having them appear in a different modality in each session. For instance, an item appearing in session one task one in the aural modality would have its roles reversed and added to session two task two in the written modality.

3.5.1.2 b) Outcome measures

The pilot mainly highlighted the need for clear instructions and examples. The eye-tracking instructions were broken down into more steps and the number of practice trials was increased from two to five. For the production tests, the leaners' scores showed that the correct tense was produced in only 45% of the items in the pre-test. This may have been because the tense cue (*now* for present simple or *since this morning* for present perfect) was not noticed as much as the other cues. To prevent this, and increase the salience of the other cues, green boxes were inserted around the question and verb cues, and double underlines were added to the tense cue. The instructions also better highlighted the cues. The scoring of the production tests in the pilot involved assigning one point for each of the following:

- Grammatical - was the response grammatical.
- Target-voice attempted - was a passive or active attempted.
- Correct verb form - was the correct past participle or gerund used.
- Correct tense.

This scoring did not take into consideration the use of *by*. Given that *by* was one of the cues targeted in the training, its usage required assessment in the tests. Therefore, scoring for *by* was added to the final scoring criteria.

A key lesson from the pilot was that the participants, who had an IELTS score comparable to that of the final study sample (mean 6), did not perform at ceiling in the pre- or post-tests. This suggested, albeit tentatively due to the small sample size, that the learners in the final study would have some room for improvement in their use of the passive.

3.5.2. Experimental procedure in the main study

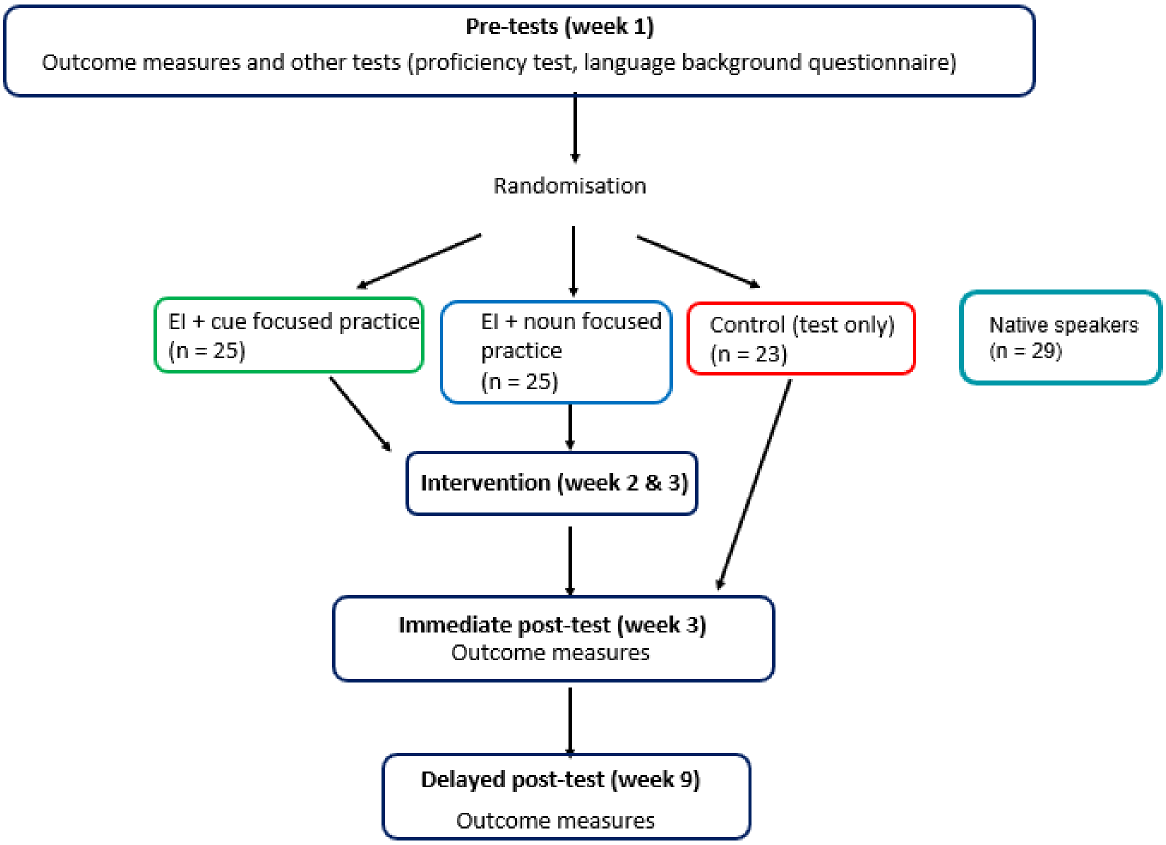


Figure 10. Experimental procedure

The proficiency test, language background questionnaire and pre-test were administered one week prior to the intervention. The intervention consisted of two sessions, once a week over a two-week period. Each session was approximately 30 minutes and was administered individually in experimental conditions. The immediate post-test followed directly after the last intervention session. The delayed post-test was administered six weeks after the final intervention session.

Care was taken that each session fell at exactly the intended interval (i.e. the first intervention 14 days after the pre-test, each intervention with similar spacing, and the tests the same amount of time after the interventions). This was important because, for example, the spacing between practice has an effect on performance in a future test. The longer the gap between the end of practice and testing, the longer the gap between practice sessions should be for optimal information retention (Cepeda et al., 2008). Also, sleep is known to affect learning (e.g. Tamminen et al., 2010; Weighall et al. 2016). Therefore, if some learners had substantially longer between practice sessions compared to others their ability to learn might differ. However, there were some occasions when this homogenous spacing between sessions was not possible due to participant cancellation. In these cases, the session was rescheduled for no later than two nights after the intended date (i.e. 14 days + one or two nights).

3.5.3 The interventions

3.5.3.1 The target features

Both interventions ([morphosyntactic] cue focused and noun focused) consisted of two sessions focusing on the present simple passive voice and the present progressive active voice (see 22a and b).

*22a. The boy is push**<u>ed by</u>** the girl.*

*22b. The boy is push**<u>ing</u>** the girl.*

*23a. The child is carr**<u>ied</u>**.*

*23b. The child is carry**<u>ing</u>**.*

Each session included passive sentences with and without *by* in both aural and written modalities. Each session was divided into two sections so that the first half practised reduced passives (without *by*) and the second half introduced full passives (with *by*) (see 23a and 22a above). Reduced passives were presented first in the training in order to focus only on the verb ending cue (-*ed* versus -*ing*). In the second half, the practice on full passives introduced *by* and EI was given on the use of *by*.

The reduced passives were contrasted with active present progressive sentences without an object. Some present progressive sentences without an object can be somewhat unusual in isolation (if the verb is normally transitive) (see 23b) but they were used to provide the contrast of verb morphology between passive and active sentences. The stimuli included both regular and irregular verbs When designing the stimuli, efforts were made to use equal numbers of regular and irregular verbs, but it was difficult to make enough stimuli to balance the different types of animacy combinations and to maintain equal numbers of verb types. Therefore, the numbers of regular and irregular verbs were slightly different (27 regular, 32 irregular).

3.5.3.2 Intervention materials

Both the morphosyntactic cue focused (henceforth ‘cue focused’) practice group and the noun focused group received two intervention sessions and the same accompanying EI prior to the practice items and main tasks (as in figure 11).

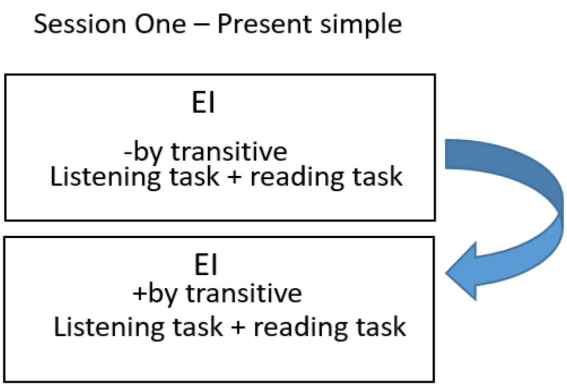


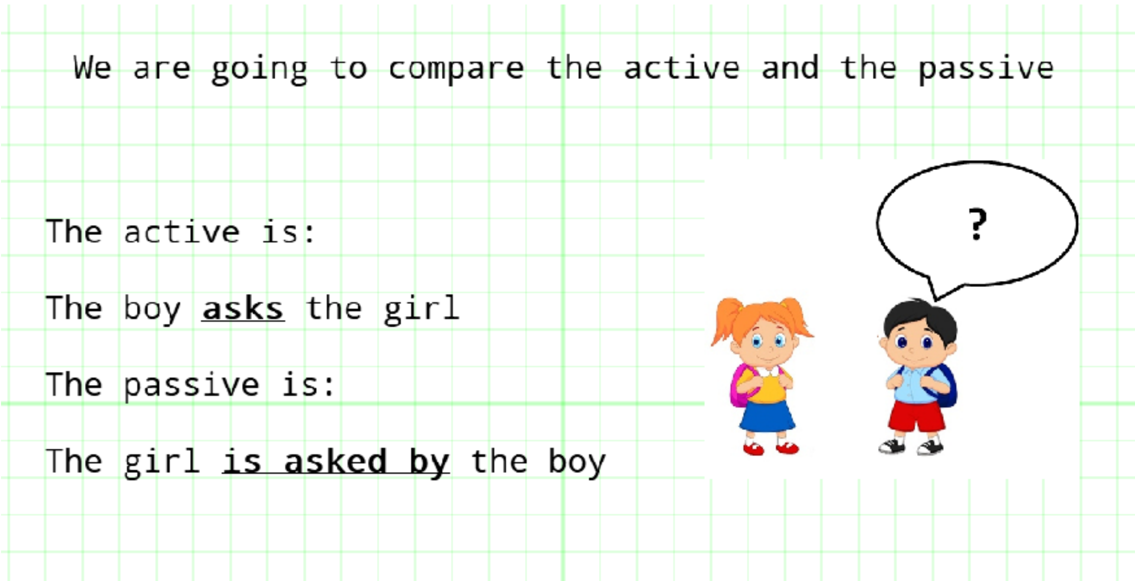
Figure 11. Example session flow

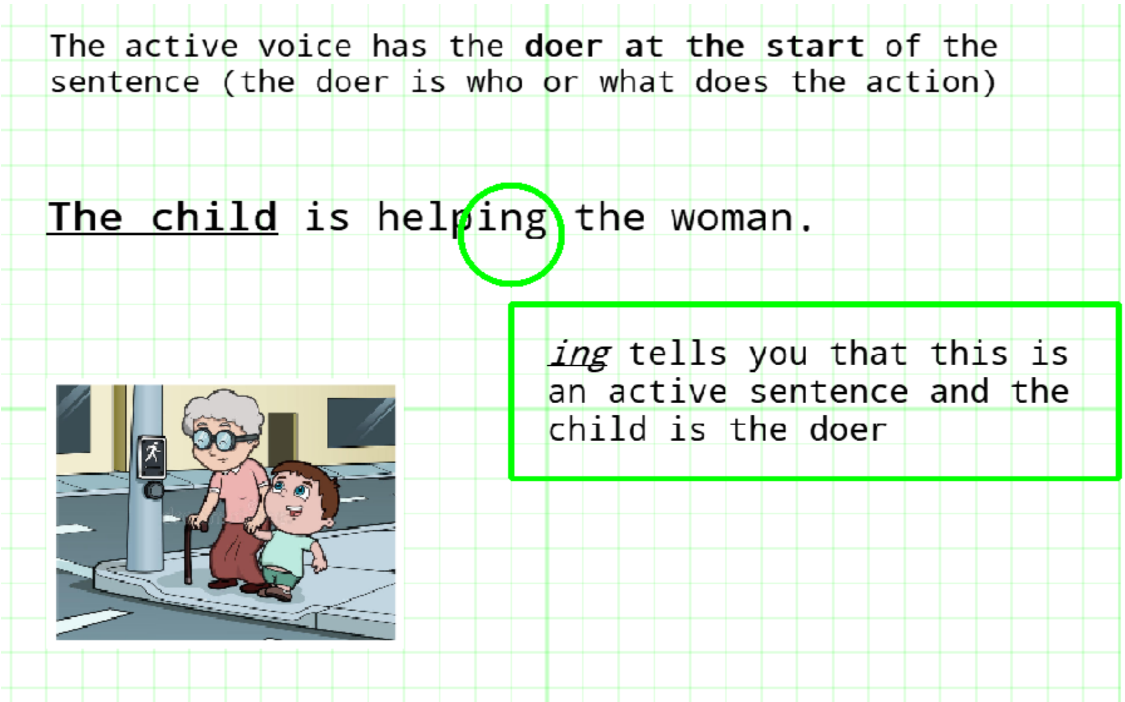
The cue focused group received extra EI in the form of aural explanations during the practice trials prior to each task. Both groups received correct / incorrect feedback after each item but only the cue focused group received metalinguistic feedback on how to use morphosyntactic cues to determine agent-patient roles.

The intervention was delivered using Open Sesame, an 'An open-source, graphical experiment builder for the social sciences' (Mathôt, Schreij, & Theeuwes, 2012). Open Sesame records RTs and accuracy rates. The RTs and accuracy scores recorded in both interventions were analysed to investigate whether participants were more accurate and quicker over time. This would allow for gains during the interventions to be recorded, as well as gains seen between the pre- and post-tests. Open Sesame also allowed the trials in each intervention task to be presented randomly for each participant. This prevented the order of the items influencing performance.

3.5.3.2 a) Explicit information (EI) in both treatment groups

The EI provided was identical for both treatment groups. The EI described and illustrated the morphological and syntactic cues that learners can use to identify the passive voice by contrasting these cues in the passive voice with the present progressive (active). The agent-patient relationship and its associated morphology was highlighted (see figure 12 and 13 below for examples from session one).





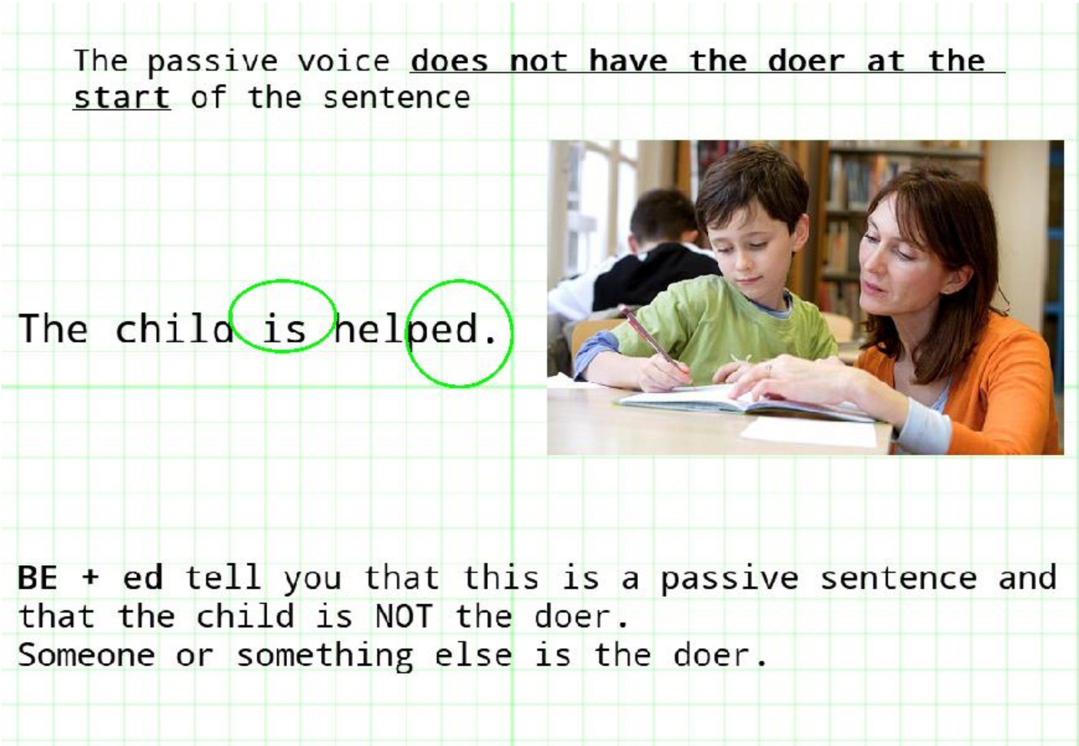


Figure 12. EI for present simple passive (without by*) and active.*

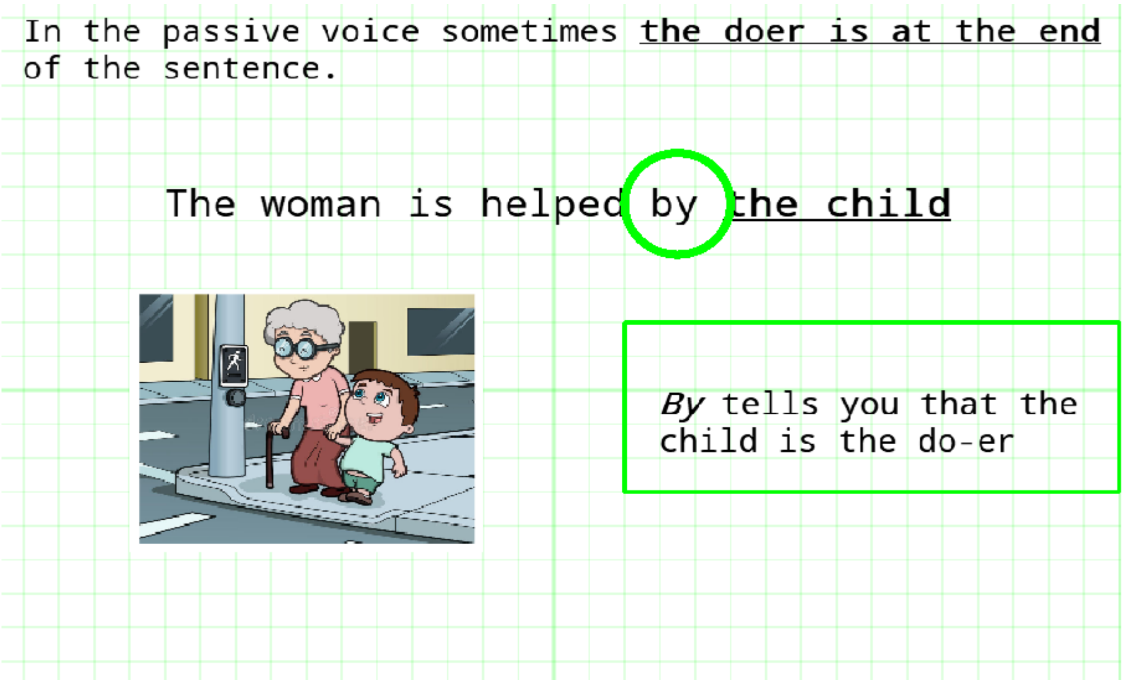


Figure 13. EI for present simple passive (with by*)*

The EI was provided four times, in both sessions, prior to the practice for reduced passive (without *by*) and prior to the practice of full passive (*with by*).

The participants could read the EI for as long as they chose, they then pressed any key to move onto the practice items. Both groups received the same number of practice sentences at the start of each task (after the EI).

3.5.3.2 b) Extra cue instruction for the cue focused practice group.

Further to receiving the EI provided at the start of each subsection, the cue focused practice group also received further information regarding cue use. The cue focused practice group heard, during the practice items, an extra aural explanation reminding them to use the cues highlighted in the EI before the tasks (for example see figure 14). This was intended to reinforce the cues that would assist with the practice activities. The image below shows the wording and timing of the extra information. After each practice item the participants saw the practice item again along with added audio (as exemplified in 14). Each segment (e.g. “the” or “boy”) appeared on the screen along with the corresponding audio (shown in speech bubbles) to walk the participant back through the item they had previously encountered.

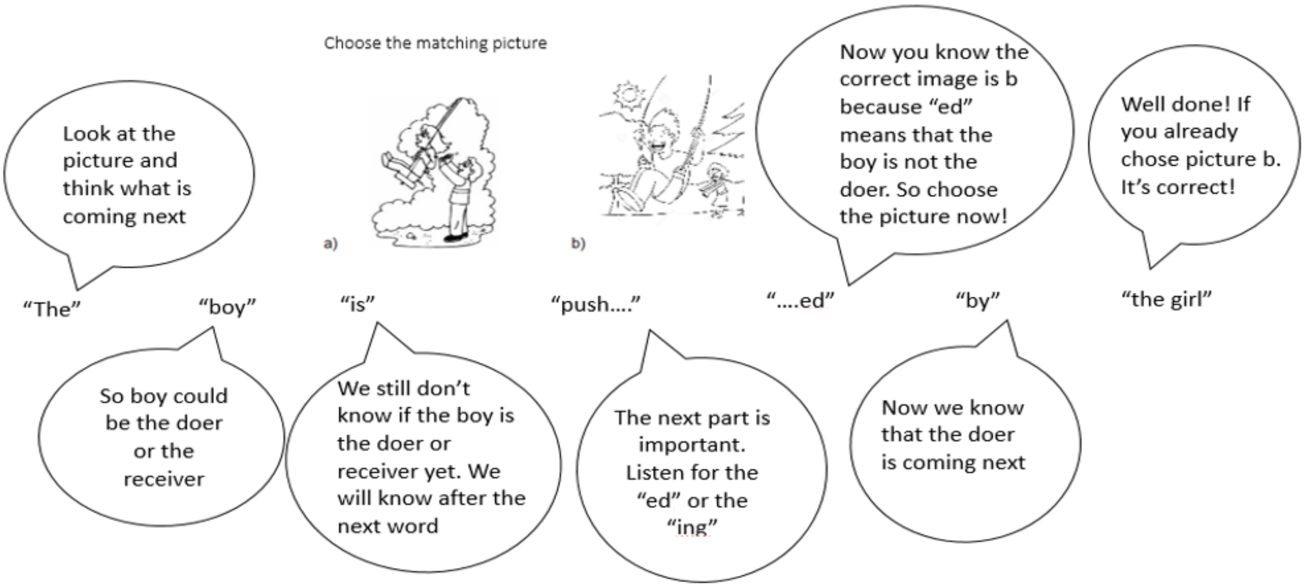


Figure 14. Example of commentary for the cue focused practice group

3.5.3.2 c) Intervention activities

Both the cue focused group and the noun focused group saw the same passive structures and the same quantity of exemplars. The manipulation was as follows: The cue focused practice activities were designed to force participants to use the morphosyntactic cues indicating the passive or active voice (i.e. *-ed* and *by,* and *-ing*). The noun focused training did not focus on these cues but presented the same number of active and passive exemplars but in tasks focusing learners' attention on either the patient or the agent, that did not require noticing of morphosyntactic cues to answer correctly. Half of the activities for both conditions were in a written modality (no audio stimuli) and half were aural (no written stimuli). All the activities contained images. Both modalities were included to increase the likelihood of the learners noticing the cues and to give them the opportunity to establish representations of both the written and oral forms.

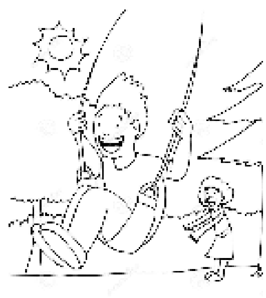
Below is an example of a cue focused practice task and its corresponding noun focused task (figure 15). In the first example participants saw the two images and heard a sentence broken up into segments. The images depict reversible agent-patient relationships so that the participants could not determine the correct image based on the nouns in each item. Between each segment was a 300-millisecond delay. A pause was inserted between the stem of the main verb and its ending (*-ed* or *-ing*) dividing it into two segments. This instructional device forced the listener to attend to the verb ending. This was also done to cause 'surprisal' when the learners encountered the verb inflection denoting a past participle (after potentially expecting to hear the more common gerund and active interpretation).

The instructions for this task were to "***hear*** *a sentence and choose the correct image* ***as soon as possible.***" The instructions were intended to encourage the participant to listen out for the verb ending and use this information to choose the correct image, prior to hearing the final noun (the agent, in passive sentences).

Cue focused - Choose the matching picture

"*The boy………is………push………ed………by………the girl*" (correct answer b)

a) 

b) 

Noun focused - Choose the matching picture

"*The boy is pushed by the girl*" (correct answer *b*)

a) 

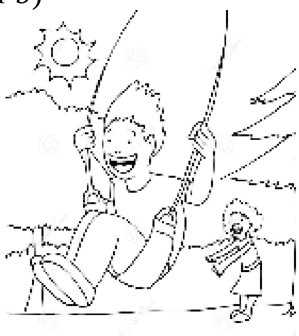
b) 

Figure 15. Example of a cue focused practice activity and the corresponding noun focused activity

The noun focused stimulus was not divided into segments (as in the second example above). The patient differed in each of the two images so that the participants were able to determine the correct image by paying attention to the first noun only (i.e. after hearing '*boy*' they knew that the correct answer was 'b' as there was no boy in picture a). Half of the items in the noun focused tasks could be answered by paying attention to the first noun (as in 15 above) and half to the final noun (see figure 16 below).

Choose the matching picture

"*The cat is paying the dog*" (correct answer b)

a) b)

Figure 16. Example of a noun focused activity focusing on final noun

The noun focused participants were also encouraged to answer as quickly as possible. Their instructions were the same as for the cue focused group; "***hear*** *a sentence and choose the correct image* **as soon as possible."**

3.5.3.2 c.i) Cue focused instruction activities

The following four task types appeared in both of the intervention sessions (one and two) in the same order (one, two, three and four). This was to ensure that exposure to different passive and active types was the same over time i.e. reduced and full sentences were seen along with EI one week before, and immediately before, the immediate post-test. Each session included four tasks with 128 trials. Four practice trials were given prior to each task (16 over the session) with extra instruction in using the verb ending cue (see figure 14), this made the total number of sentences participants were exposed to in each session 144, totalling 288 exemplars (144 passive, 144 active) over the two intervention sessions. Each session lasted around 30 minutes. Tasks one and three were in the oral modality; tasks two and four in the written modality.

Tasks one (oral) and two (written) (Both without 'by' practice)

Task one. This required learners to look at a picture and listen to a sentence stem followed by two possible final segments. Tasks one and two forced attention to the morphosyntactic cues (*-ed* and *-ing*). A pause was inserted between the noun and the verb to give the learners time to pre-empt the upcoming verb inflection. Participants had to choose the final segment to match the picture, for example, in figure 17 below.

"The child is ……" (correct answer a)

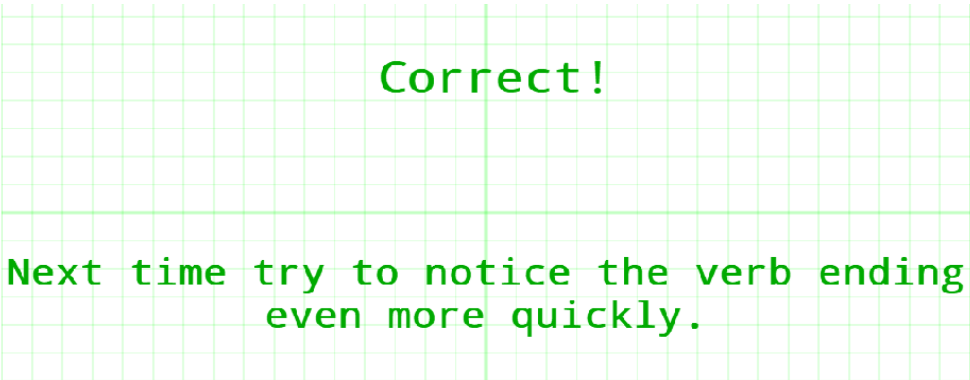
a) *"Carried"*
b) *"Carrying"*

Aural feedback if incorrect: "*The elephant is DOING the asking so the lion IS NOT the doer. We need to choose the lion is... ASKING".*
Feedback if correct: "*CORRECT! WELL DONE"!*

Figure 17. Example cue focused task one trial and feedback

The participants indicated the correct segment by pressing "z" on the keyboard for the first segment they heard and "m" on the keyboard for the second segment. Half the pictures corresponded with a passive final segment and half with an active (present progressive). In half of the trials, the correct answer was heard first and in half the correct answer was heard second. There were 32 trials (16 passive, 16 active). Animacy was balanced across trials so that eight trials had animate subjects and objects (four passive, four active), eight trials had inanimate subjects and objects (four passive, four active), eight animate subjects with inanimate objects (four passive, four active), and eight inanimate subjects with animate objects (four passive, four active). Yes / no feedback was given for correct and incorrect responses. In addition, audio feedback was given if the response was incorrect. This feedback highlighted the verb ending and the agent-patient roles (see 17 above). This metalinguistic feedback was given as research has suggested this can be more beneficial than only yes / no feedback (Carroll & Swain, 1993; Dracos, 2012; Ellis et al., 2006).

Participants were encouraged to select the ending as quickly as possible and were given accuracy and speed feedback at the end of the task (see figure 18 below). This was intended to encourage them to respond as soon as they heard the correct ending (-*ing* or the past participle ending), with a view to serving any automatisation process that may be happening.



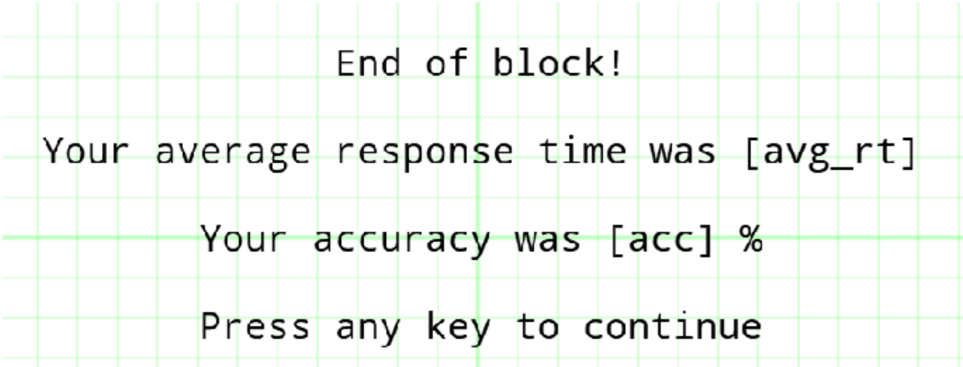
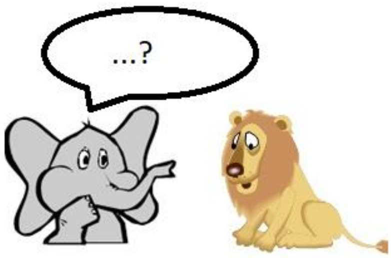


Figure 18. Example cue focused feedback to encourage speed

Task two. This task also required the participants to choose the end of a sentence by selecting the ending that matched the picture, but the input was written. Participants selected the correct answer by pressing “z” on the keyboard for the left-hand ending and “m” for the right-hand ending, for example in figure 19 below.

The lion is……. (correct answer a)



a) *asked* (press Z) b) *asking* (press M)

Feedback if incorrect: *The elephant is DOING the asking so the lion IS NOT the doer. We need to choose the lion is... ASKING.*
Feedback if correct: *CORRECT! WELL DONE!*

Figure 19. Example cue focused task two trial and feedback

As with task one there were 32 trials, the position of the correct verb was balanced across trials and conditions and there were the same proportion of passive (16) and active trials (16) and types of animacy as in task one. This task was designed to draw attention to the written verb ending. Written feedback was given for incorrect responses to highlight the agent-patient roles and the verb cue (see figure 19). Overall accuracy and speed feedback were also given at the end of all trials (as in figure 18 above).

Tasks three (oral) and four (written) (Both with ‘by’ practice)

Task three. In task three the cue *by* was introduced. In this task, attention was drawn to the verb ending and *by* using audio input with a 300-millisecond delay inserted between each segment to emphasise and isolate the verb base and the ending. For example, the sentence “*the boy is pushed by the girl*” had a gap between each word and *push* and *-ed*. This drew attention to the *ed* and *by* (see figure 20 below) resulted in ‘surprisal’ if the learners were expecting an active sentence. In this task, the participants chose one picture, from two, that matched the audio. Participants heard the sentence and pressed “z” to choose the left-hand picture and “m” for the right-hand picture.

“*The boy.....is……….push……..ed………..by………..the girl*.” (aural) (correct answer b)

a)

b)

Figure 20. Example cue focused training task three trial

As with the previous tasks there were 32 trials, 16 active and 16 passive. Animacy was balanced across the trials as in the previous tasks. The correct image appeared on the left in half the trials (eight passive and eight active) and on the right in half (eight passive and eight active). Feedback was given after each trial to inform participants if their response was accurate. If the response was incorrect, aural feedback was given (see 19 above). the target was presented on screen and the cues were highlighted e.g. “*the boy is push**ed by** the girl*”. Feedback for overall response time and accuracy were again given after all trials.

Task four. Task four was similar to task three except that the segments were provided in written form. There were 32 trials, with voice, animacy and correct image position balanced as in the other tasks. The participants saw passive and active sentences divided up into segments appearing for 1000ms per segment. Using the “z” and “m” keys participants selected a picture matching the sentence they saw, for example in figure 21 below.

“*The dog…….is………....chas……..ed……....by……....the boy*.” (written) (correct answer b)

a) b)

Figure 21. Example cue focused training task four trial

Participants were able to select the picture after any segment and were given written feedback informing them if their choice was accurate or not. Feedback rewarding the participants for speed was given if they were able to select the correct image prior to seeing *by*. This was to discourage reliance on hearing the second noun before deciding the correct images and therefore encourage cue use to assign roles (retrospectively for the first noun, and predictively for the second noun). If the learners provided an incorrect response, they saw the remainder of the sentence and then received feedback pointing out the correct answer and highlighting the verb ending. Feedback was also given if the participants chose the picture too early to be accurate, i.e. after the first and second segments (the first noun and the auxiliary *is*). This reminded participants to pay attention to the target cues. Overall accuracy and speed feedback were given at the end of the task (as in figure 18).

3.5.3.2 c.ii) Noun focused intervention activities

The noun focused activities were preceded by the same EI as in the cue focused practice (see section 3.5.3.2a). The noun focused activities were designed to expose the participants to the same number of exemplars as the cue focused group but without drawing attention to the morphosyntactic cues highlighted by the tasks in the cue focused training. The same set of target images were used for both interventions (distractor images differed). If participants chose incorrectly in any of the tasks, they were told the answer was incorrect, but no explanatory feedback was given (unlike the cue focused training in which metalinguistic feedback was given). Each session lasted around 20 minutes; this was slightly shorter than the cue focused practice due to the nature of the tasks.

Tasks one (oral) and two (written) (Both without 'by' practice)

As in the cue focused practice, tasks one and two focused on the reduced passive, i.e. without *by* and the second noun phrase, by contrasting it with an active present progressive (without an object). In the first task participants saw two pictures and heard a sentence matching one of the images (see figure 22 below). In order to correctly choose the matching image in this task participants needed to attend to the agent or patient in the sentence and not the verb morphology. The second task was the same as task one, but the stimuli were written rather than aural. In task one and two, as in the cue focused practice, there were 32 trials, 16 active and 16 passive. Animacy was balanced across the trials. The correct image appeared on the left in half the trials (eight passive and eight active) and on the right in half (eight passive and eight active).

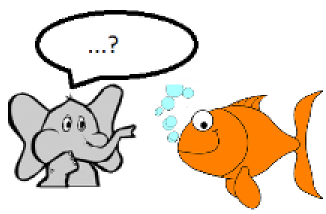
"*The dog is carried*." (aural stimuli) (correct answer b)

a) 

b) 

The lion is asking (written stimuli) (correct answer a)

a)



b

Figure 22. Example trial from noun focused task one (aural) and two (written)

Tasks three (oral) and four (written) (Both with ‘by’ practice)

As in the cue focused training, tasks three and four introduced passives with *by* contrasted with full (i.e. transitive with object) present progressive active sentences. Tasks three and four were the same format to tasks one and two (see figure 23).

“*The boy is pushed by the girl*.” (aural input) (correct answer b)

a) 

b) 

The dog is chased by the boy. (written input) (correct answer b)


a)


b)

Figure 23. Example trial from noun focused tasks three (aural) and four (written)

In both tasks three and four, as in the other tasks, there were 32 trials, 16 active and 16 passive. Animacy was balanced across the trials. The correct image appeared on the left in half the trials (eight passive and eight active) and on the right in half (eight passive and eight active). As in the cue focused practice tasks, RTs and accuracy were recorded, and feedback about accuracy and speed was given to the participants (as in figure 18). The full stimuli list used in the interventions is in appendix 8.

3.5.4 Outcome measures

Outcome measures were adapted from the exploratory study; the picture decision task (an eye-tracking task in the main study), a written untimed acceptability JT, and additionally a written test and an oral production test.

Order of the tests. The eye-tracking test was intended to elicit principally implicit knowledge of the passive and active voice (at least in the pre-test when participants had not yet been exposed to training). The JT was intended to test principally explicit knowledge of the passive and active voice. For this reason, the eye-tracking test was the first test in the battery and the JT was the last test. The eye-tracking was first so as to prevent the other tests resulting in awareness of the target structures at pre-test, and so resulting in possible activation of explicit knowledge. The eye-tracking test was followed by the oral production test and written test. The oral production test followed the eye-tracking test since oral production tends to allow for less planning during performance than written. The written test was assumed to activate explicit knowledge more than the oral production test and the eye-tracking. Therefore, the written test followed the oral production test and eye-tracking. As mentioned, the JT was the final test (see figure 24 below for test order and duration).

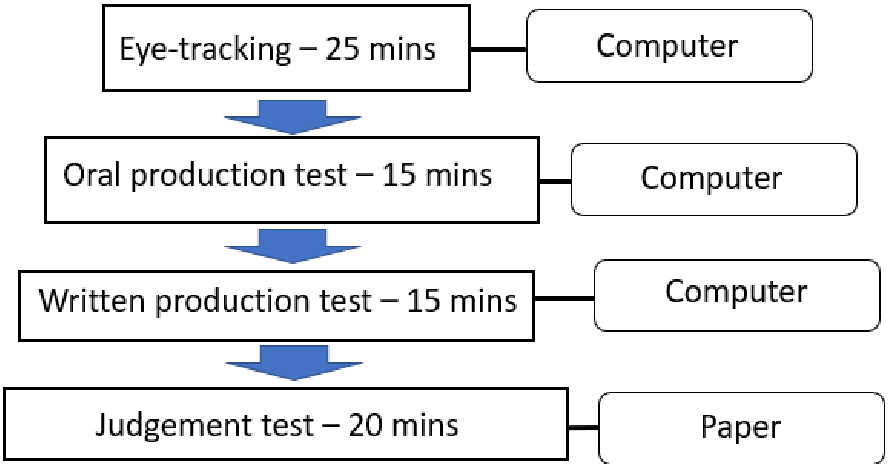


Figure 24. Order, timing and medium of outcome measures

3.5.4.1 Visual world eye-tracking

3.5.4.1 a) Stimuli design

Audio stimuli – critical items

In reporting the design and procedure of the visual world eye-tracking task I endeavoured to meet the minimum requirements for reporting where possible in order to increase transparency and reproducibility (as suggested by Fiedler et al., 2019) The visual world eye-tracking task used images and lexical items based on the exploratory study. Items were designed to have ambiguous agent-patient relationships up until the point of the verb ending (*-ed* (or other past participle) / *-ing*). For example; *The boy is pushed by the girl.* A listener was not able to determine the agent-patient relationship based on semantics or word order (first noun) since both agent-patient roles were possible and grammatically correct. Eye-tracking was used because it allowed for a fine-grained measure of the point during input at which participants were able to interpret the agent-patient relationship after the point of disambiguation. For example, from the offset of the verb (see segments in 24).

24. The boy *is* *push | ed* *by* *the girl.*

There were 48 critical items. 12 of these critical items were created with animate subjects and animate objects (see example 25a below), 12 with inanimate subjects and inanimate objects (see example 25b), 12 with animate subjects and inanimate objects (see example 25c) and 12 with inanimate subjects and inanimate objects (see example 25d). To test generalisation of cue sensitivity, half the sentences were the present perfect tense (see 25b and c for examples), and half were the present simple (as taught in the interventions) (see 25a and d). Across the 48 items, 24 contained regular verbs (e.g. 25a) and 24 contained irregular verbs (e.g. 25d).

25a. The boy is pushed by the girl.

25b. The wheel has been chased by the cat.

25c. The boy has been hidden by the rock.

25d. The car is hit by the bike.

For the 48 critical items six versions were created (full passive, reduced passive, active progressive and the same with sentence types nouns reversed e.g. 26 below) counterbalanced across six lists (see appendix 9 for list logic). This was done to limit the potential impact of any test effect occurring, due to the participants having completed the same outcome measures three times (Cohen et al., 2017; Marsden & Torgerson, 2012).

26a. The boy ***is pushed*** *by the girl for a while.*

26b. The boy ***is pushed*** *for a while.*

26c. The boy ***is pushing*** *the girl for a while.*

26d. The girl ***is pushed*** *by the boy for a while.*

26e. The girl ***is pushed*** *for a while.*

26f. The girl ***has been pushing*** *the boy for a while.*

These sentences were recorded and were presented alongside two cartoon images constructed to show either a passive or an active interpretation of the stimulus (see 27 below for an example). The recording was "at a slow-to-moderate pace with neutral intonation" (following Hopp & Lemmerth, 2018, p. 182, and Ito et al., 2018) in order to "create optimal conditions for predictive eye movements" (Ito et al., 2018, p. 253). The audio recordings were made by the researcher using natural intonation and a neutral British English accent. This accent was one the learners would be familiar with as a result of the pre-sessional classes they were enrolled in. This was important to ensure that the nature of the audio recordings themselves did not affect the learners' ability to comprehend the input (for full stimuli list see appendix 10).

At the end of each item, a wrap-up phrase was included to attempt to mitigate for prosodic wrap-up cues on the reduced passives. By adding a phrase after a reduced passive, e.g. *The boy was hit* ***over there*** the learners should not have been aware of the absence of *by* due to the speaker's intonation on the verbal phrase. *Over there* and *for a while* were chosen as these phrases do not require a visual correlate to make sense unlike a phrase such as *on the hill.* Two phrases were used as one phrase did not work semantically with all of the scenarios. For example, in *The man has been fixed by the robot,* ***over there*** makes more sense than ***for a while***.

27. *"The child is carried by the dog over there."*

a) 

b) 

Visual stimuli – critical items in the visual world eye-tracking

The images used for the critical items were partly taken from Kasprowicz and Marsden (2017) accessed from the IRIS database and partly created for the experiment by the researcher. Where possible previously tested images were used as these were more likely to represent the situation intended. However, because the current study required unusual scenarios depicting various animacy combinations (i.e. inanimate-inanimate nouns) new images needed to be created for a number of items. The images were all created using freely available clipart or by doctoring existing free images. Every effort was made to ensure that all images used were freely available and copyright was not breached.

As recommended by Godfroid (2019) images appearing together on screen were similar in terms of visual salience. Efforts were made to ensure this was the case as the two images were reversible semantically, so where possible the images used the same visual objects in different agent-patient roles. For example, in the below images (see figure 25) the differences are small, and each image is of similar visual salience. The images were the same for the six items they referred to, only the audio varied (following Dijkgraaf et al., 2017, adapting Altmann and Kamide, 1999). So, the following two images would appear with the following audio stimuli in different trials:

The man is helped by the woman (b)

The man is helped (b)

The man is helping the woman (a)

The woman is helped by the man (a)

The woman is helped (a)

The woman is helping the man (b)


a)


b)

Figure 25. Example eye-tracking images and aural stimuli (matching image in parentheses)

All images were cartoons (no photos were used) to remove the influence of real-world likelihood.

Image size was controlled so that each image was 4cm x 4cm. Objects appeared on a white background. Colour images appeared together, and black and white images appeared together. In other words, if a black and white image was used, both images in that presentation were black and white. This was important because visual salience, size, and location of images can affect the chance of a fixation (Godfroid & Hui, 2020; Henderson, 2017; Henderson & Ferreira, 2004)

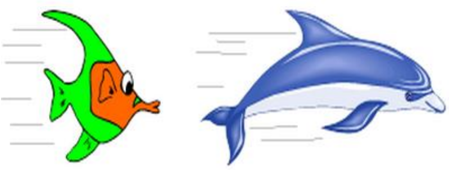
Image position was also controlled for so that the correct image appeared on the left of the screen in half the tests and on the right in the other half. This was to control for the fact that people who read from left to right, and top to bottom (e.g. English and modern Chinese speakers) tend to look to the left of the screen first, even in non-reading tasks (Godfroid, 2019).

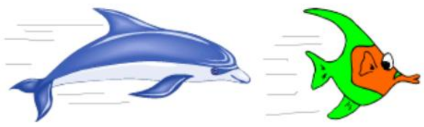
Non-critical items

The critical items were pseudo-randomised and were presented alongside 44 distractors (relative clauses with reversible agent-patient roles) and 24 fillers. The distractors and fillers were the same in each experimental list. Fillers were included to prevent the task itself generating awareness of the target structure by disguising the experimental manipulation (Godfroid, 2019), an issue with some previous research into instruction (as pointed out by Andringa & Curcic, 2015), though it is acknowledged that once the instruction had been experienced, it becomes increasingly likely that participants would know the purpose of the tests.

The distractors focused on the subject-object relationship but not on the passive morphology (adapted from Niu, X., in progress). Half of these were animate-animate and half inanimate-inanimate subject-objects (see figure 26 below for examples). All sentences were active sentences, 14 were present simple, 14 present progressive, and 14 present perfect. Half were subject relative clauses and half object relative clauses. These distractors were chosen because they focused on the subject-object roles and were complex enough to be comparable to the critical sentences.

The dolphin that has been chasing the fish is blue. (correct answer b)





a) b)

The girl that the boy kisses is over there. (correct answer b)

a) b)

Figure 26. Example distractors used in eye-tracking

The 24 fillers did not focus on the subject-object relationship in the stimuli, but focused on the lexical items used, they were also not semantically reversible (see figure 27 below). 12 of these had animate subjects and 12 inanimate, and used various tenses (eight present simple, eight present progressive, eight past simple). Animacy was controlled to ensure that the filler sentences were not noticeably different to the critical items and distractors. Tense also varied for this reason (for full list of fillers and distractors see appendix 11).

The boy does the homework. (images focus on the subject) (correct answer a)

a)  b) 

The woman is driving the car. (focus on the object) (correct answer a)

a)  b) 

Figure 27. Example fillers used in eye-tracking

In total, there were 68 non-critical items and 48 critical items. Keating and Jergerski (2015) recommend 75% non-critical items for eye-tracking studies, the current eye-tracking test had

60% non-critical items. The lower percentage in the current study was decided upon due to the high number of critical items (48) and the length of test. The test was 25-30 minutes long, so to achieve a balance between masking the critical items and mitigating for fatigue, the number of non-critical items was less than the ideal but greater than the minimum. In Godfroid's (2019) recent review of eye-tracking study design, of the 15 studies that reported using non-critical items, the mean percentage of fillers and distractors was 58% (SD = 12) with the percentage of non-critical items ranging from 33%-85% of the total items. The percentage in the current study is therefore in line with recent eye-tracking research.

Comprehension questions

In order to ensure that the learners were attentive during the task, and that they had understood the tasks' instructions, comprehension questions were included. 35 of the non-critical items were followed by a comprehension questions (see appendix 12). 17 of the comprehension questions focused on the audio and 18 on the images. The questions required yes / no responses by pressing z for "yes" and m for "no" (see example 28 below). Half of the comprehension questions require a response of "Yes" and half "No". The critical items were not followed by comprehension questions to further disguise the target feature and goal of the study prior to the intervention sessions.

28. Example comprehension questions - comprehension questions focusing on the audio input

"The boy does the homework."

b)  b) 

Comprehension questions: *Was there some homework?* (correct answer "YES")

The non-critical items were the same in all six lists as were the comprehension questions. The critical items were counterbalanced across the lists and randomised. These were then inserted around the non-critical items to create six tests. At each test time, pre-test, immediate post-test and delayed post-test each participant saw a different version of the test. This ensured that no participant saw the same test twice, and all versions of the test were used during the experiment.

e.g.

	Pre-test	Immediate post-test	Delayed post-test
Participant 1	1	2	3
Participant 2	2	3	4
Participant 3	4	5	6
Participant 4	5	6	1

3.5.4.1 b) Eye-tracking test procedure

The eye-tracking test was designed and compiled with Experiment Builder software (SR-Research, 2011), and the participants' eye movements were recorded with a desktop-mounted EyeLink 1000 eye-tracker. The signal from the eye-tracker was sampled every millisecond. A chin rest was positioned 80cm from the screen to minimize the participants' head movements. The computer screen was 48x27cm and 1024x768 pixels. An ASIO-compatible sound card was used on the display computer to ensure that the audio timing would be accurate. The setting of the test was a dark, quiet room within the Education department at the University of York.

After calibration, the participants were shown instructions (see appendix 13). The learners were instructed to listen to the sentence and answer a question about it if one appeared. They were not asked to look at the image that matched the sentence. Five practice trials, that did not include the target features, were provided to familiarise the learners with the test procedure. An opportunity for questions was given between the practice trials ending and the main experiment beginning. The structure and timings of the test were as follows:

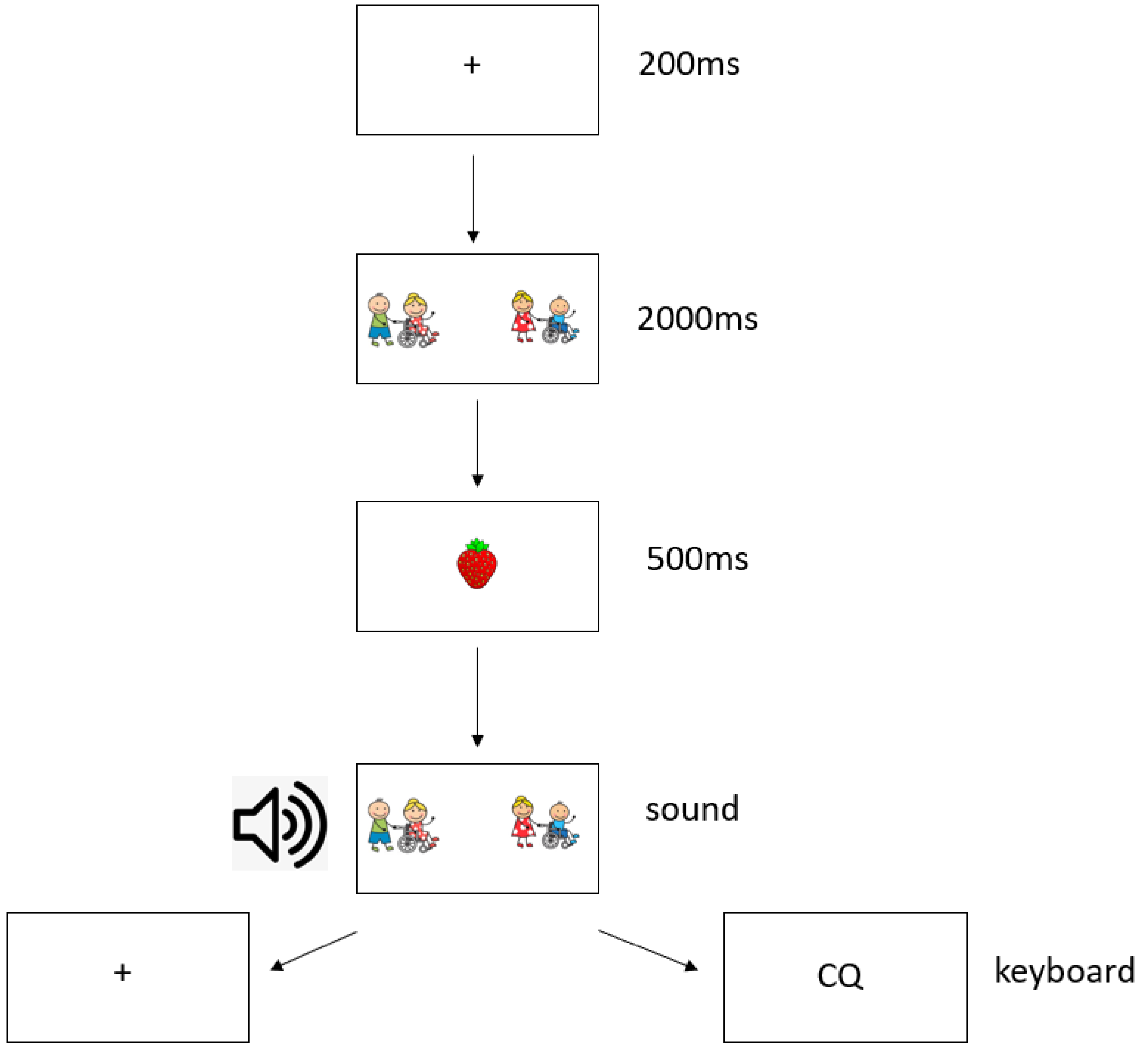


Figure 28. Example eye-tracking trail.

Following other studies which investigated anticipatory processing, a fixation cross was included at the start of each trial (Godfroid, 2019). A preview of the images was provided to ensure that both possible scenarios has been equally considered. In order to produce a mental representation of the images, participants should see images before hearing an audio stimulus. This is because of the linking hypothesis which states that there is overlap between activated mental representations and the following audio input that produces an eye-movement (Altmann & Kamide, 2007). Furthermore, research has shown that, without a preview, anticipatory effects are not seen (Ferreira et al., 2013). Since in the current study it was important that both images were considered by the learners to be equally possible prior to the audio, the preview allowed for anticipation of either scenario to occur.

Having collected the duration of image previews from 36 studies using eye-tracking, we found that 2000ms was the mean and modal duration of the previews used (see appendix 14 for full data from a review of eye-tracking design). Therefore, 2000ms was used. The 'strawberry' slide (in lieu of a fixation cross) following the preview was used to ensure that the learners' eye-movements upon hearing the audio originated from the centre of the screen thus making it easier to determine which image the participants preferred upon hearing the first noun (see figure 28 above).

After the fixation strawberry disappeared the images reappeared, and the audio began. The images remained on the screen for the length of the audio plus 300ms. For some non-critical trials, the audio was followed by a comprehension question. For the critical trials, the audio was followed by a fixation cross and a new trial.

3.5.4.1 c) Data preparation and cleaning

Before conducting the analyses, it is standard practice in eye-tracking studies to first remove trials with incorrect responses to comprehension questions (though these were only on non-critical trials in any case in the current study), and to remove data for looks neither to target nor distractor (i.e. looks away from the screen) (Keating & Jegerski, 2015). 17.31% data points were removed due to not containing fixations. No participants were removed due to low scores on the comprehension questions - all participants scored over 75% on the comprehension questions, mean 89% (SD = 4). Although this does not imply comprehension of the critical items specifically, it does suggest that the learners were paying attention throughout the test. Five participants could not complete the eye-tracking due to technical issues, so these were removed from the final data set for all test phases. Extra participants were recruited to make up for the resulting reduction in sample size.

The data were analysed as the proportion of looks to target compared to the proportion of looks to distractor. This was done using areas of interest (AOI) applied to the data in the Data Viewer program (SR-Research, 2011). The AOIs were slightly larger than the images to account for any drift issues occurring during the eye-tracking test (5x5cm). Occasionally this occurred for a number of trials and the eye-tracker needed to be recalibrated. The AOIs were the same size and location on each trial and did not overlap.

The data were initially extracted in 20ms time windows which were then collapsed into 100ms time windows and log transformed for further analysis. The log transformed data were analysed using a mixed-effects model for each separate time window (following Cunnings, Fotiadou, & Tsimpli, 2017; Dijkgraaf et al., 2017; Flecken et al., 2015; Ito et al., 2018; Kim, Montrul, & Yoon, 2015; Kohlstedt & Mani, 2018; Schumacher, Roberts, & Järvikivi, 2017). More detail about the model is given in the Results chapter.

3.5.4.2 Written untimed grammaticality judgement test (JT)

An untimed written JT was used to assess learners' explicit knowledge of passive and active sentences as this is an appropriate test for measuring explicit knowledge (see section 3.1.1

for more detail). The JT tested grammaticality judgements on passive and active sentences. The same 48 critical items seen in the eye-tracking task were presented in the JT (though with grammatical anomalies inserted, as described below). If a participant saw list one in the eye-tracking they saw a JT also containing items from list one. This was to elicit data during processing of the same items under time constraint (online) and whilst accessing explicit knowledge (offline). 24 fillers were included and were the same as those used in the eye-tracking test. Distractors from the eye-tracking were not included since the test would have been very long and could have resulted in participant fatigue, particularly because the JT was the last test in the battery. Gass and Mackey (2012) suggest that giving participants 50-60 sentences is the maximum to avoid fatigue, the current test included 72 sentences so was slightly longer than ideal in order to include all of the critical and filler items from the eye-tracking.

Participants were asked to judge the sentences as grammatical or ungrammatical and to correct those judged as ungrammatical. To find out whether type of error would affect judgements, three error types were included following Spada et al. (2015) and Li et al. (2016). These were missing auxiliary verb (k = 16) (29a), base form of the main verb (k = 16) (29b) and gerund form of the main verb (k = 16) (29c). The matching image was included to ensure the meaning of the sentence was clear. The images were also used in the eye-tracking test. Half of the 48 critical items appeared in a grammatical form and half ungrammatical (with one of the three errors described above). Item grammaticality and error type were counterbalanced across the three lists so that a grammatical item in list one would be ungrammatical in another list (see appendix 15 for lists).

*29a. The child * being helped by the teacher.*

29b. The dog is being walk by the boy.*

29c. The person is being following by the police.*

Participants were instructed to rate each sentence based on its grammaticality. A Likert scale of one-five was used, one being “ungrammatical” and five being “grammatical”, participants could also choose “*don’t know*”. If the sentences were judged “ungrammatical” participants were asked to circle the problem word or words, and then correct them by changing or adding one word (see figure 29 below) (for a discussion of the advantages of using multiple judgement tasks see Schütze, 1996,). Participants were asked to identify and correct the error because this would give insight into their awareness and understanding of the grammatical structure and encourage them to rely on explicit knowledge. Asking participants to identify, correct, or describe errors results in reliance on explicit knowledge (as found by Bialystok, 1979, 1982).

The woman pushed by the door. 1 2 3 4 5 don’t know Ungrammatical Grammatical	

Figure 29. Item from JT (ungrammatical passive item).

The JT was presented in paper form, to allow the participants to circle and correct the error. Instructions and examples were provided in order to prevent participants becoming confused between grammaticality and plausibility. Some of the sentences were not very plausible in the real world but were constructed grammatically. Confusion between grammaticality and plausibility has been observed to be an issue with JTs (Mackey & Gass, 2015). The researcher also verbally checked the participants understood the task before they began and that they understood the concept of "grammaticality".

3.5.4.2 a) Scoring and reliability

Judgements were scored dichotomously, 1 = correct and 0 = incorrect. For grammatical items, "correct" included Likert rating scores 5 and 4, for ungrammatical items "correct" referred to scores of 1 and 2. Over all tests and conditions only 3% of learners' responses were scored 2 and 4, and only 1% were scored 3 or "don't know". Error identification was also scored dichotomously. If the correct word was circled or underlined (or indicated in any other clear way) then the response scored 1. Incorrect words circled, or multiple words circled, received 0. Corrections were scored in the same way (1 = correct correction, 0 = incorrect correction).

Revelle's (2018) omega was calculated as a measure of test reliability as it is thought to be appropriate for binomial, unit-weighted scales, which do not meet the assumption of unidimensionality (Kasprowicz, Marsden, & Sephton2019; McNeish, 2018; O'Reilly & Marsden, 2020). The more commonly used reliability index, Cronbach's alpha, was not appropriate for the data since it does not account for minor dimensions and non-unidimensionality (i.e. tau equivalence) (McNeish, 2018). Revelle's omega also takes into account differences in the degree to which individual items measure the construct in question (i.e. grammaticality). Furthermore, although Cronbach's alpha is often used to asses instruments such as JTs or production tests, its estimates can be overly conservative due to empirical data often not meeting the assumptions required (ibid). Since each item in the JT provided data for various variables (i.e. voice, animacy, tense) the data could be assumed to be non-unidimensional data (i.e. does not meet tau equivalence). McNeish (2018) first proposed the use of omega and other reliability indices in L2 research. The current study adds to recent L2 research that has followed McNeish (2018) in order to investigate reliability using omega (e.g. Kasprowicz et al., 2019; O'Reilly & Marsden, 2020). Since the use of various reliability indices in L2 research is so recent, further studies are needed to confirm which are most appropriate for different instruments and data types. For the current JT, Revelle's omega results suggested that the reliability of each version of the JT was approaching a reasonable level (ungrammatical items mean $\omega = .71$, grammatical items mean $\omega = .72$) (based on the 25th, 50th (median), and 75th percentiles of instrument reliability from a meta-analysis of reliability coefficients i.e. .74, .82, .89 (Plonsky & Derrick, 2016).

To check interrater reliability, a second marker was recruited and trained to score a subset of the data (20 JTs) taken pseudorandomly from the three groups and test times. Interrater agreement score was calculated using Cohen's (1960, p. 201) equation:

$$\frac{\text{Number of actual agreements}}{\text{Number of possible agreements}} \quad \text{x } 100$$

The agreement score was 88%. Mackey and Gass (2016) advise that 75% agreement is good, and 90% agreement is ideal. So, the scores in the current study were reliable.

10 NSs were also tested. They scored a mean of 91% (SD = 3.59) (range = 85%-100%). The high NSs' accuracy score suggests that the items in the JT were acceptable to the NSs and elicited the intended target response. In other words, the NSs' scores suggested strong test validity.

3.5.4.3 Oral production test

The oral production test assessed participants' accuracy in producing full passives. To elicit the full passive a picture cue task was used (based on Seyednejad & Gholami, 2017). The picture description task provided participants with two images and a sentence stem. The test was made up of 36 items - 18 present perfect passive sentences and 18 present simple passive sentences. The 36 passive items had three versions, passive voice, active voice and then the alternative tense (see example 30). The present perfect items were included as they were untrained (i.e. not in the interventions), so the learners' ability to generalise cue use from trained constructions to untrained constructions was examined.

30. Item from picture description task in three versions

30a. The picture has been drawn by the boy.

(present perfect passive - untrained)

30b. The boy has been drawing the picture.

(present perfect active - untrained

30c. The picture is drawn by the boy.

(present simple passive - trained)

In line with the other tests animacy was manipulated in the oral production test. Half of the items had reversible agent patient roles (*k*=18), as both nouns were animate (31a). The other half were non-reversible animate-inanimate or inanimate-animate items (*k*=18) (as in 31b and c). Half of the reversible and non-reversible items were present simple (trained) and half present perfect (untrained).

31a..Animate-animate reversible item: The dog is asked by the girl

31b. Animate-inanimate non-reversible items: The man is offering the biscuit

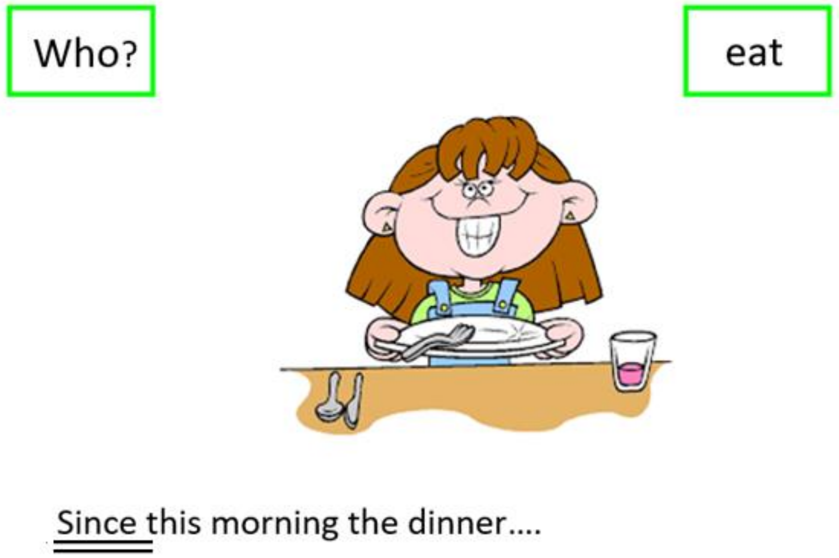
31c. Inanimate-animate non-reversible items: The tower is built by the girl

These items were used to create three tests with different versions of each item so that the pre, post and delayed post-tests were different (see appendix 16).

To elicit passives, participants were asked to finish the sentence using the stem (e.g., *since this morning the dinner*) by answering the question "*who?*". To do this correctly they had to produce a full passive. In the example below this would be "*The dinner has been eaten by the girl*". To elicit actives, participants were asked to finish the sentence using the stem by answering the question "*what*?" (see figure 30 below). The sentence stem (e.g., *since this morning the dinner*) contained the patient in the passive sentences and the agent in the active sentences. The tense was elicited using "*since this morning*" for the perfect tense (untrained) and "*now*" for the present tenses (trained).

The picture cue task was carried out using Experiment Builder (SR Research). Responses were recorded using a microphone. The image remained on the screen until the participants responded orally, 3000ms after speaking the trial ended and the next one began.

This was to prevent the learners from revising their answers, in order to promote spontaneous production.



Correct response: *(Since this morning the dinner) has been eaten by the girl*

Figure 30. Picture description test example.

Prior to the critical trials, the participants were given instructions on screen followed by four practice trials. The practice trials were designed so as not to elicit the target form – the passive. In these practice trials, "*what?*" was used to avoid eliciting the passive voice, and "*yesterday*" was used to elicit the past simple, rather than the present simple or present perfect as in the critical trials.

3.5.4.3 a) Scoring and reliability

Scoring assessed the use of the correct verb inflection and the use of *by*. These elements were focused on because they were the cues focused on in the EI (for both conditions) and the cue focused training. The correct verb form had to include both the correct auxiliary and correct verb form. For example:

[time adverbial provided: Since this morning] The picture has been drawn by the boy scored 2 (1 point for the correct form of the verb and 1 point for the correct use of *by*)

The picture has been drawing by the boy scored 1 point for correct use of *by*

For scoring the production of active constructions (scored separately to passives) a progressive sentence scored 1 point for the absence of *by*, e.g.

The girl is paying the man scored 2 points (1 for the correct verb and 1 for not using *by*)

The girl is being paid by the man was also acceptable for the active sentences.

As in the JT, a second marker was recruited and trained to score a subset of the data (20 oral production tests) taken pseudorandomly from the three groups and test times. Interrater agreement score was 91%. Compared to the benchmark of 90% the scores in the current study were ideal (Mackey & Gass, 2016). 10 NSs also took the test and scored a mean of 92% (SD = 6.00) (range = 82% - 99%). This was not 100% because two of the NSs ignored the prompt "*who*?" and the sentence stem on some trials and converted the sentence into an active when it should have been a passive. For example, instead of finishing the sentence stem "*The award*……" with the target "….*is presented by the man*" the participants responded with "*The man presented the award*". This was unexpected since the participants were given examples and the sentence stem and cues were highlighted. None of the learners responded in this way. It is interesting that these NSs felt that the active interpretation was more acceptable than the passive in some cases.

As for the JT, to assess the reliability of the oral production test, Revelle's omega was calculated as it is appropriate for unit-weighted scales, which do not meet the assumption of unidimensionality (McNeish, 2018). Revelle's omega suggested that the reliability of each version of the oral production test was approaching a high level (version 1 $\omega = .82$, version 2 $\omega = .77$, version 3 $\omega = .90$) (based on Plonsky & Derrick, 2016).

3.5.4.4 Written production test

The written production test was the same format as the picture cue test but instead of completing the sentence orally, the participants typed the end of the sentence (see figure 31). As in the picture description test the written production test had 36 items (see appendix 17 for items). The items in the written test were the same as in the picture cue oral production test. Different versions of the oral production test and the written test were used for each participant in each test session e.g. if a participant saw oral production test version one, they saw writing test version three in the same test session. This meant that two identical items would never be seen in the same test session in both production tests. Using the same items across test sessions resulted in all the lexical items being used across both the oral production measure and the written production measure.

The written production test was administered using Open Sesame (Mathôt et al., 2012). The first screen (figure 31 below) gave the sentence stem with a matching picture and prompts to elicit the voice and tense. In the below example "*Who*?" elicited a passive sentence, "*throw*" provided the verb to be used in the sentence, and "*Since this morning*" indicated the tense – present perfect. This screen was visible for 3000ms. The screen that followed reminded the participants of the cues and the image and included a text box to complete the sentence stem (figure 32). This screen remained visible until the participants finished typing and pressed the enter key. They had the opportunity to delete and revise their typed answer before pressing enter to continue.

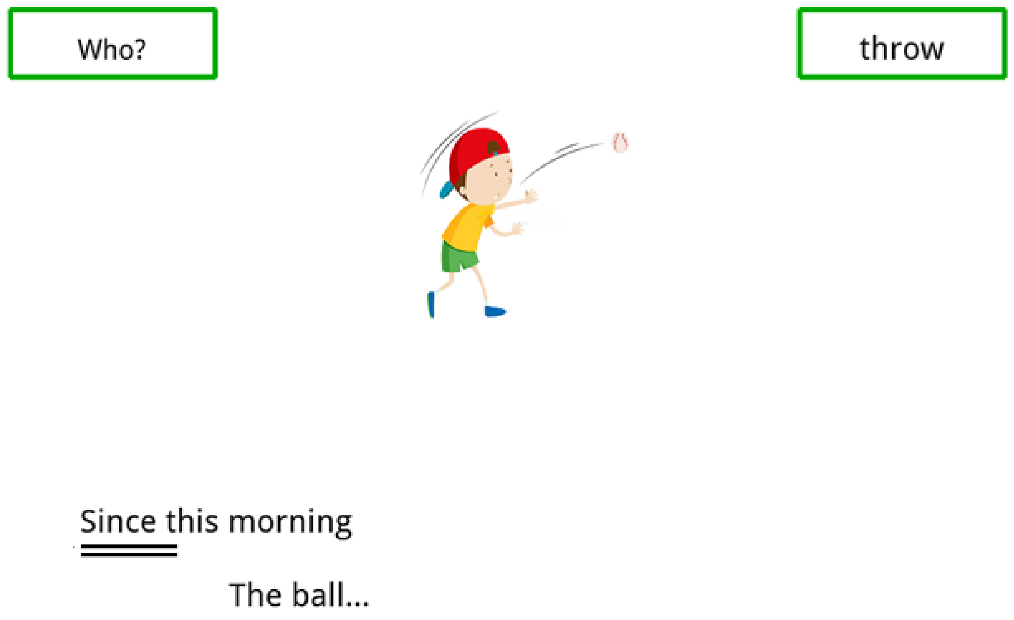


Figure 31. Example written production test trial - preview of items (3000ms)

Who?

throw

Since this morning the ball...

Correct response: …………… was thrown by the boy

Figure 32. Example written production test trial – response screen

Prior to the main trials the participants received instructions and two practice trials followed by an opportunity to ask the researcher questions before beginning the main test. The researcher observed the practice trials to ensure the participants were following the instructions correctly and intervened if necessary, with further instructions. The practice trials did not elicit the target forms. "*What?*" was used to avoid eliciting the passive voice and "*yesterday*" was used to elicit the past simple, not the present simple or present perfect (as in figure 33 below). Therefore, the test procedure was practised, but the target language was not.

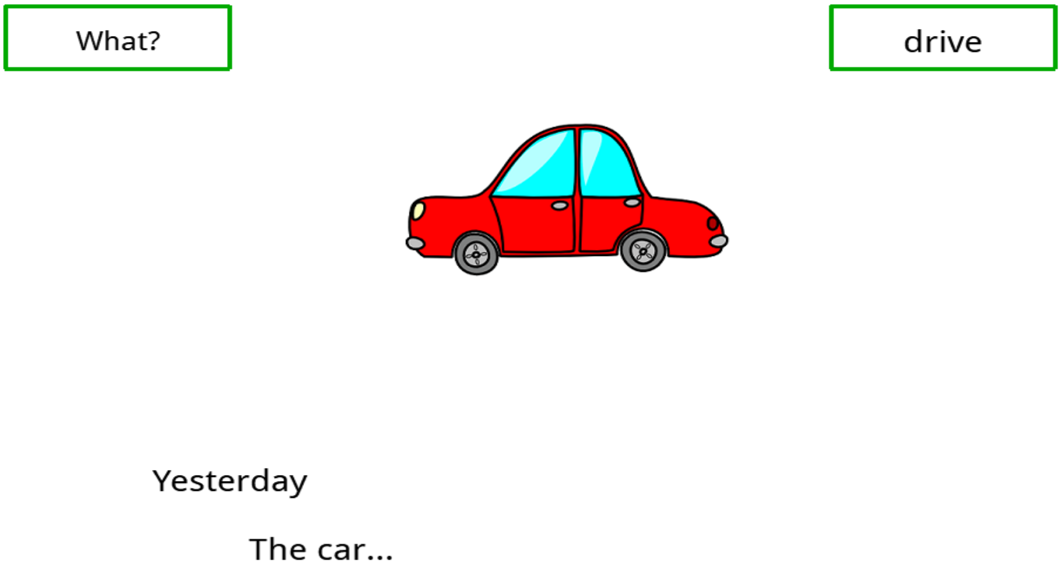


Figure 33. Practice trial written production test.

3.5.4.4 a) Scoring and reliability

As in the oral production test, scoring assessed the use of the correct verb form and the use of *by*. These elements were focused on because they were the cues focused on by the cue focused training. For example:

[time adverbial provided: Since this morning] The picture has been drawn by the boy scored 2 (1 point for the correct form of the verb and 1 point for the correct use of *by*)

The picture has been drawing by the boy scored 1 point for correct use of *by*

For scoring the production of active constructions (scored separately to passives) a present progressive sentence scored 1 point for the absence of *by*, e.g.

The girl is paying the man scored 2 points (1 for the correct verb and 1 for not using *by*)

The girl is being paid by the man was also acceptable for the active sentences.

Again, the second marker scored a subset of the data (20 WTs) taken pseudorandomly from the three groups and test times. The interrater agreement score was high (93%). The NSs scored a mean of 91% (SD = 6.00) (range = 79%-100%). This was not 100% because some of the participants produced reduced passives, not full passives, despite the prompt "*who*?". This occurred in half of the tests (5/10) but not all of the trials. It did not appear to be an issue caused by particular trials.

Since the writing test data were similar to the oral production test, Revelle's omega was again used to assess instrument reliability. Reliability was found to be acceptable (version 1 ω = .81, version 2 ω = .84, version 3 ω = .79). On average the reliability index suggested that the writing test was fairly reliable but there may have been issues with some of the items on some of the tests.

Chapter 4: Results

In the following chapter, the results of the interventions and outcome measures are presented in order to respond to the study's RQs. NS data from the visual world eye-tracking task are analysed and compared to the learners' eye-tracking data. The effect of the interventions on cue sensitivity, as evidenced by the eye-tracking task, is compared with the test-only group. As is performance on the offline measures (oral and written production tests, and a JT). The effects of animacy and trained / untrained constructions on production and comprehension are also analysed. First, intervention performance is presented, followed by the eye-tracking results for the learners and NSs, and finally the offline test results are given.

4.1 Intervention results

Prior to the results of the outcome measures, it is useful to examine performance during the training conditions themselves, to inform us about whether the training was completed as expected by the learners.

4.1.1 Response accuracy during the interventions

4.1.1.1 Overall intervention score – passives and actives combined

For each trial in the intervention the learners scored 1 for a correct response (key press) and 0 for an incorrect response. All the trials required dichotomous responses, either a choice of one of two pictures, or one or two words (see figure 34 below for an example). Total accuracy score is presented as a percentage of the total trials (see table 2).

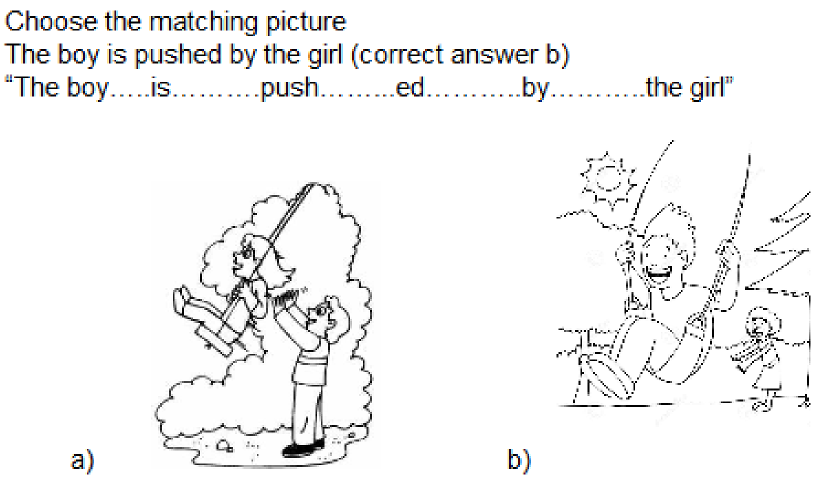


Figure 34. Example intervention task

Both the cue focused group and the noun focused group’s mean percentage score approached ceiling level in both sessions of the intervention. Each session contained *k* = 128 practice items, making the total denominator 256 for the percentage calculation.

Table 2. Mean % intervention session score (SDs)

Group (*n*)	Session 1 (*k* = 128)	Session 2 (*k* = 128)
Cue (25)	89 (6)	90 (8)
Noun (25)	91 (5)	91 (6)

The data violated the assumption of normality (Shapiro-Wilks p = .009) so non-parametric tests were used. There was no difference between training sessions overall (*z* = -627, *p* = .53) or groups at the different sessions (session 1 *U* = 247, *p* = .20; session 2 *U* = 288, *p* = .64).

4.1.1.2. Intervention score - Passive and active items

Table 3 below shows very little difference between passive and active items in both intervention sessions and for both groups.

Table 3. Mean scores % (*SDs*) for intervention sessions by voice and group

Group (*n*)	Session 1		Session 2	
	Passive (*k* = 64)	Active (*k* = 64)	Passive (*k* = 64)	Active (*k* = 64)
Cue (25)	89 (7)	89 (7)	88 (9)	91 (7)
Noun (25)	92 (6)	91 (6)	90 (9)	91 (6)

4.1.2 Response time during intervention tasks (reaction times)

Both groups reduced their reaction time (RT) by session two in all four tasks.

Table 4. Mean RTs in ms for cue focused group (mean and* SDs*)

	Session 1	Session 2	Difference between session 1 and 2
Task 1 (*k*=32)	4,338 (686)	3234 (571)	-1104
Task 2 (*k*=32)	2,370 (546)	1972 (726)	-398
Task 3 (*k*=32)	3619 (558)	3101 (906)	-518
Task 4 (*k*=32)	2741 (206)	2736 (347)	-5

As table 4 above shows there was a reduction in RTs for the first three tasks. The fourth task was more complex (requiring participants to read a sentence word-by-word and choose a correct image as soon as possible) so it may not have been surprising that the average RT did not decrease as compared to previous sessions.

Table 5. Mean RTs in ms for noun focused group (mean and* SDs*)

	Session 1	Session 2	Difference session 1 and 2
Task 1 (*k*=32)	2736 (520)	2045 (470)	-691
Task 2 (*k*=32)	2421 (471)	1814 (519)	-607
Task 3 (*k*=32)	1103 (387)	794 (311)	-309
Task 4 (*k*=32)	1054 (128)	962 (96)	-92

To summarise the intervention results, there was no difference between the two groups in terms of accuracy, with both groups' accuracy being high during both intervention sessions. In terms of RTs, since the tasks were different in each intervention type it is not possible, or meaningful, to compare statistically the RTs of the cue focused group and the noun focused group. Nevertheless, descriptively, the sets of RTs showed a similar trajectory of speeding up between trials in each of the first three tasks (with the cue focused group taking a lot longer on the first task than the noun-focused group), and then both groups completing task four in about the same time as task three.

4.2 Passive and active morphosyntactic cue sensitivity: Online processing

This section presents the data for RQ1 followed by RQ3:

RQ1) To what extent do second language learners show sensitivity during processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

Is cue sensitivity affected by:

a) Training (either cue focused or noun focused) as compared to a test-only control group?
b) First and second noun animacy?
c) Trained and untrained constructions (present simple and present perfect?)

RQ3) To what extent do native English speakers show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

To what extent is native speaker processing of the passive different to learners?

The first research question is addressed by analysing the visual world eye-tracking test. To analyse the eye-tracking data a mixed-effects model was used. Proportion of fixations to target were calculated in 100ms time windows and then log transformed. Then a model was run for each time window (TW) over the time course of the eye-tracking data. Following a maximal approach (Cunnings, 2012) to multi-level modelling and a maximal random-effects structure (Barr et al., 2013), a model considering all possible variables was run. The maximal model had proficiency as a control variable, fixed effects of animacy, test time, group, voice (active and passive) and reduced / full passive, and random intercepts for animacy, voice, reduced and test time, and random slopes for subject and item, as follows:

*Maximal model: TW800_720 ~ prof + animacy * time * group * voice * reduced + (time * animacy * voice * reduced | sub) + (voice * time | item)*

This model did not converge, so random slopes were removed stepwise until a model which converged was found. This was done by first changing interactions to additive effects for the random intercepts and then slopes e.g. (time * group | sub) became (time + group | sub). Following this, random intercepts were removed first, followed by random slopes e.g. (time * group | sub) became (time | sub) (see appendix 18 for all attempted models). The predictor variables were sum coded so that the intercept moved to the middle of the categories. This is the conceptual version of "centering" for categorical predictors (following Winter, 2019).

The resulting model was the basis for all models used to analyse the eye-tracking data. *P* values were calculated using Satterthwaite approximation calculated in lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). Likelihood ratio tests (to give *p* values by comparing the full model with the model excluding the effect of interest) were not appropriate given that a separate model was run for each time window (TW). The control variable was proficiency, the fixed factor predictors were animacy, test time, group, voice and reduced / full passive. The random slopes were subject and item. The following model was built in R software (R Core Team).

*Final model: TW ~ prof + animacy * time * group * voice * reduced +(1 | sub) + (1 | item)*

The trained constructions, those practised in the intervention sessions, were analysed separately from the untrained constructions. The predicted eye-tracking behaviour for the native speakers was as follows.

If the native speakers behaved as expected, based on the Given-new theory of processing and previous research into predictive cue use (see sections 2.1.2.1 and 2.1.2.2), then the eye-movements seen in the charts would be as follows. It would be expected that, upon hearing the first noun, there would be some anticipatory looks to either the image depicting the passive interpretation or the active interpretation. The first noun would generate expectations about the likely interpretation of the upcoming input. These would be reassessed at the point of disambiguation (the verb inflection). In active sentences, since an active interpretation would have been more likely, the verb inflection (-*ing*) would be integrated with ease and processing would be facilitated. For actives, at the verb ending, a rise in looks to the picture matching the active interpretation would be expected. For passives, the unexpected verb inflection (-*ed* / other past participle) would be expected to cause processing difficulty because the original interpretation of the sentence would need to be reassessed. At this point, for passives, looks to the image matching the passive sentence would not be expected to increase. Having integrated the new information (the verb inflection) the following preposition *by* would be expected. So, it would be at this point that processing would be facilitated and looks to the passive image would be expected to increase.

The description above of the native speakers' predicted behaviour is intended to aid interpretation of the following charts and findings by providing a benchmark for what might have been expected if incremental processing, and predictive cue use, were employed by the native speakers and / or learners in the eye-tracking task.

4.2.1. Trained constructions

This section focuses on the trained constructions (e.g. 32 below). The trained constructions were either present simple passive (*k* = 24) or present progressive (active) (*k* = 24).

32a. The dog is carried by the boy

32b. The dog is carrying the boy

The analysis addresses the following variables:

- Training (either cue focused or noun focused) or no training (test-only control group) at each test time (pre-test, immediate post-test, and delayed post-test).
- Passive and active voice.
- Reduced versus full passives
- Animacy

The figure below (35) gives a basic overview of any statistically significant effects ($p < .10$) or interactions found by the mixed-effects model for trained items. Significance was $p < .10$ in this figure in order to show all time windows in which an effect may have occurred. For example, the arrow indicating the effect of animacy captures time windows where $p < .05$ and < .10. The arrows demonstrate approximately the time windows these effects were seen in. Nonsignificant effects (and any effects not reported in this chapter) can be found in appendix 19).

These interactions and effects, for the trained items, are described in detail in the following sections. 0ms in the following figures indicates the offset of the verb (either *-ed* (or other past participle ending), or irregular past participle inflection, or *-ing*). Effects of the verb and subsequent features are described taking into consideration 220ms needed to plan and execute an eye-movement (Wright & Ward, 2008). In other words, effects of the offset of the verb would be expected around 200-300ms rather than at 0ms. Statistical significance for results reported in the text was $p < .05$

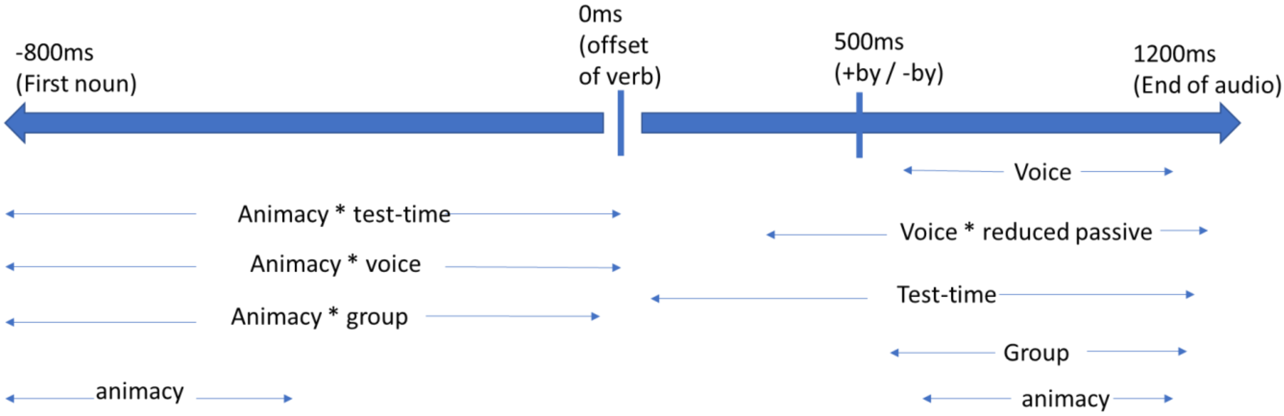


Figure 35. Overview of main statistically significant effects and interactions for trained items

4.2.1.1 The effect of voice and group at each test time (pre, immediate and delayed post-test)

The following figures show the mean proportion of looks to target and looks to distractor, for passives and actives separately at each test time, for the three groups. The mean proportion of looks does not total 1.00 because looks away from the two AOIs are not shown in the figures e.g. looks to other areas of the screen. As mentioned in the method chapter, data points in which there was no data (no fixation) were removed prior to analysis (see 3.5.4.1c). In order to show behavioural differences at each test time (pre, immediate post, and delayed post-test), figures for each test are presented separately. However, the mixed-effects models analysed the combined data from all three test times with test time as a predictor variable. Significant findings from the models are reported alongside interpretation of the figures. Nonsignificant results from the models are in appendix 19.

4.2.1.1 a) The effects of voice and group - Pre-test

For passives, after the verb offset, all three groups looked more to distractor than to the target for the passive items (see figure 36 below). For actives, all three groups looked more to target than distractor (see figure 37 below).

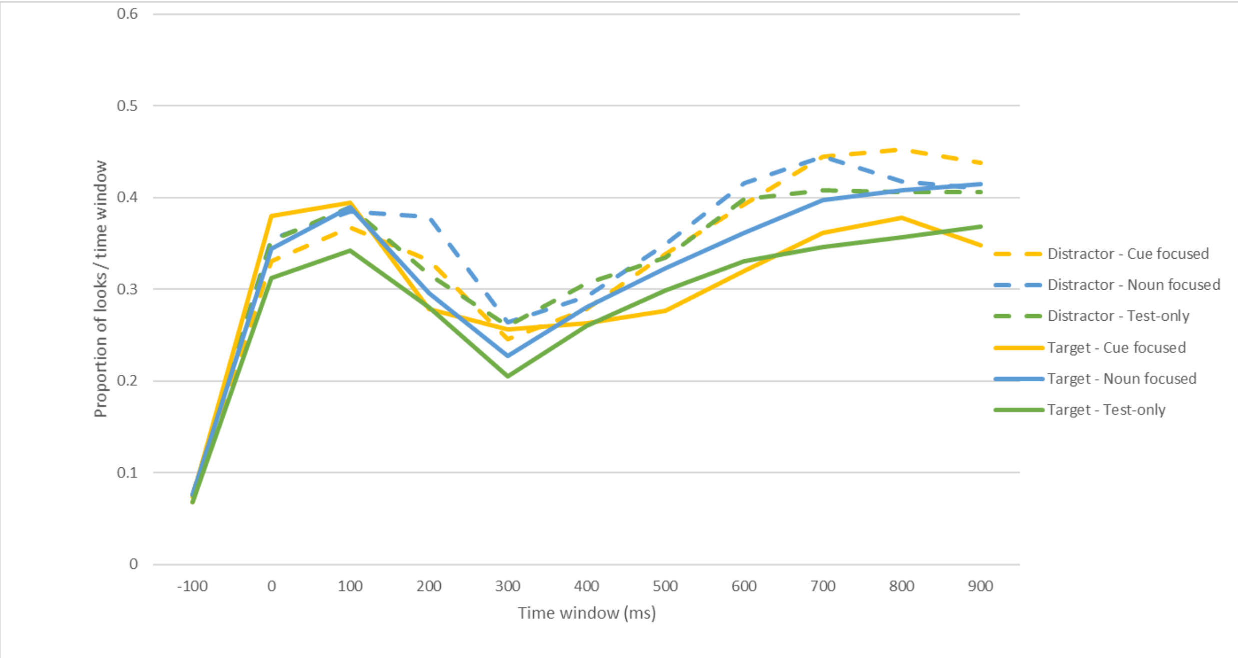


Figure 36. Eye-tracking trained passives and group at pre-test

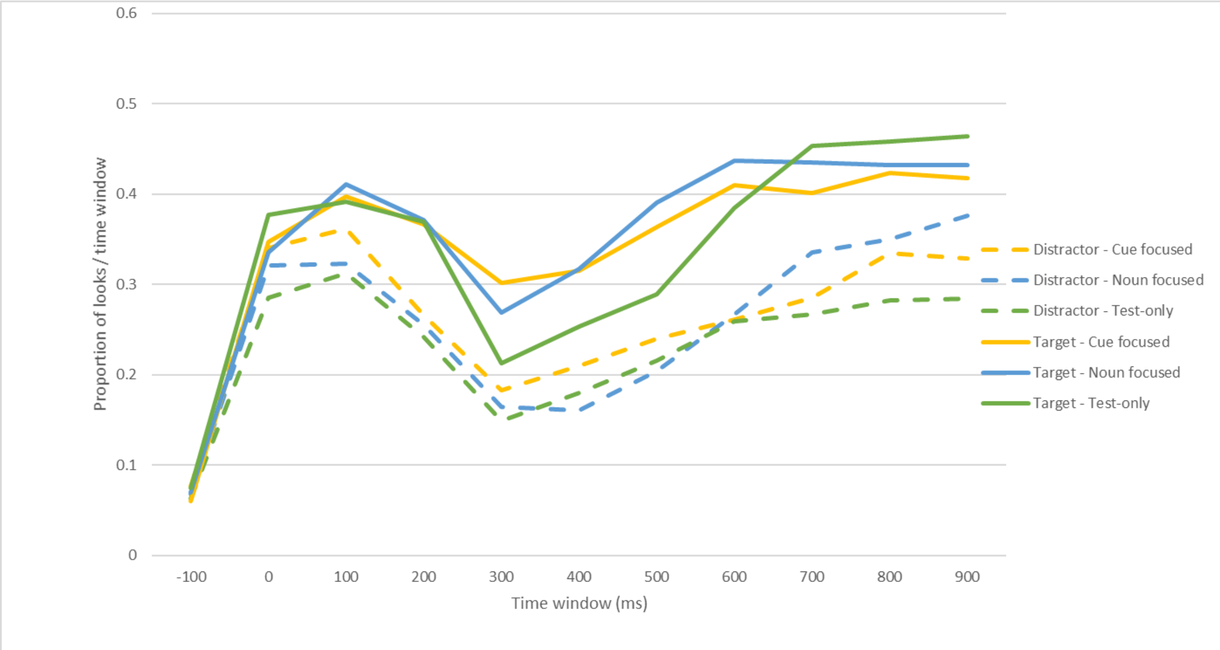


Figure 37. Eye-tracking trained actives and group at pre-test

4.2.1.1 b) The effects of voice and group - Immediate post-test

The cue focused group, for passives, looked more to target than distractor around 300ms (the same point as the noun focused group), but with a lower proportion of looks overall compared to the noun focused group (see figure 38 below). The model found a significant interaction for the cue focused group and passive items at 700ms (estimate = 6.03e-01, S.E. = 2.76e-01, t = 2.19, p = .03). For the cue focused group, a similar pattern of looks to target was shown for active items.

The noun focused group, for passives, looked much more to target than distractor compared to pre-test, and compared to the other two groups (figure 38 below). The model found a significant interaction for the noun focused group and passive items at 700ms (estimate = 5.74e-01, S.E. = 2.71e-01, t = 2.12, p = .03). For actives, the noun focused group's looks to target clearly diverged more from distractor at around 200ms – sooner than at pre-test.

The test-only group, for passives, looked more to distractor throughout at immediate post-test showing no difference in processing behaviour compared to pre-test. For actives, looks to target surpassed distractor but to a lesser proportion than the intervention groups (see figure 39 below).

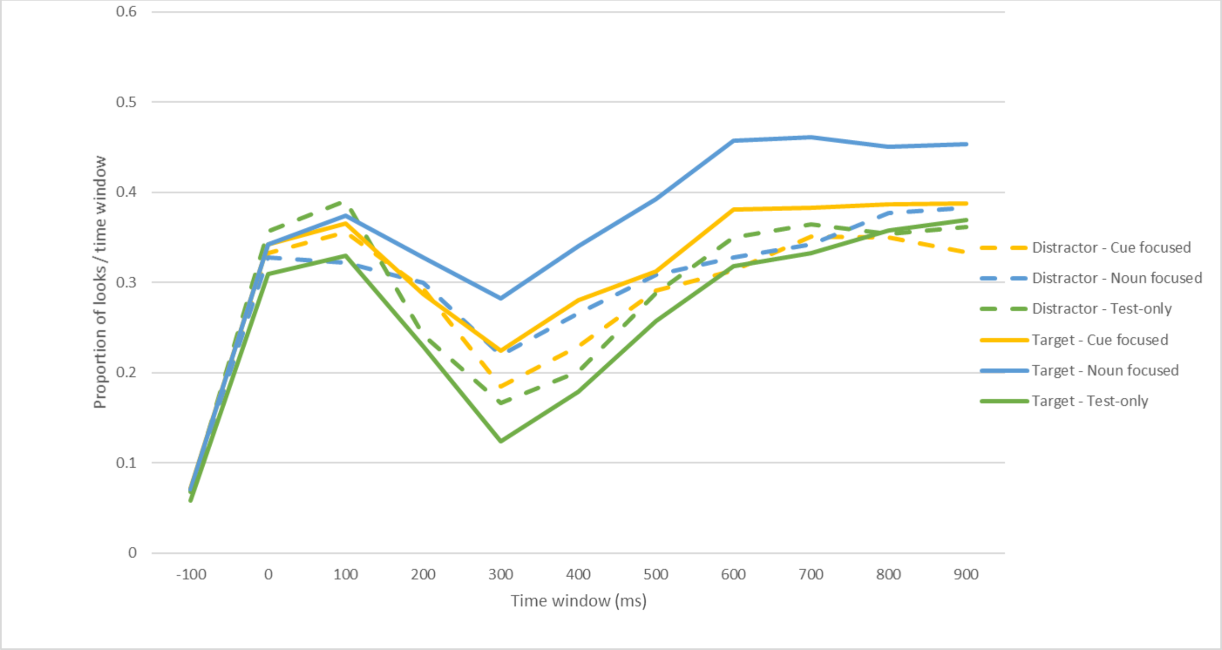


Figure 38. Eye-tracking trained passives and group at immediate post-test

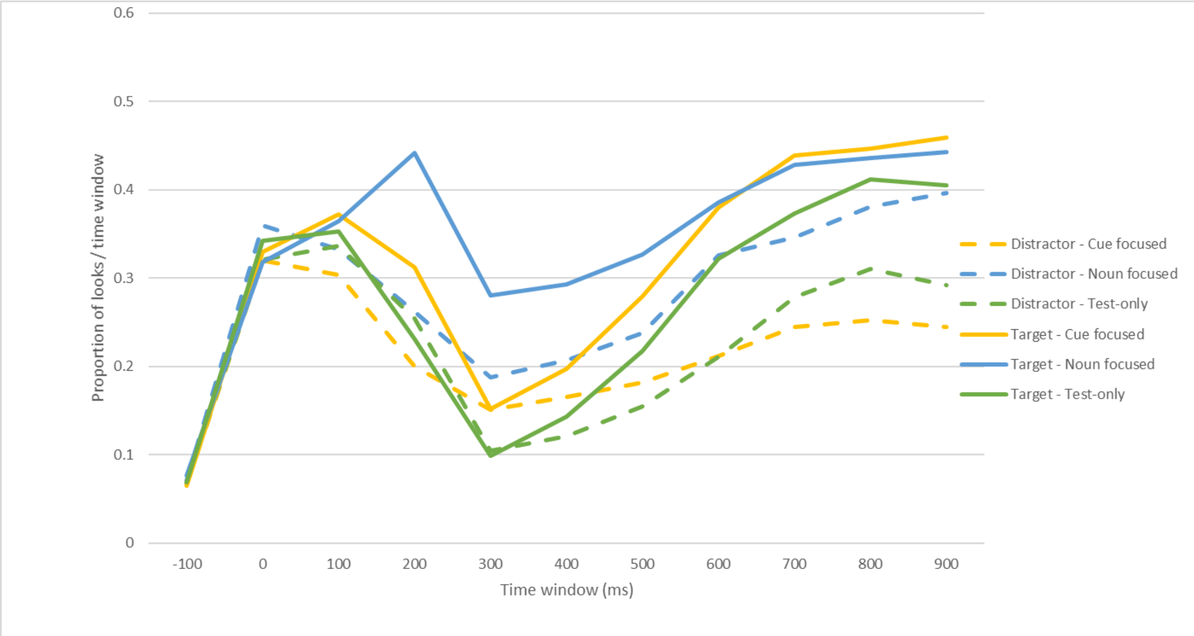


Figure 39. Eye-tracking trained actives and group at immediate post-test

4.2.1.1 c) The effects of voice and group - Delayed post-test

The cue focused group, for passives, looked to target sooner compared to the other two groups. This was also sooner than at the other test-times. However, this visible trend, shown in figure 40 below (p.187), was not supported by the model as no effect or interaction with group was found for passive items and test time in the time windows immediately following the verb (see appendix 19 for model results). For passives, after initially looking more to target, the cue focused group looked more to distractor for the remaining time windows. For actives, the cue focused group looked to target less than distractor at all time windows (see figure 41 below). This was a marked change from immediate post-test in which looks to target were greater than to distractor.

The noun focused group, for passives, looked proportionally more to distractor than to target at 100-300ms (directly after the verb). Their looks to target only surpassed looks to distractor around 700ms (around the second noun). This was much later than at immediate post-test. For actives, the noun focused group looked to target slightly later than at immediate post-test. Looks to target far surpassed the other groups in the time windows following the verb (300ms onwards).

This difference in looks to target between the noun focused group and cue focused group at delayed post-test compared to the immediate post-test was reflected in the model by significant interactions of delayed post-test and passive items found around 100-200ms (Estimate = 6.74e-01, S.E. = 2.97e-01, $t = 2.27$, $p = .02$) and 700-800ms (Estimate = 5.63e-01, S.E. = 2.72e-01, $t = 2.07$, $p = .04$).

The test-only group, for passives, also looked more to target earlier (100ms) at delayed post-test compared to immediate post-test, and looked more to target than distractor after the verb for all time windows, albeit proportionally less than the other groups. For actives, at delayed post-test, the test-only group looked more to target earlier (100ms), compared to immediate post-test, and again after 500ms (the absence of *by*). However, these observed differences between test times were not found to be significant by the model.

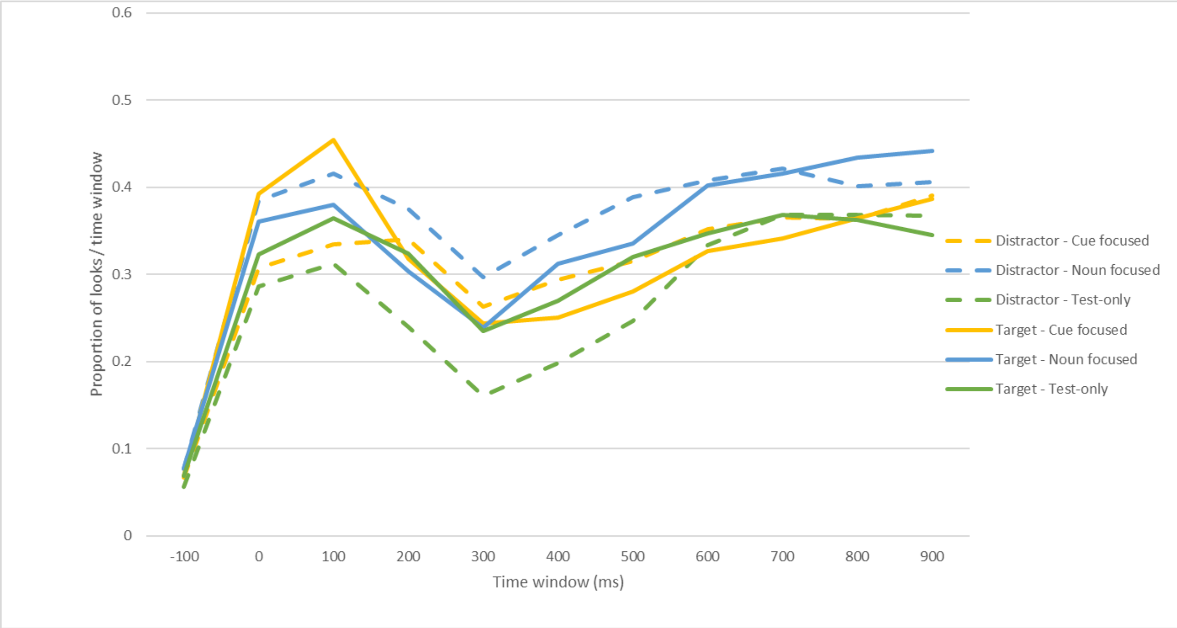


Figure 40. Eye-tracking trained passives and group at delayed post-test

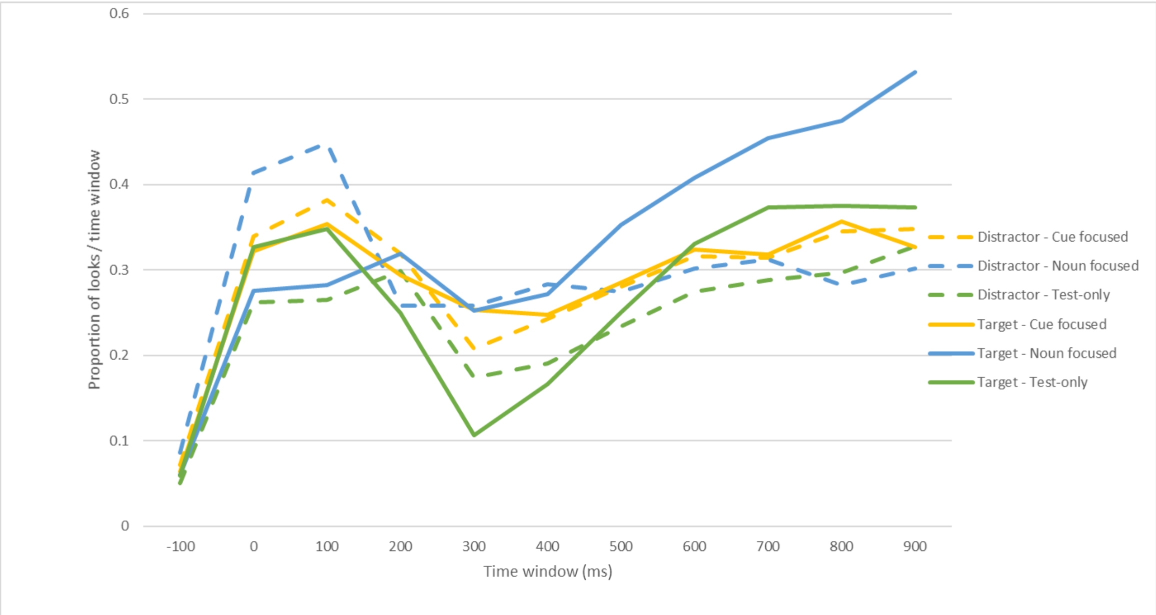


Figure 41. Eye-tracking trained actives and group at delayed post-test

4.2.1.2 The effect of type of trained passive (reduced / full passives)

Reduced passives were included to investigate how learners would deal with sentences without the cue of the preposition *by* e.g.

33. The car is chased

The following section presents figures at each test time showing each groups' looks to target and distractor for reduced and full passives.

4.2.1.2 a) Reduced / full passives - Pre-test

At pre-test, there were no clear differences between reduced and full passives after 500ms. Whether or not passives were reduced or full was not found to be a significant main effect at pre-test at any time window (see figures 42 and 43 below).

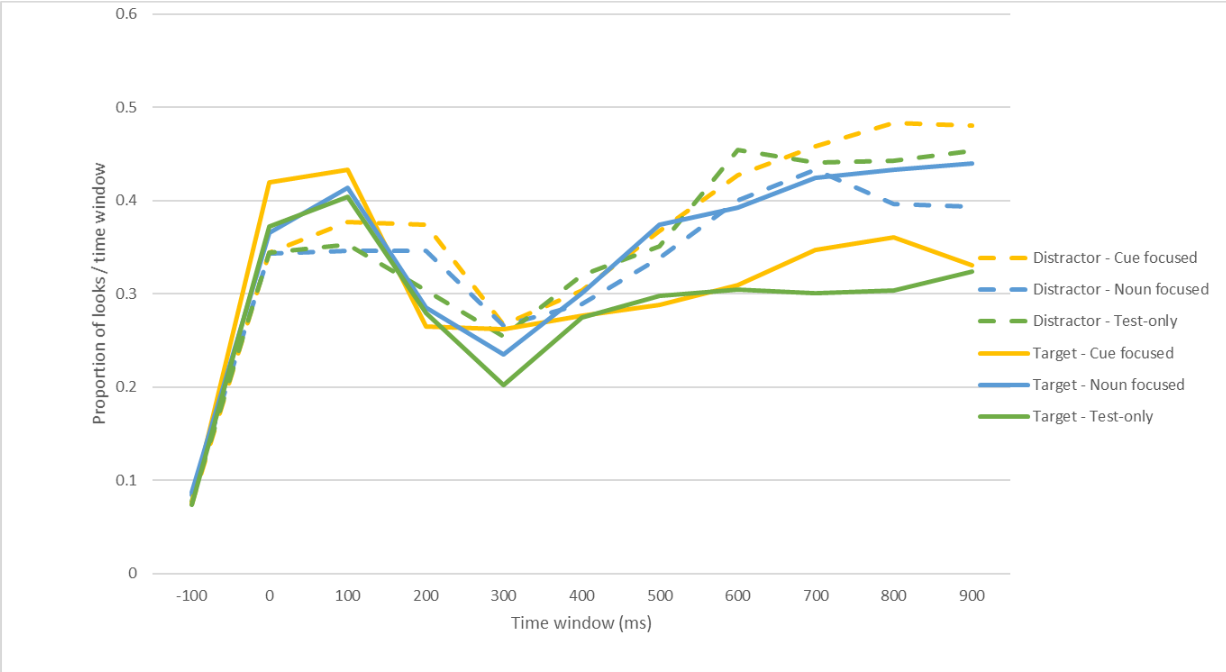


Figure 42. Eye-tracking – full passives and groups at pre-test

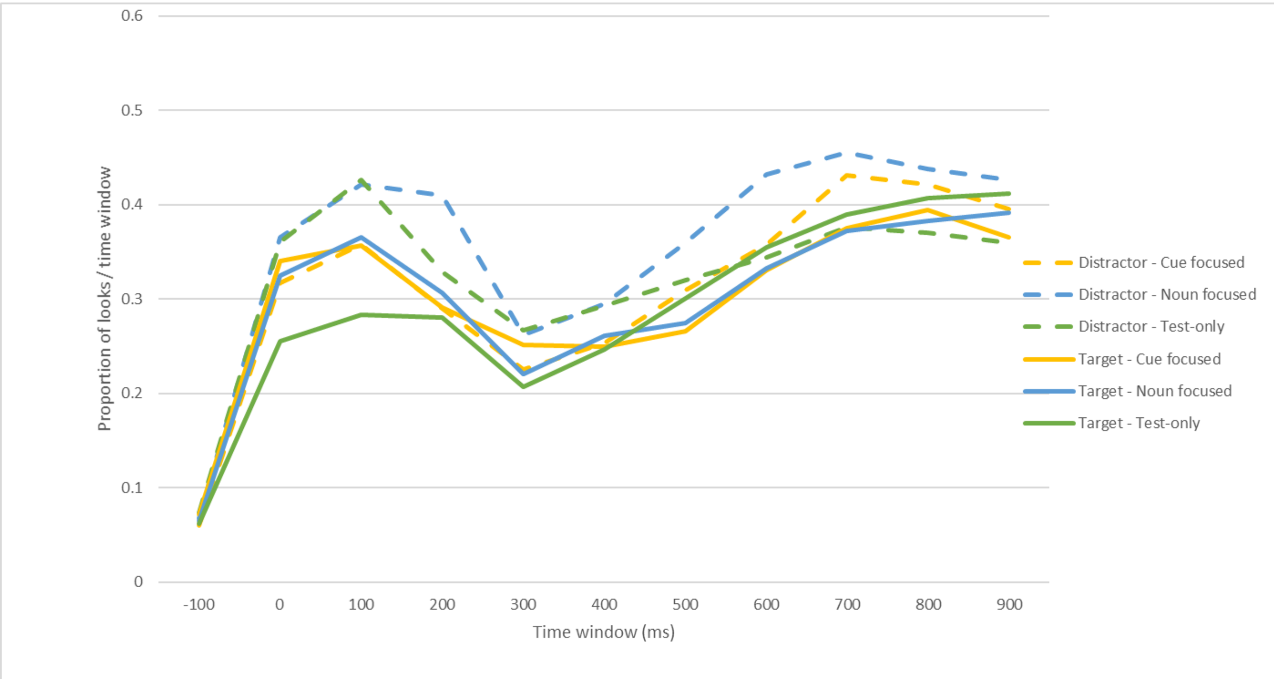


Figure 43. Eye-tracking – reduced passives and groups at pre-test

4.2.1.2 b) Reduced / full passives - Immediate post-test

At immediate post-test, only the noun focused group appeared to respond noticeably differently to reduced passives compared to full passives. For reduced passives, looks to target rose to the absence of *by* and then fell (see figure 45). This difference at 500-600ms between pre-test and immediate post-test and reduced and full passives for the noun focused group was found to be significant by the model (estimate = 5.56e-01, SE = 2.73e-01, t = 2.05, p = .04). The fact that the noun focused group's looks decreased after the absence of *by* perhaps suggests some sensitivity to its absence.

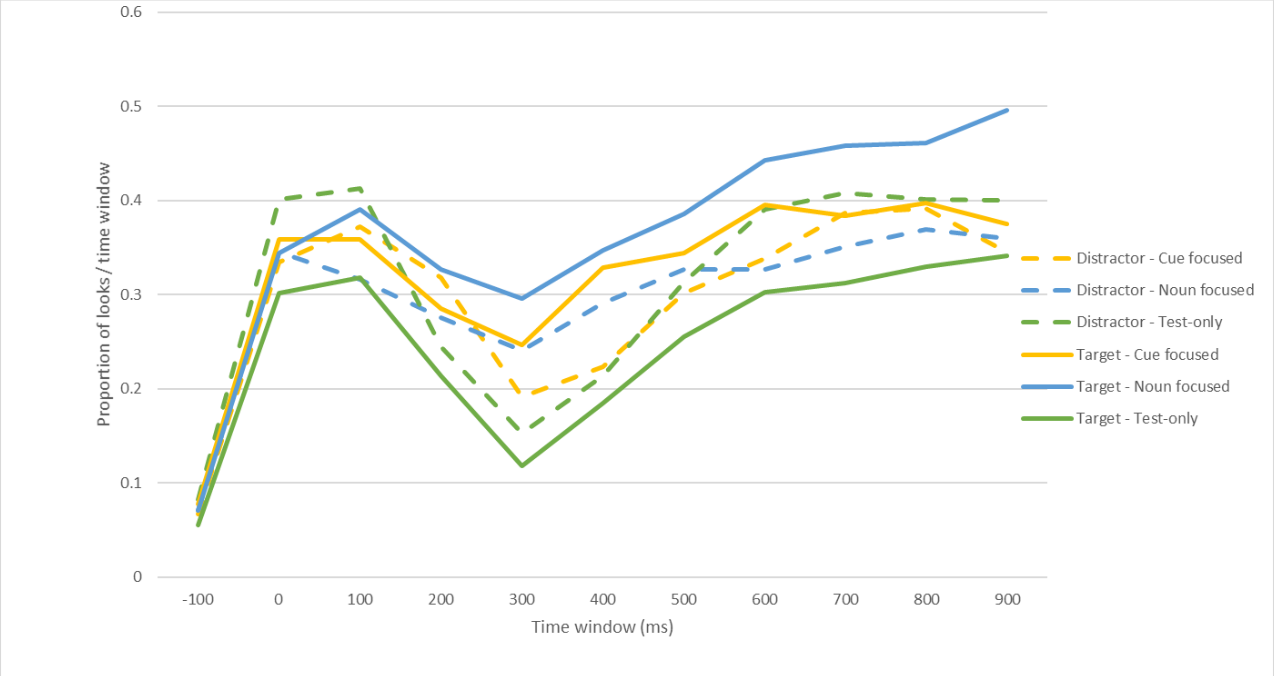


Figure 44. Eye-tracking – full passives and groups at immediate post-test

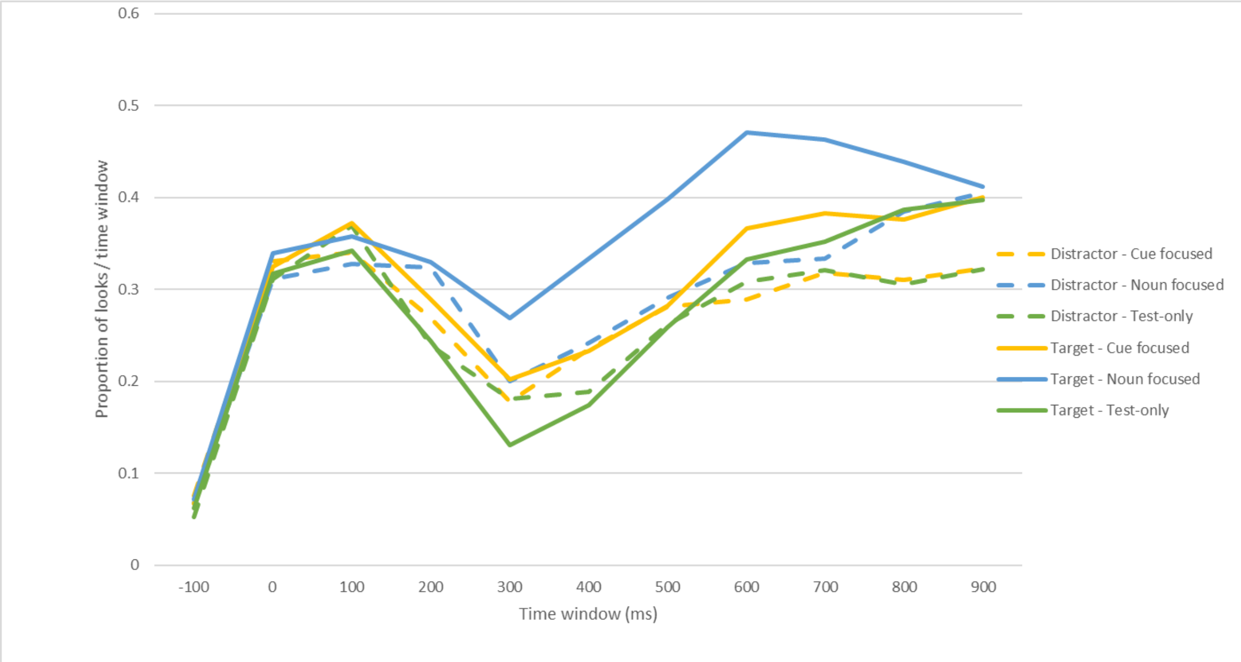


Figure 45. Eye-tracking – reduced passives and groups at immediate post-test

4.2.1.2 c) Reduced / full passives - Delayed post-test

Reduced passives were found to be significantly different to full passives at delayed post-test (700-800ms: Estimate = -6.03e-01, SE = 2.71e-01, t = 2.22, p = .03).

The cue focused group appeared to be more sensitive to the verb ending cue because looks to target peaked at this point for both full and reduced passives (see figure 46 and 47). *By*, or its absence, did not appear to affect looks to target for the full passives, but for the reduced passives a slight increase was seen around the second noun (700ms).

The noun focused group's looks to target exceeded looks to distractor around 500-600ms in the full passives and the reduced passives. This was later than at immediate post-test. At delayed post-test, this happened around *by*, or its absence, for both types of passive.

The test-only group showed a similar pattern for reduced and full passives with looks to target increasing around the verb offset. This was a similar pattern to the cue focused group but looks to target were proportionally less than the cue focused group. For the test-only group, looks to target did not diverge from looks to distractor after by or its absence.

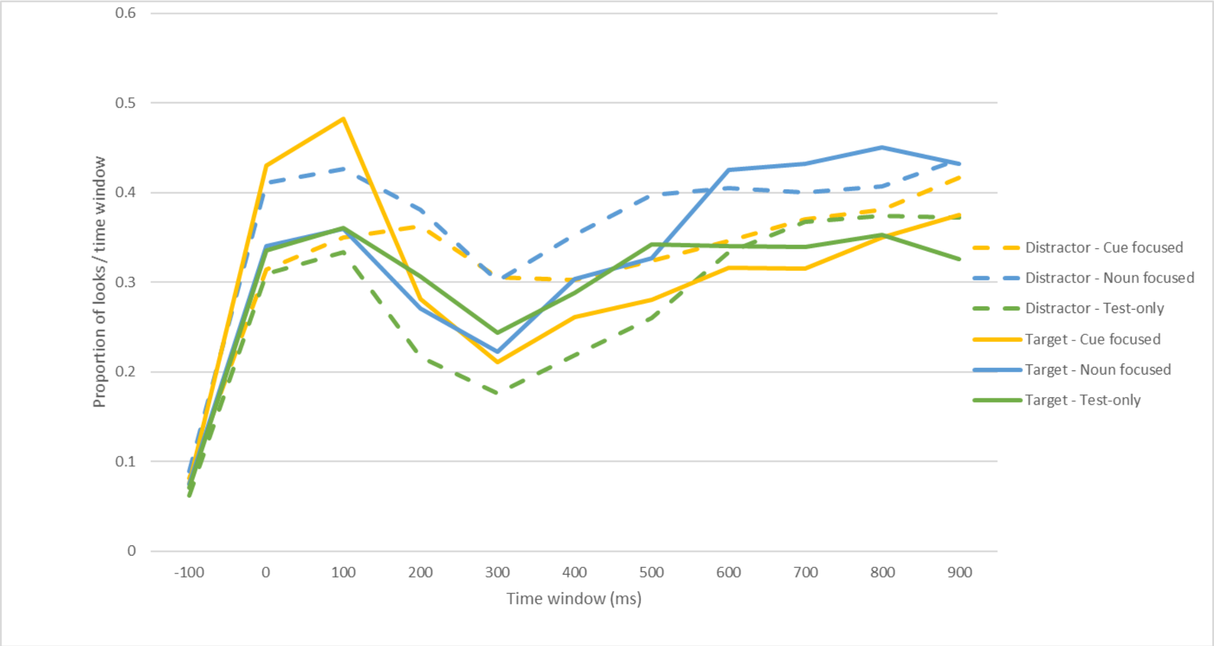


Figure 46. Eye-tracking – full passives and groups at delayed post-test

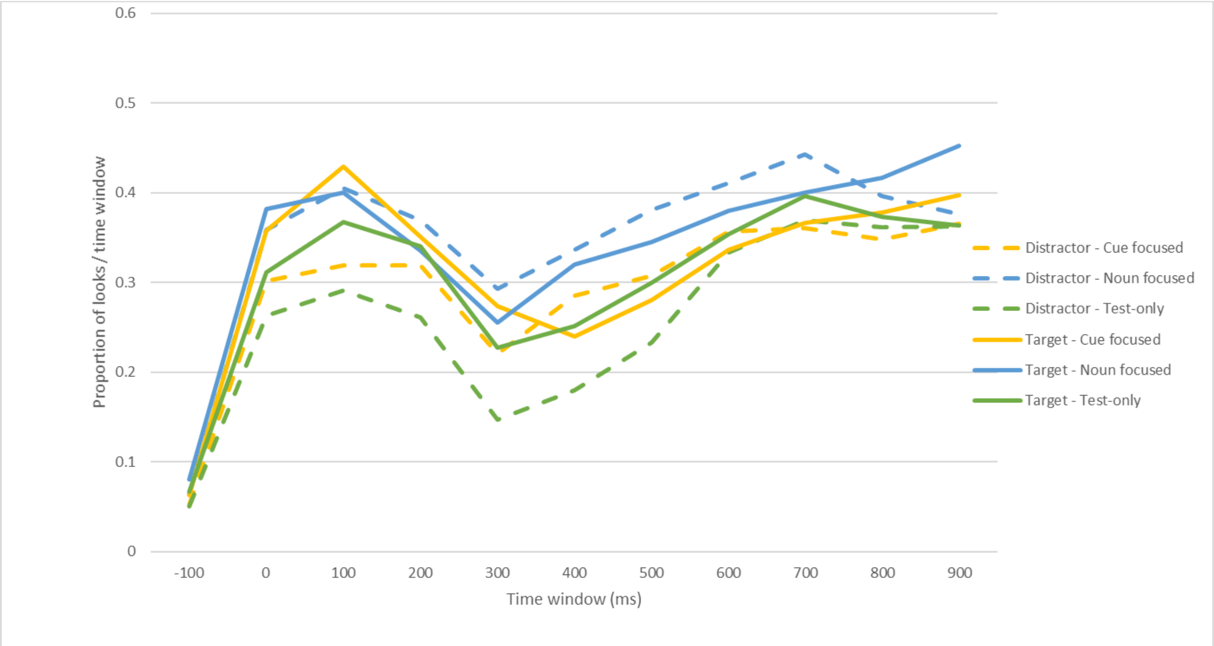


Figure 47. Eye-tracking – reduced passives and groups at delayed post-test

4.2.1.3 The effect of first and second noun animacy on the trained sentences

Four types of agent-patient noun combinations (both passives and actives) were included in the eye-tracking alongside images with reversible agent and patient roles (for example as in 34).

34a. Animate-animate: The boy is pushed by the girl OR The boy is pushing the girl

34b. Animate-inanimate: The boy is hidden by the rock OR The boy is hiding the rock

34c. Inanimate-inanimate: The car is hit by the bike OR The car is hitting the bike

34d. Inanimate-animate: The wheel is chased by the cat OR The wheel is chasing the cat

Noncanonical sentences such as actives with inanimate first nouns (as in 33c and d above) would be expected to cause more processing difficulty than more common constructions, such as passives with inanimate first nouns. Actives with inanimate first nouns could be misinterpreted as passives and vice versa. Since the cue focused training focused on the morphosyntactic cues more so than the noun focused training, it was expected that animacy effects would be reduced for this group. Therefore, it was expected that the cue focused group would look more to target for all animacy types after training. The effect of animacy after the verb is presented first. This analysis showed how animacy affected the processing of the morphosyntactic cues. The effect of first noun animacy was also analysed and is presented following section 4.2.1.3b.

4.2.1.3 a) Post-verbal animacy effects and voice (passive and active)

This section presents a comparison between the passive and active voice at pre-test for all participants combined. The below figures show that after the verb (300ms) active sentences were processed more accurately for all animacy combinations compared to passives (see figures 48 and 49).

For passives, only animate-animate (A-A) and inanimate-animate (I-A) items produced higher looks to target compared to distractor at any point in the sentence. Inanimate-inanimate (I-I) and animate-inanimate (A-I) sentences resulted in similar proportions of looks to target and distractor at all time windows.

For actives, inanimate-animate items seemed to cause the most processing difficulty. For these items, looks to target diverged from distractor upon hearing the verb, but then did not for the remaining time windows. This difference between passive and active inanimate-animate items was significant around 500ms (estimate = 2.71e-01, S.E. = 1.30e-01, t = 2.08, p = .04)

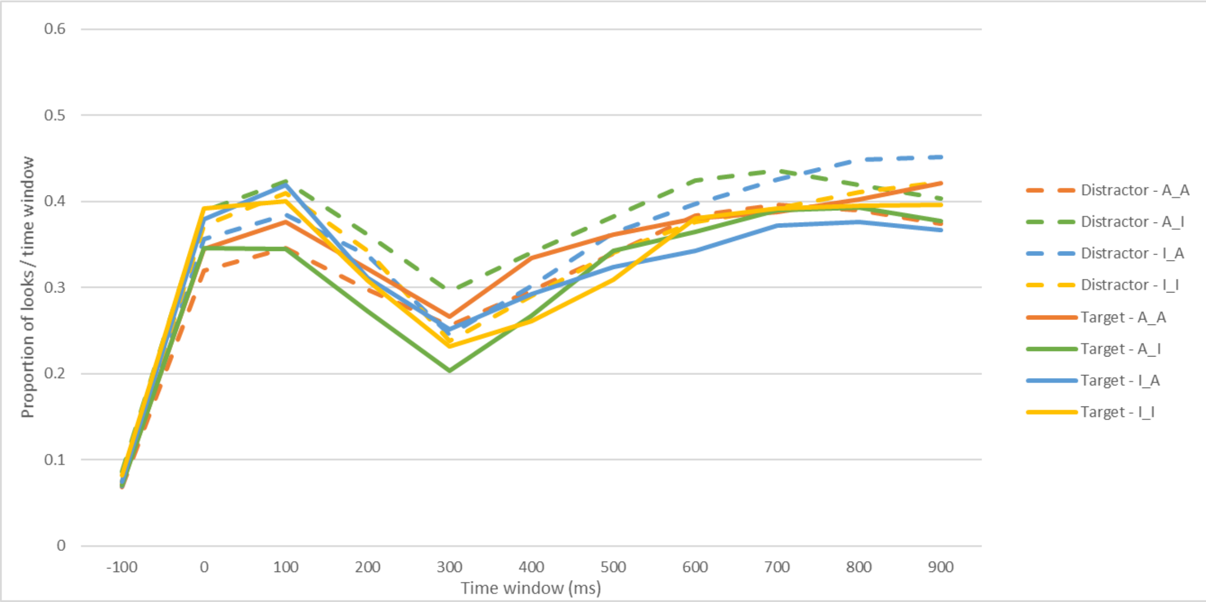


Figure 48. Eye-tracking - Post-verbal animacy and passive voice at pre-test

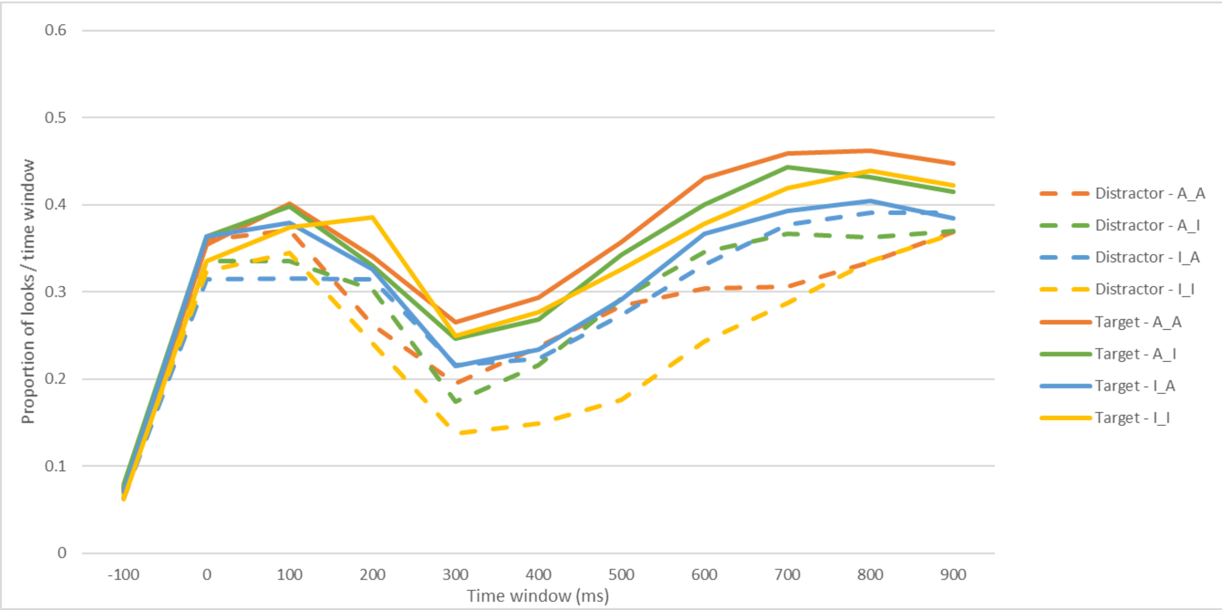


Figure 49. Pre-test - Eye-tracking - Post-verbal animacy and active voice at pre-test

4.2.1.3 b) Post-verbal animacy effects and voice before and after training

Immediate post-test – passives

The cue focused group, at immediate post-test, appeared to process inanimate-inanimate passives with more ease after the verb than at pre-test (as shown by an increase in looks to target – see figure 50 below). This was also indicated by a significant interaction of inanimate-inanimate, immediate post-test and passive voice found by the model at 500-600ms (Estimate = 6.58e-01, S.E. = 2.73e-01, t = 2.41, p = .01).

The noun focused group did not show this change between tests (see figure 51). For the noun focused group, changes were seen for animate-inanimate and inanimate-animate passives as looks to target increased after the verb (at pre-test looks to target did not surpass those to distractor – see figure 48 on the previous page). This was reflected in the model by a significant interaction of inanimate-animate items and noun focused group at 500-600ms (Estimate = 3.76e-01, S.E. = 1.85e-01, t = 2.04, p = .04).

Compared to the intervention groups, the test-only group's looks to target were much lower at immediate post-test (see figure 52 below).

Figure 50. Eye-tracking: Cue focused group at immediate post-test- Post-verbal animacy and passive voice

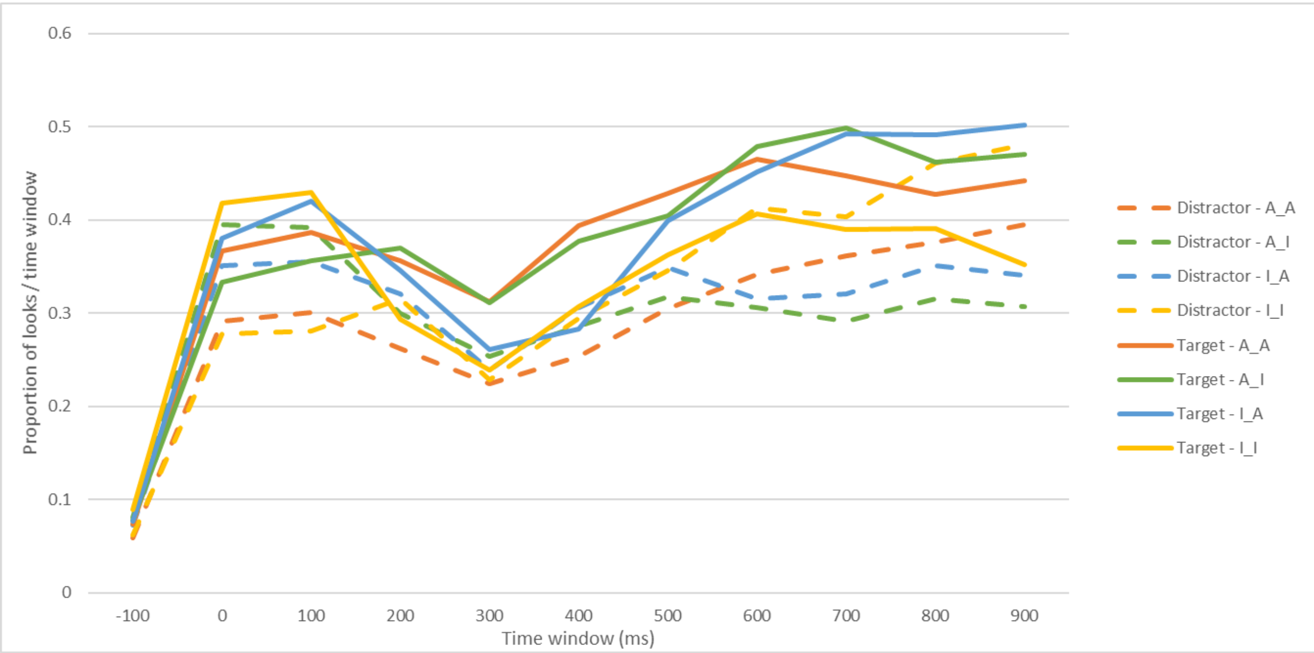


Figure 51. Eye-tracking: Noun focused group at immediate post-test- Post-verbal animacy and passive voice

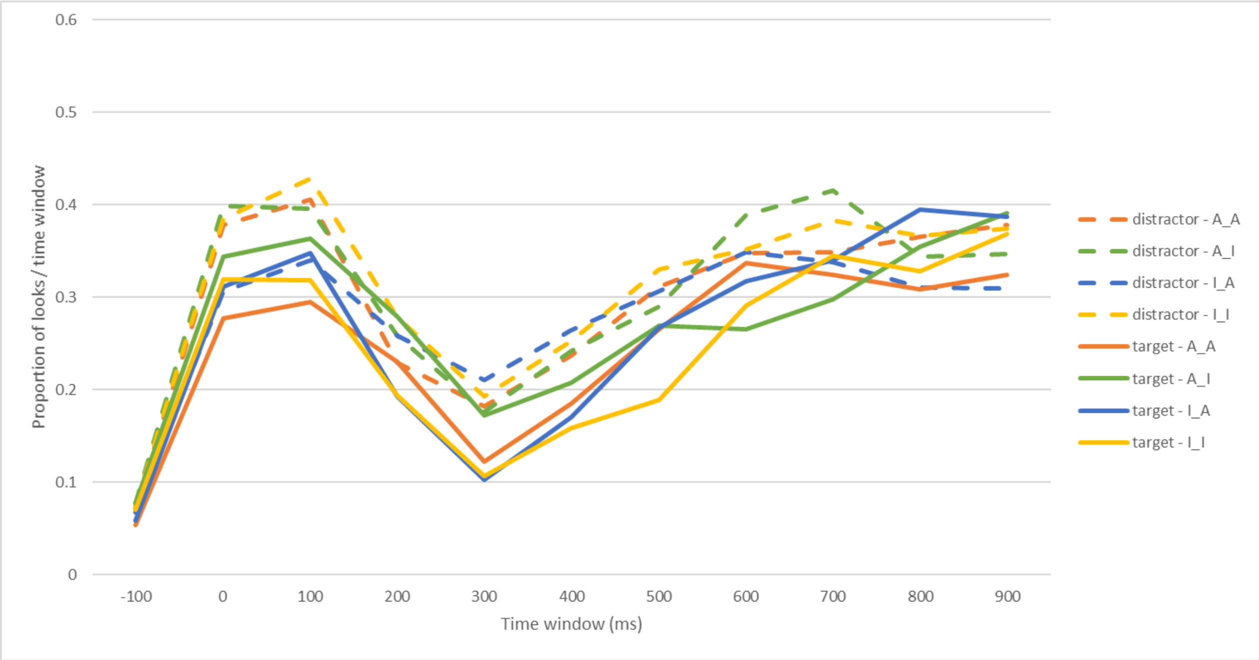


Figure 52. Eye-tracking: Test only group at immediate post-test- Post-verbal animacy and passive voice

Immediate post-test – actives

For all three groups there were no significant differences between pre and immediate post-test for active items (see appendix 19 for model results).

Delayed post-test – passives

For the cue focused group, at delayed post-test, for animate-inanimate passives, looks to target and distractor did not surpass looks to distractor until initially at 300ms (verb offset) and then more so at 500ms (*by*). This was reflected in the model in which a significant interaction of animate-inanimate items, delayed post-test and passive voice was found at 500-600ms (Estimate = 6.09e-01, S.E. = 2.79e-01, t = 2.19, p = .03). For passive animate-animate items, and those with inanimate first nouns, looks to target surpassed looks to distractor noticeably sooner than in the previous tests (from 0ms onwards) (see figure 53 below). This was not the case for the noun focused group.

For the noun focused group, looks to distractor were higher than looks to target till after 500ms (around *by*) for all animacy types, except animate-animate items. For sentences with inanimate first nouns, looks to target were lower at delayed than immediate post-test. This difference was significant at 100-200ms (estimate = -8.67e-01, S.E. = 3.95e+03, t = -2.16, p = .03).

For the test-only group, looks to target surpassed looks to distractor upon hearing the verb for all items except animate-inanimate. The test-only group did not show any significant change between test times.

Figure 53. Eye-tracking: Cue focused group at delayed post-test- Post-verbal animacy and passive voice

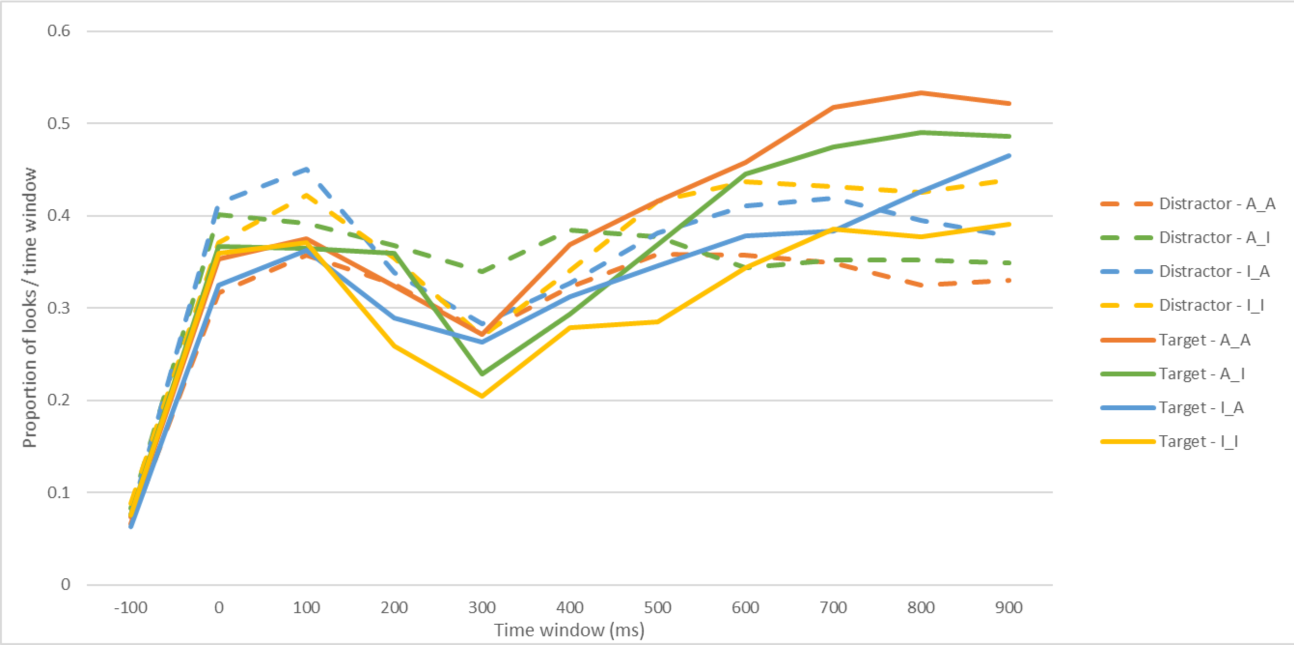


Figure 54. Eye-tracking: Noun focused group at delayed post-test- Post-verbal animacy and passive voice

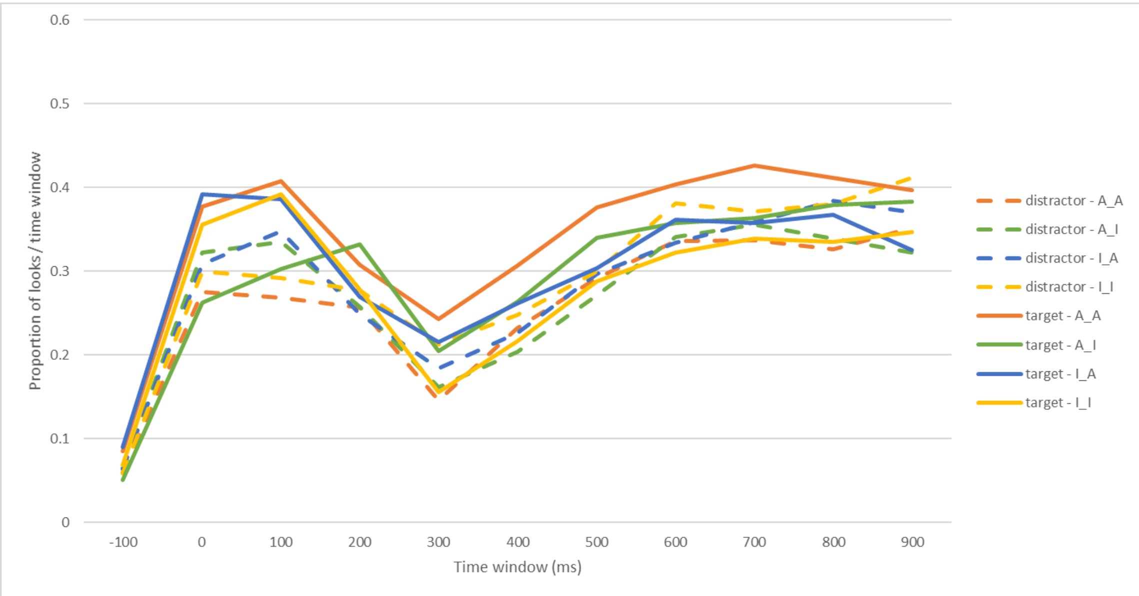


Figure 55. Eye-tracking: Test-only group at delayed post-test- Post-verbal animacy and passive voice

Delayed post-test – actives

For the cue focused group, the interpretation of active items changed greatly between pre-test and delayed post-test (as shown in figure 56). Animate-animate items went from being processed most accurately and soonest to causing the most processing difficulty. Inanimate-inanimate and animate-animate sentences seemed to tend to be misinterpreted as passives (given the finding that looks to distractor surpassed looks to target). Inanimate-animate active sentences were processed more accurately than at the other test times. Although these differences are apparent in the figures, they were not statistically significant.

For the noun focused group, for actives, there was very little change between tests. Figure 57 below suggests that for animate-inanimate actives, looks to target diverged from distractor sooner compared to immediate post-test. This observed change was not found to be significant by the models.

For the test-only group, as in the immediate post-test, looks to target surpassed distractor for animate-inanimate and inanimate-animate sentences (see figure 58). These sentences appeared to cause the learners less problems across the test times.

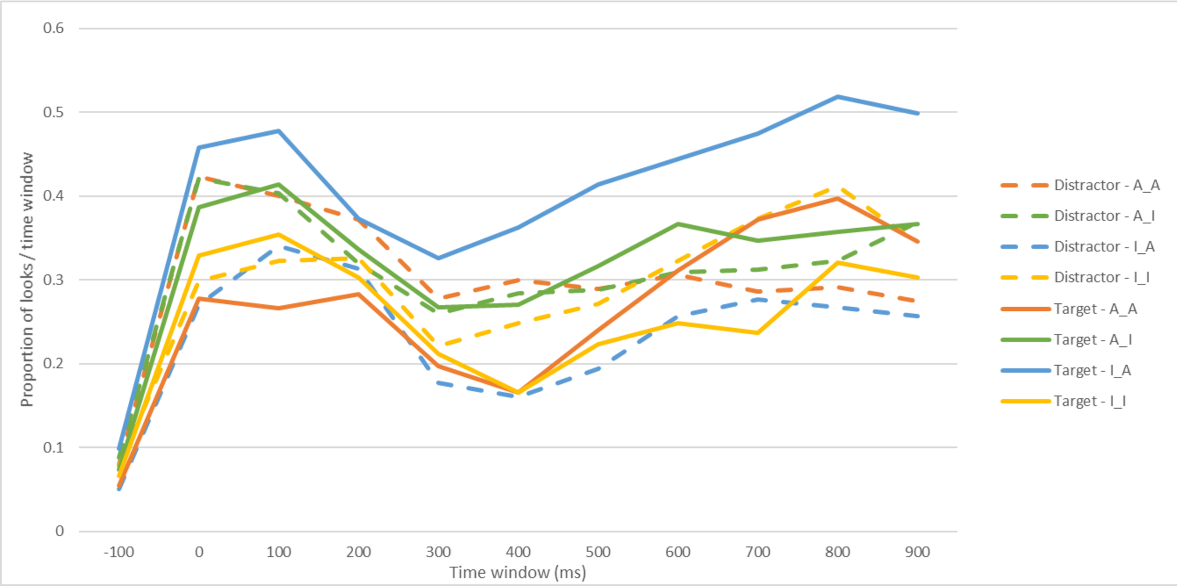


Figure 56. Eye-tracking: Cue focused group at delayed post-test- Post-verbal animacy and active voice

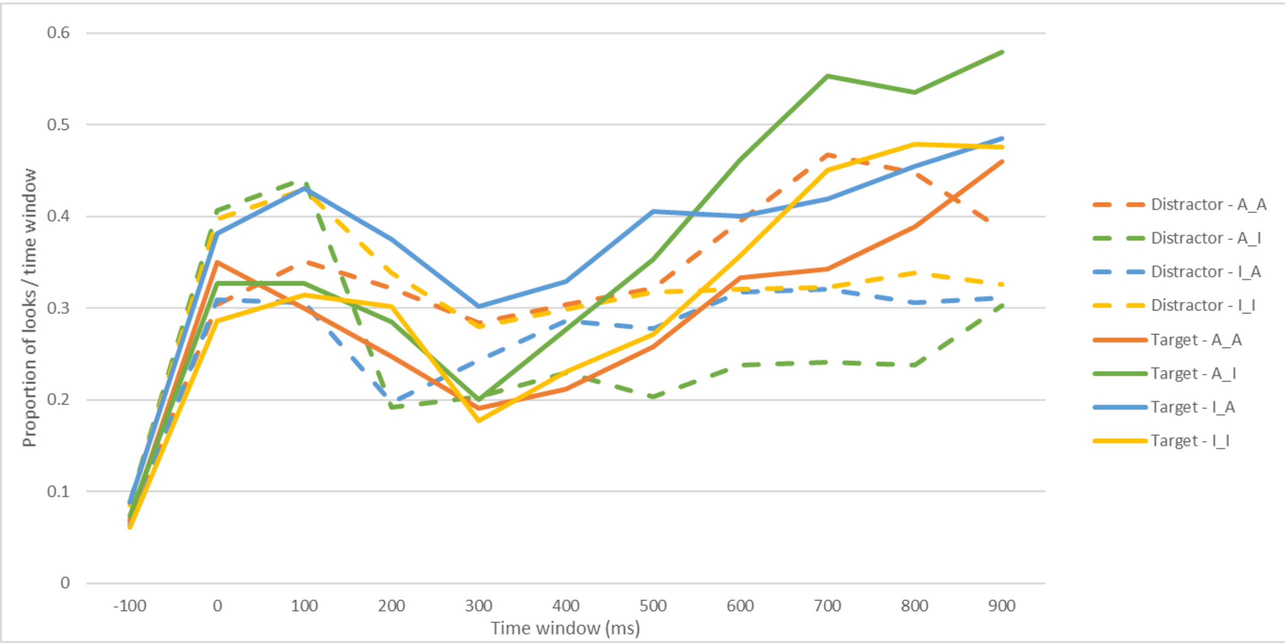


Figure 57. Eye-tracking: Noun focused group at delayed post-test- Post-verbal animacy and active voice

Figure 58. Eye-tracking: Test-only group at delayed post-test- Post-verbal animacy and active voice

4.2.1.3 c) Pre-verbal first and second noun animacy effects

This section presents the data from before the verb in order to investigate the effect of the animacy of the first noun on its interpretation.

There was a main effect of inanimate-animate items before the verb (-700ms-0ms). Significant interactions of inanimate-animate items and passive voice were also found before the verb as were interactions with group and time (see appendix 19 for model results). The following figures illustrate the statistically significant interactions found by the model.

The figures below show the difference between active and passive items for each animacy type (at pre-test collapsed across groups). Animate-inanimate items seemed to be processed more accurately upon hearing the first noun for active items (figure 59) than passive items (figure 60) - looks to target diverged from looks to distractor prior to the verb (around -300).

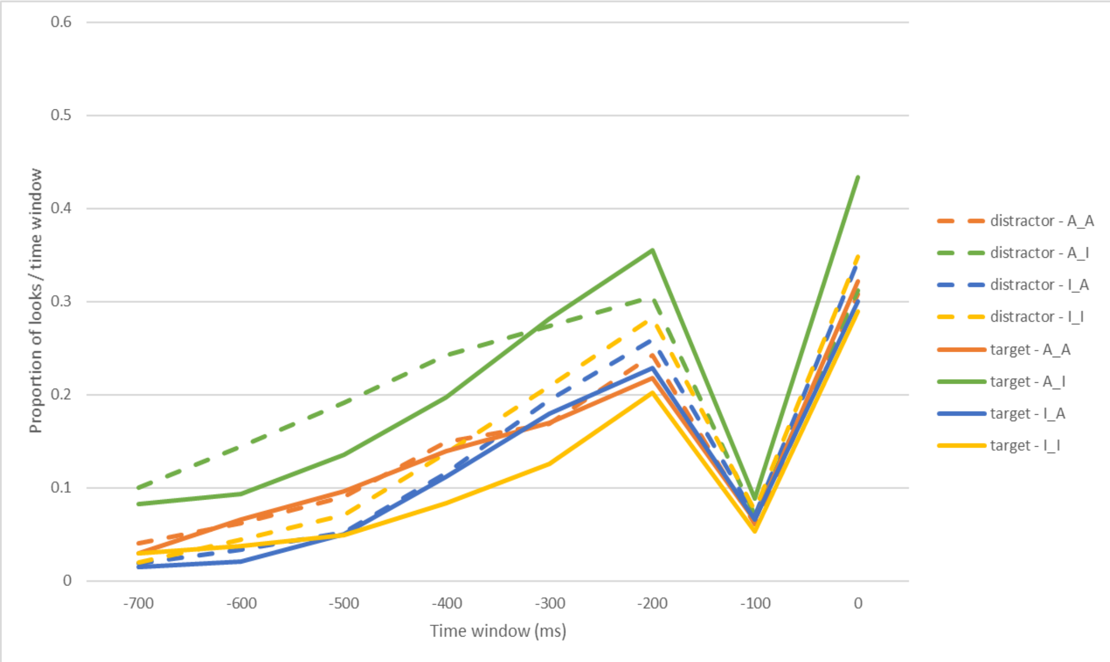


Figure 59. Eye-tracking: Pre-verbal animacy and active voice at pre-test

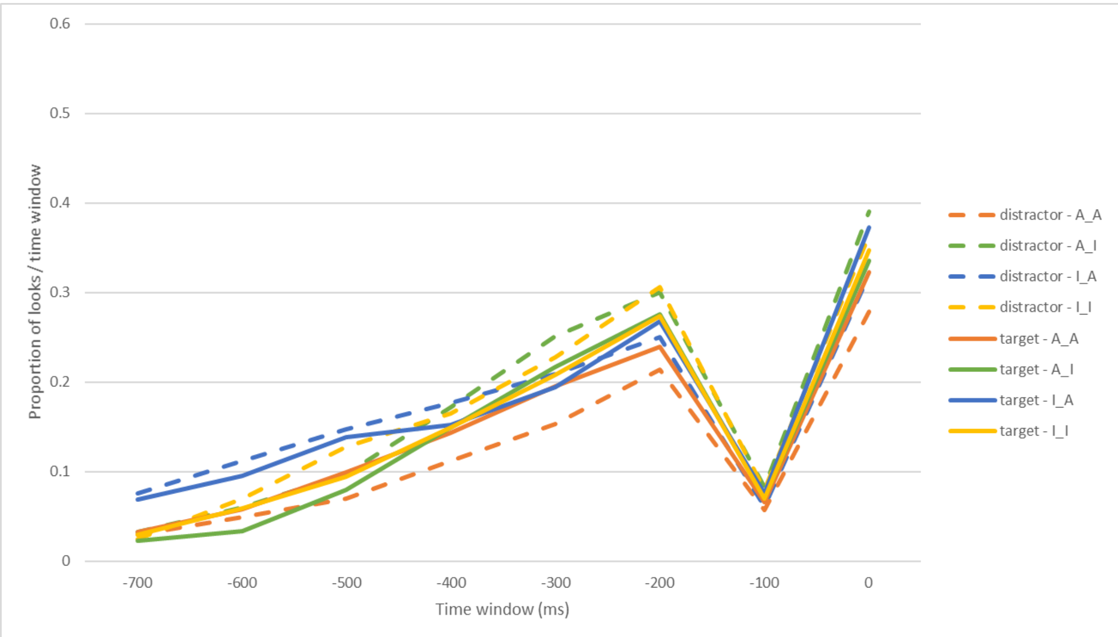


Figure 60. Eye-tracking: Pre-verbal animacy and passive voice at pre-test

Immediate post-test – passives

For passives, the cue focused group looked more to target than distractor for inanimate-animate passive items than the noun focused or test-only group around the first noun (see figures 61 and 63). This was shown by a significant main effect of the noun focused group at -700ms (estimate = -1.03e+00, S.E. = 3.93e+03, t = -2.29, p = .02), -600 (estimate = -1.17e+00, S.E. = 4.73e-01, t = -2.467, p = .01) and -500 (estimate = -1.12, S.E. = .49, t = -2.269, p = .02). This suggested that at immediate post-test the cue focused group entertained the possibility of a passive interpretation of inanimate agents more so than the noun focused group.

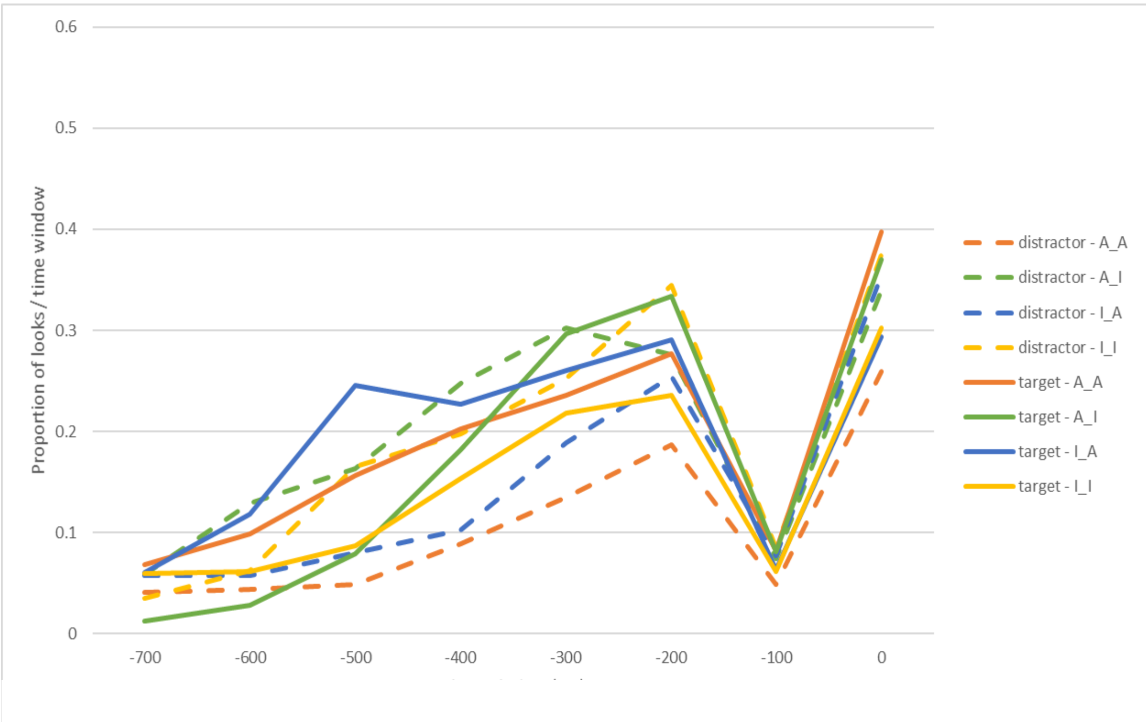


Figure 61. Eye-tracking: Cue focused group at immediate post-test- pre-verbal animacy and passive

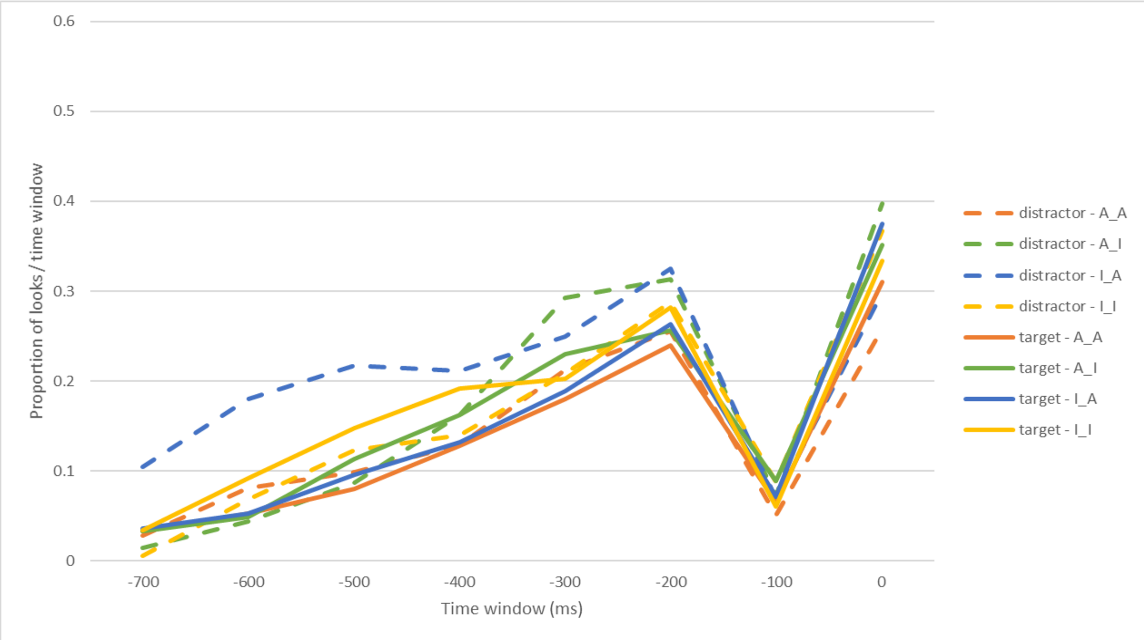


Figure 62. Eye-tracking: Noun focused group at immediate post-test- pre-verbal animacy and passive voice

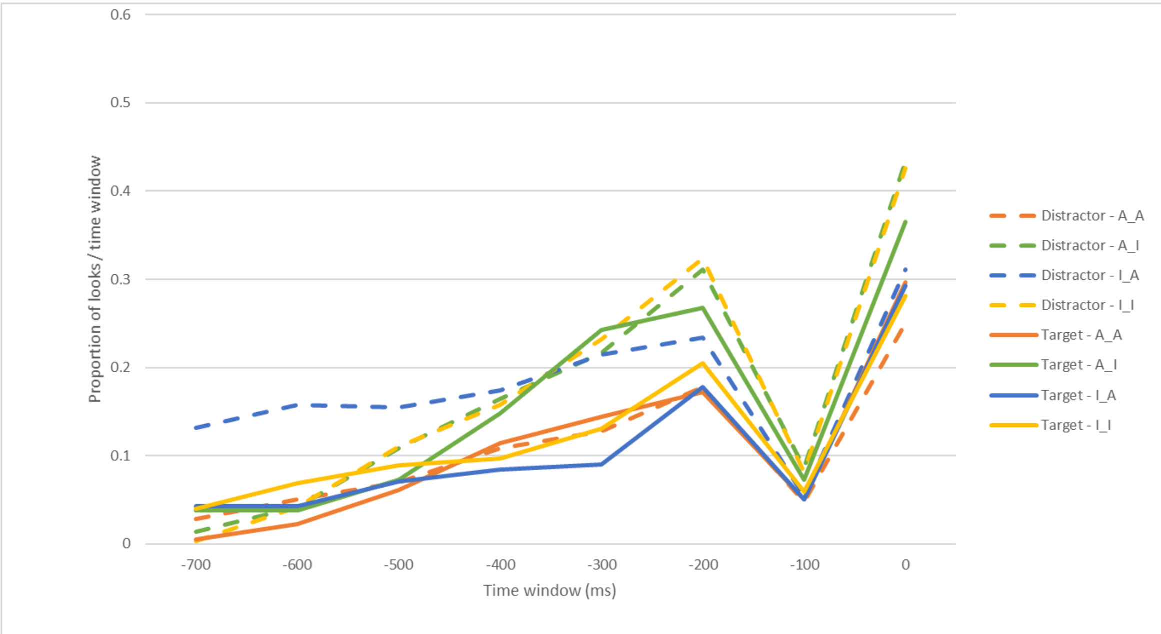


Figure 63. Eye-tracking: Test-only group at immediate post-test- pre-verbal animacy and passive voice

Immediate post-test – actives

For actives, the cue focused group was the only group for which looks to target surpassed distractor, but only for animate-inanimate actives. This was not found to be significant by the model. The other groups did not look more to target than distractor any animacy combination (see appendix 20 for figures).

Delayed post-test – passives and actives

For both passives and actives, these differences between groups were not maintained to delayed post-test – all three groups' looks to target did not clearly diverge from distractor for all animacy types (see appendix 20 for figures).

4.2.2 Untrained constructions

The untrained constructions, those not practised in the intervention sessions, were analysed separately from the trained constructions. The untrained constructions were either present perfect passive or present perfect progressive (active).

e.g. 35a. The dog has been carried by the boy

35b. The dog has been carrying the boy

This section focuses on the untrained constructions and address the following variables;

- Training (either cue focused or noun focused) or no training (test-only control group) at each test time (pre-test, immediate post-test, and delayed post-test).
- Passive and active voice.

The untrained items were included to investigate whether learners could generalise what they had learnt in the training and transfer it to untrained items. In order to answer the RQs, the learners' eye-tracking data for active and passive sentences is presented here. It was not an intention of this study to investigate untrained items and animacy effects. Animacy effects for untrained items can be seen in the model results in appendix 19 but are not discussed in this thesis.

The mixed-effects models used for the untrained constructions were the same as the models used for the trained constructions. Again, the predictor variables were sum coded so that the intercept moved to the middle of the categories. The control variable was proficiency, the fixed factor predictors were animacy, test time, group, voice and reduced / full passive. The random slopes were subject and item. The following model was built in R software using the lmer function (R Core Team).

*Final model: TW ~ prof + animacy * time * group * voice * reduced +(1 | sub) + (1 | item)*

4.2.2.1 The effect of voice (active and passive) on untrained constructions

The following figures present a comparison of passive and active untrained items (present perfect passive) at each test time. At pre-test, the three groups performed similarly on the untrained items as they did at pre-test on the trained simple passive items (figures 36 on p.182 and 64 below). Looks to target did not diverge much from looks to distractor. Active untrained items, (present perfect continuos) were judged less accuratey than the trained items (figure 37 on p.183 and figure 65 below).

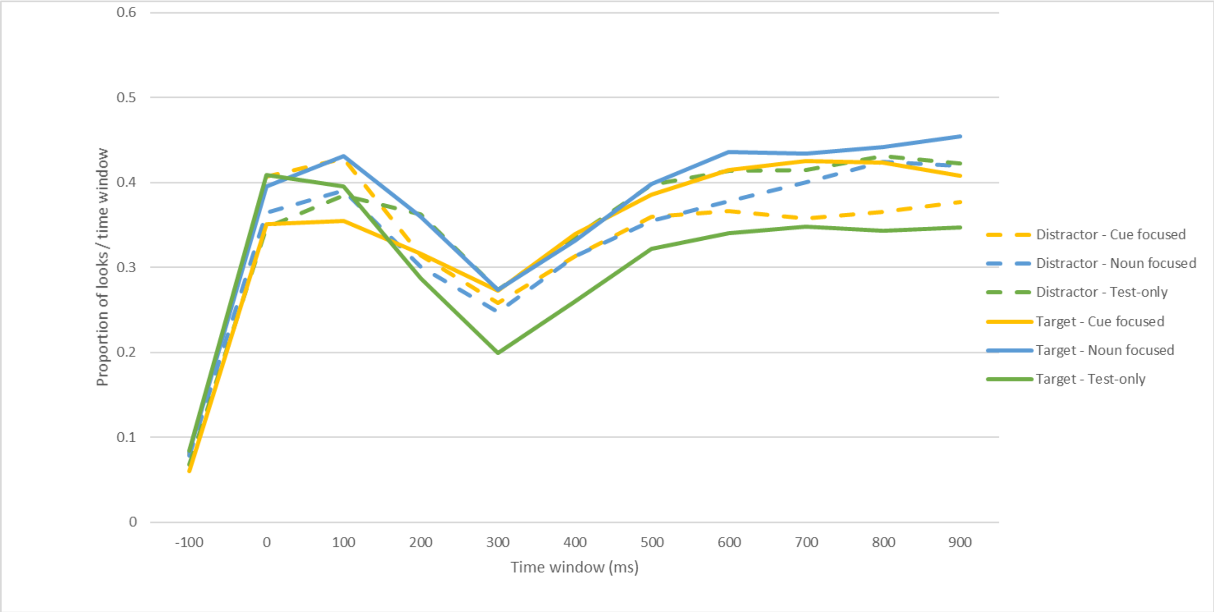


Figure 64. Eye-tracking: Untrained passives and group at pre-test

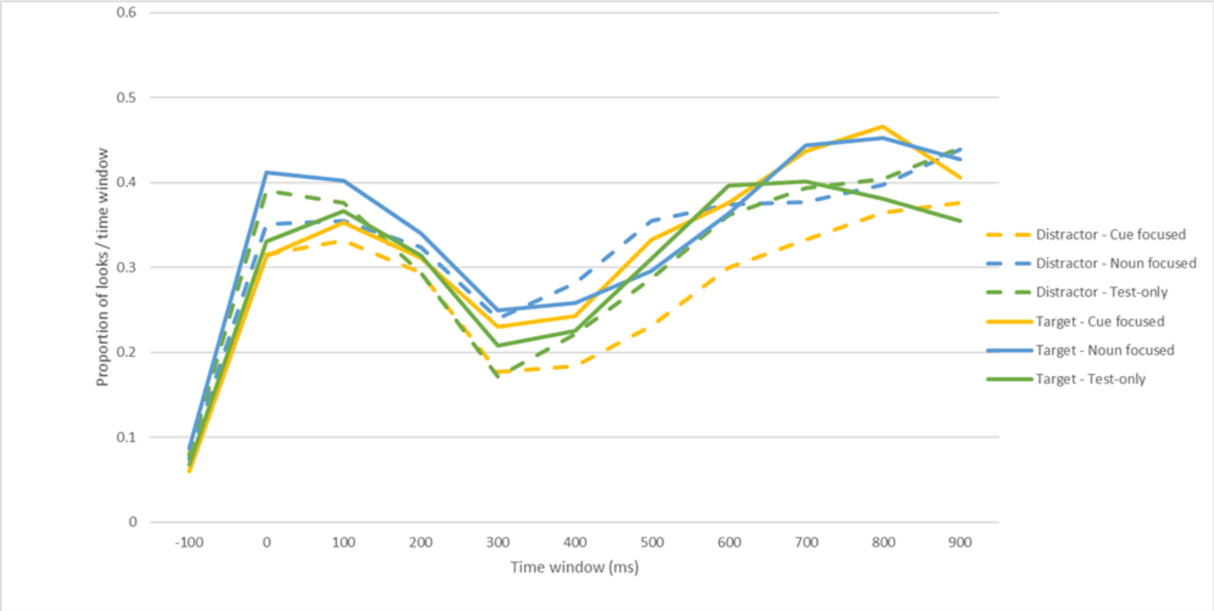


Figure 65. Eye-tracking: Untrained actives and group at pre-test

At immediate post-test, the cue focused group, for both passives and actives, looked to target slightly sooner than at pre-test (0ms – 300ms during the verb) but then appeared to misinterpret actives as passives after hearing the verb (300ms onwards).

The noun focused groups' looks to target diverged most from looks to distractor from 500ms onwards for both passives and actives (around *by* or the absence of *by*) (figure 66 and 67).

The test-only group consistently looked to distractor more than target throughout for passives. For actives, looks to target and distractor were roughly the same.

Despite the visual difference seen in the figures none of the groups were found to be statistically significantly different to each other.

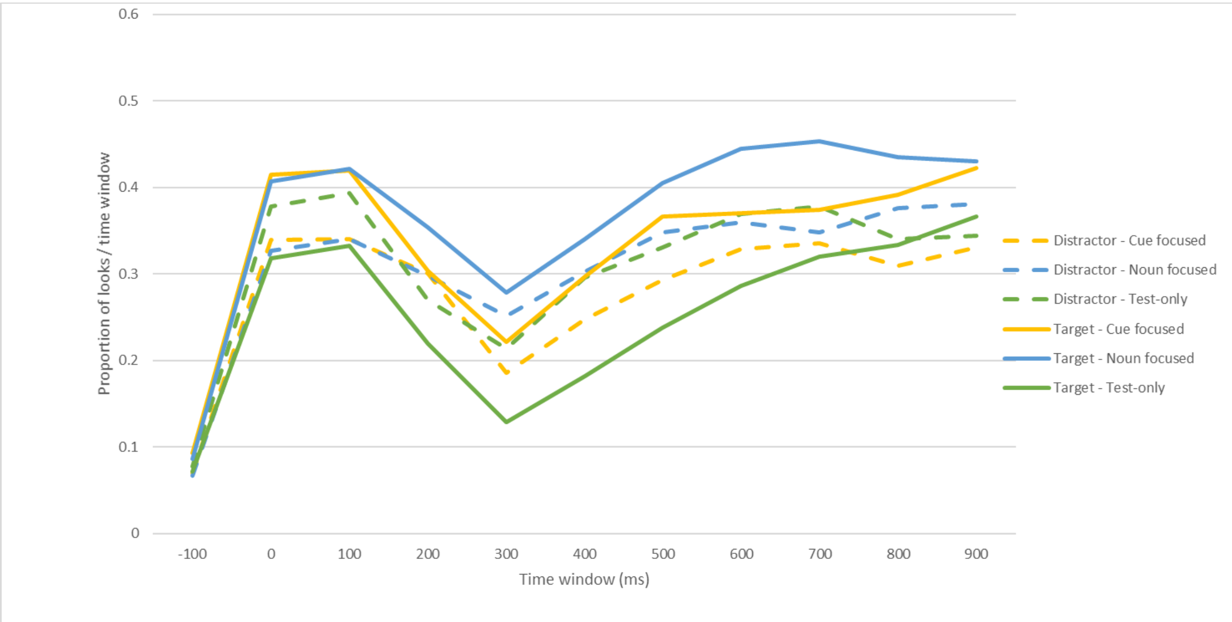


Figure 66. Eye-tracking: Untrained passives and group at immediate post-test

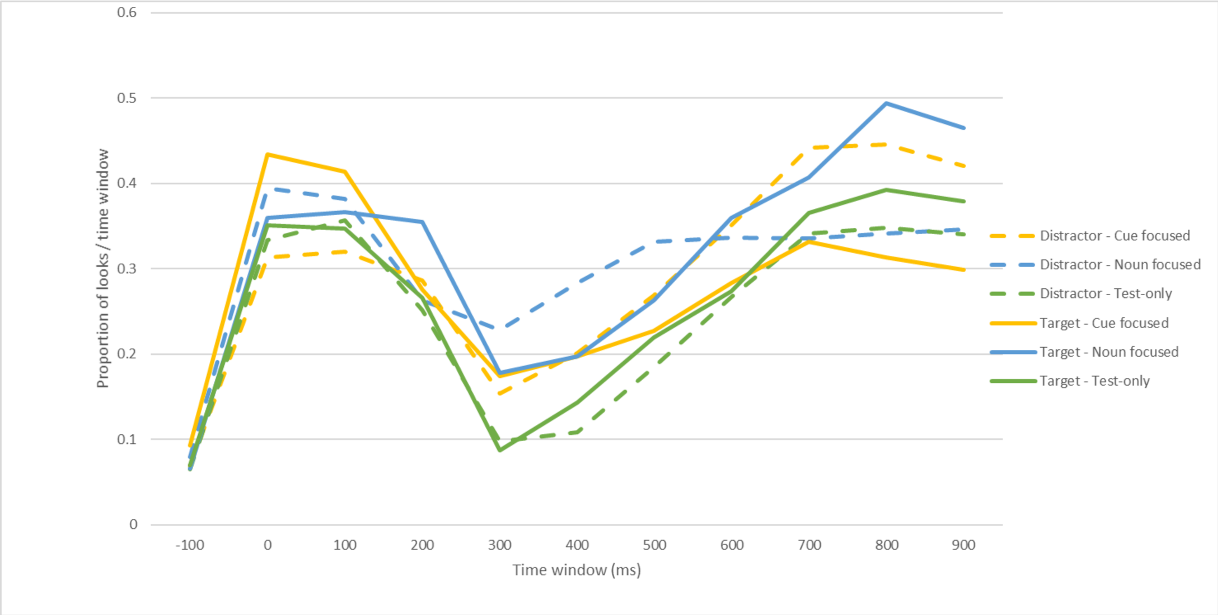


Figure 67. Eye-tracking: Untrained actives and group at immediate post-test

At delayed post-test, for passives, the cue focused group looked more to target than distractor, after the verb, compared to the other groups (300ms) (see figure 68 below).

The noun focused group waited till after *by* (or its absence, in the active sentences) (500ms). An interaction at 500ms between the noun focused group, delayed post-test and voice was found to be approaching significance by the model (Estimate = 4.92e-01, S.E. = 2.79e-01, t = 1.77, p = .08). This interaction was not seen for the trained passives.

The test-only group's looks to target did not diverge much from looks to distractor in any time window after the verb.

For actives, all three groups looked most to target during the verb (see figure 69 below). All three groups showed uncertainty straight after the verb as looks to target did not surpass looks to distractor.

Compared to the other groups, the cue focused group looked more to target than distractor after the verb ending and throughout the remaining time windows. The noun focused group appeared to wait until the absence of *by* to look to target after the verb (as they did for passives). The test-only group did not look more to target than distractor after the verb.

Despite the differences observed in the figures, there were no significant differences between times and groups found by the model.

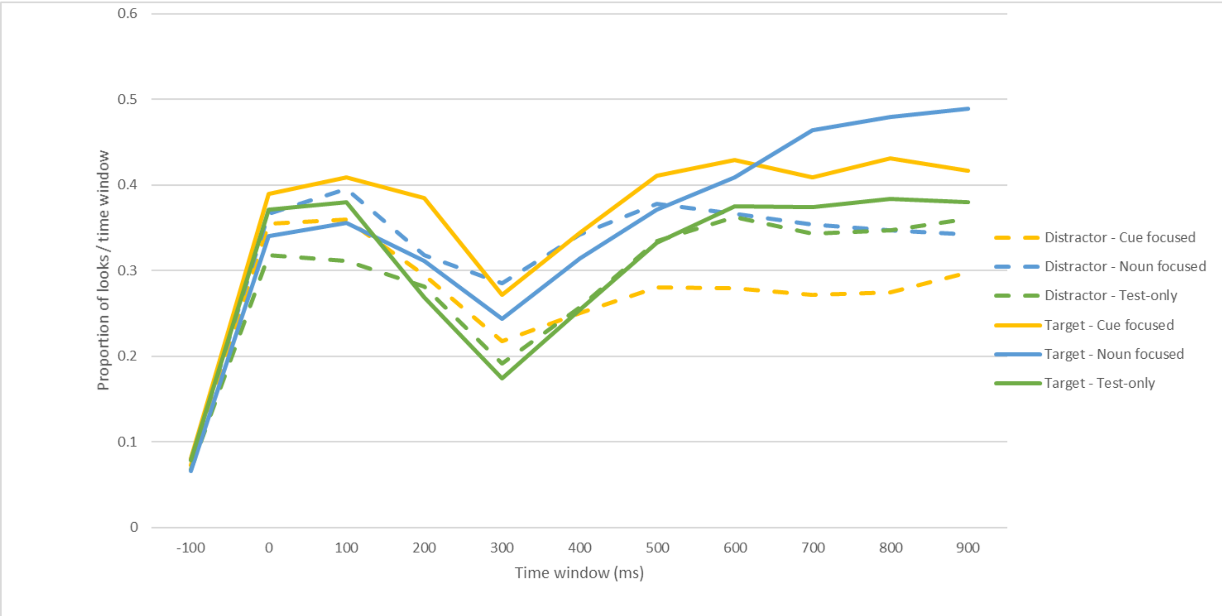


Figure 68. Eye-tracking: Untrained passives and group at delayed post-test

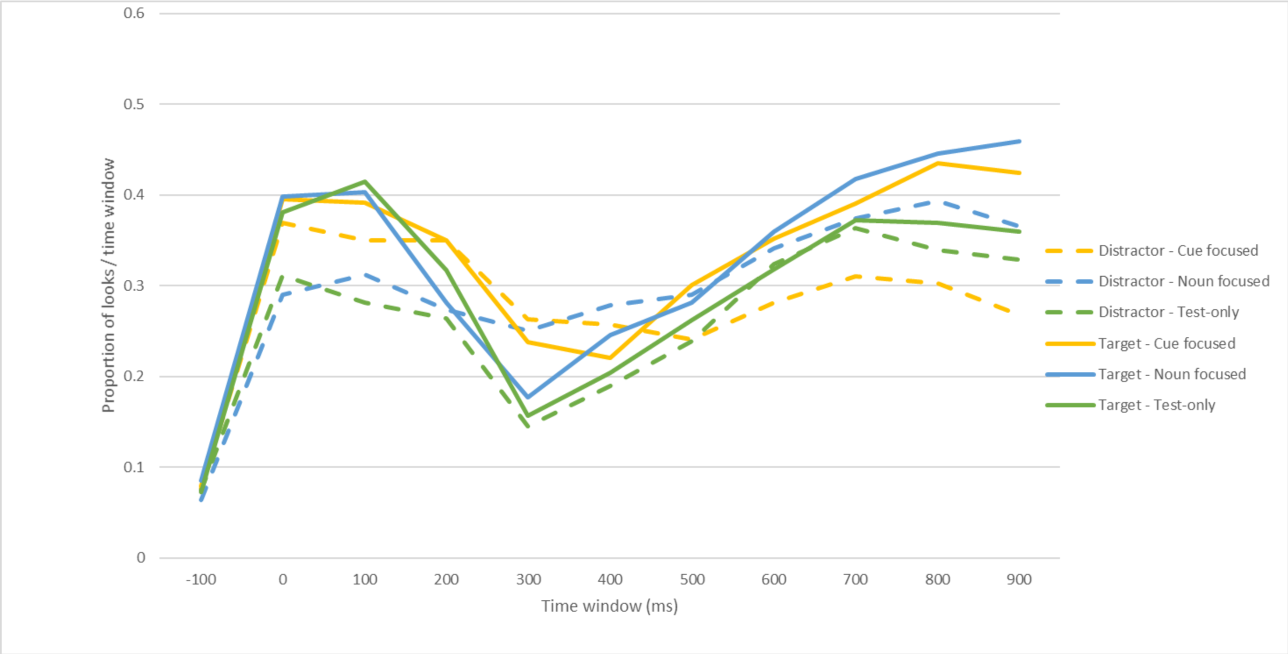


Figure 69. Eye-tracking: Untrained actives and group at delayed post-test

4.2.2.2 The effect of type of untrained passive (reduced / full passives)

There were no significant main effects of, or interactions, for full versus reduced passives at any test time. This lack of significant effects was also the case for the trained passives (full versus reduced). A main effect for the noun focused group was found for all test times. This is indicated by a difference in this group's behaviour for both types of passive compared to the other groups. The noun focused group's looks to target fell after *by* for reduced passives but rose at this point for full passives (see appendix 21 for figures). This was also seen for the trained items at immediate post-test (see section 4.2.1.2b figure 44). Since this was the only finding of interest for untrained full / reduced passives the figures are not presented here. The figures for reduced and full untrained passives are in appendix 21.

4.2.3 Native speaker processing of the passive voice – evidence from eye-tracking

This section presents eye-tracking data from the NS group. 29 NSs did the eye-tracking test once. The passive and active voice are compared and animacy effects are presented in order to respond to the third research question (see below). Reduced / full passives and tense (trained / untrained) did not have a significant effect on the NSs' behaviour therefore these are not presented in detail. Model results can be found in appendix 19.

RQ3) To what extent do native English speakers show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

To what extent is native speaker processing of the passive different to learners for:

It was expected that passive and active sentences might have been processed differently since the passive voice is less common and more complex. Animacy effects were expected since noncanonical sentences such as actives with inanimate first nouns would be expected to cause more processing difficulty than more common constructions, such as passives with inanimate first nouns. Actives with inanimate first nouns might have been expected to be misinterpreted as passives and vice versa prior to hearing the verb.

Mixed-effects models were used to analyse the NS data. Fixed effects were animacy, voice, reduced / full passives and tense (trained / untrained) and random slopes were included for subject and item, as follows:

*Model > TW ~ animacy * voice * reduced * tense + (1 | sub) + (1 | item)*

4.2.3.1 Native speaker processing – comparing passive and active items

Although the model did not find a significant main effect of voice the below figures show that active and passive items might have been processed differently by NSs. For passive items, the NSs appeared to wait to hear the verb before looks to target diverged from looks to distractor. There was also a slight increase in looks to target around 500ms (the offset of *by*). Looks to target for active items increased after the verb inflection (0-300ms) and were high prior to the verb. This may be because NSs expected items to be active because this construction is more common.

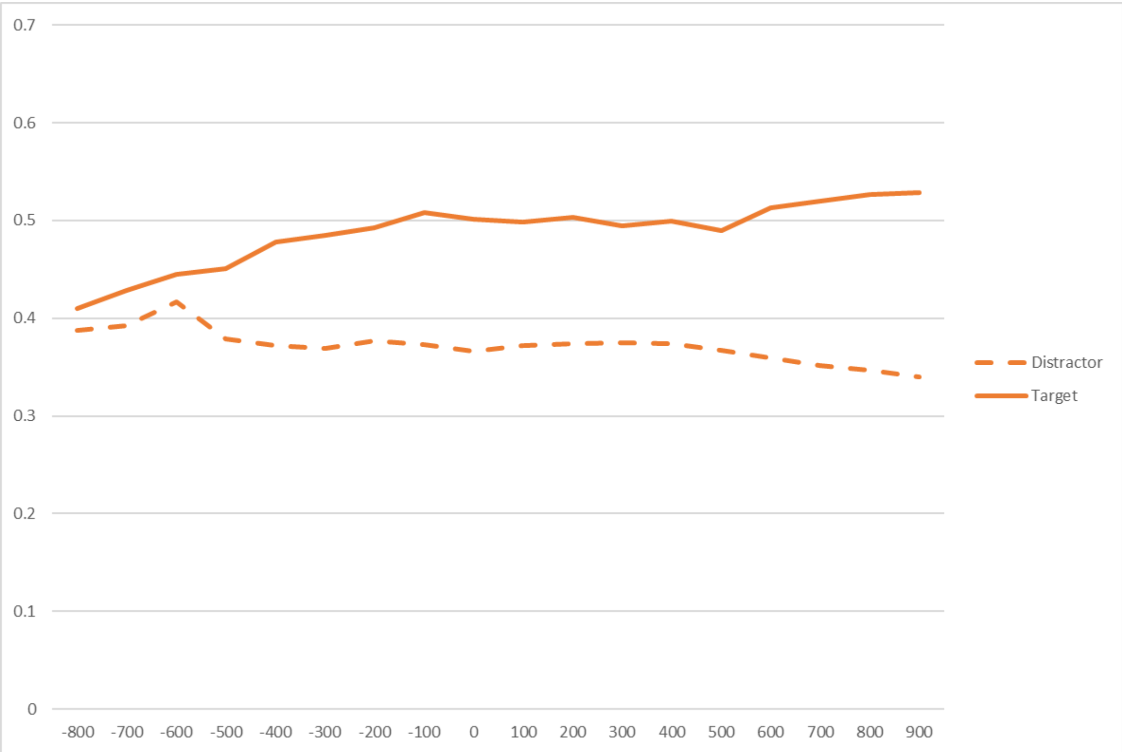


Figure 70. Eye-tracking: Native speakers and the passive

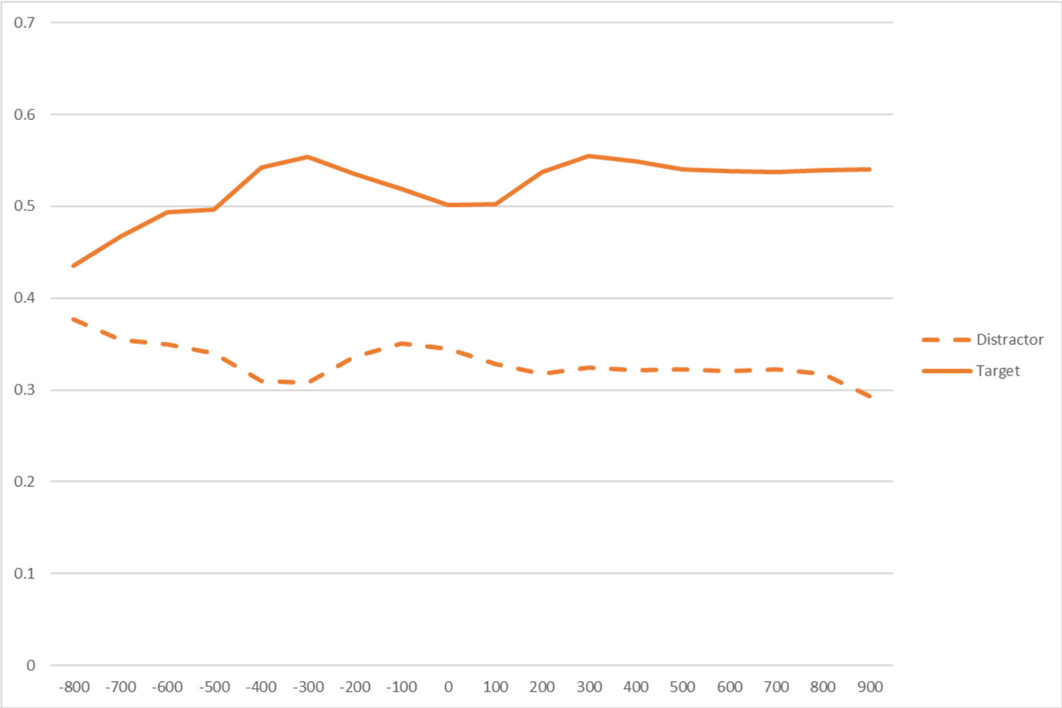


Figure 71. Eye-tracking: Native speakers and the active

4.2.3.2 Native speaker processing and the effects of first noun animacy

The above difference between passive and active items prior to the verb may have been driven by the animacy effects of the first noun. For active items, all items, except animate-inanimate items, were interpreted correctly prior to the verb (see figure 72 below).

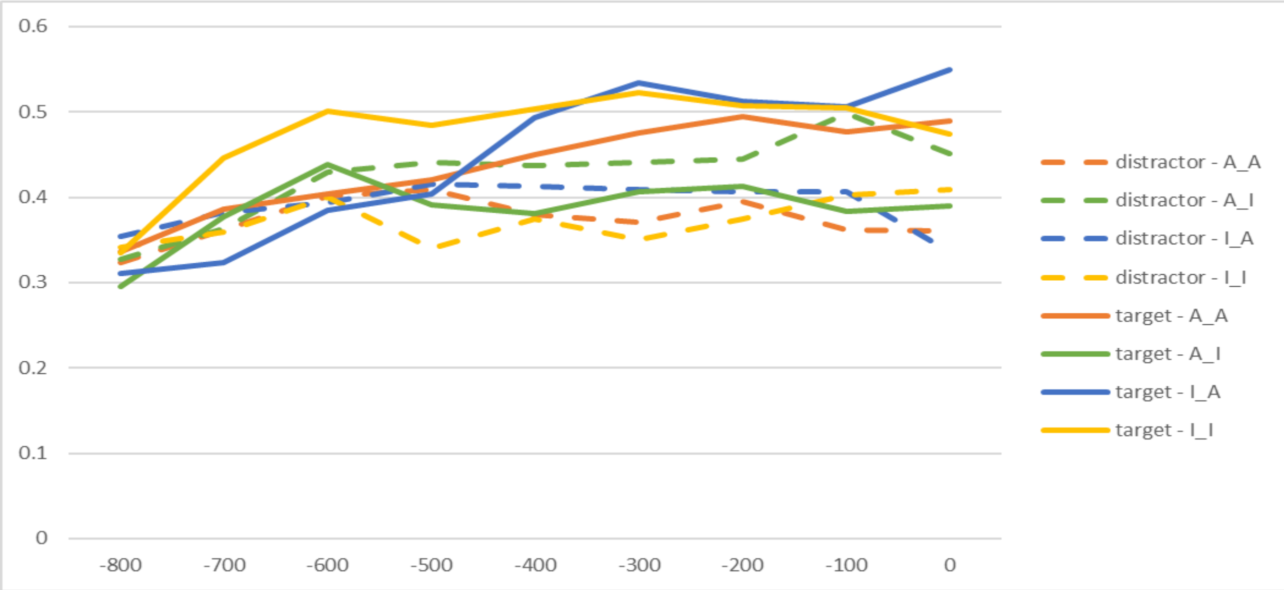


Figure 72. Eye-tracking: Native speakers and pre-verbal animacy effects and active items

For passives and most animacy combinations, looks to target were lower than distractor than for active items (see figure 73 below). This suggests that the participants assumed most items to be active sentences prior to the verb. This was not true for inanimate-animate items. Looks to target were greater than to distractor for these items. These were the most likely to be passives, since an inanimate first noun is most likely to be a patient.

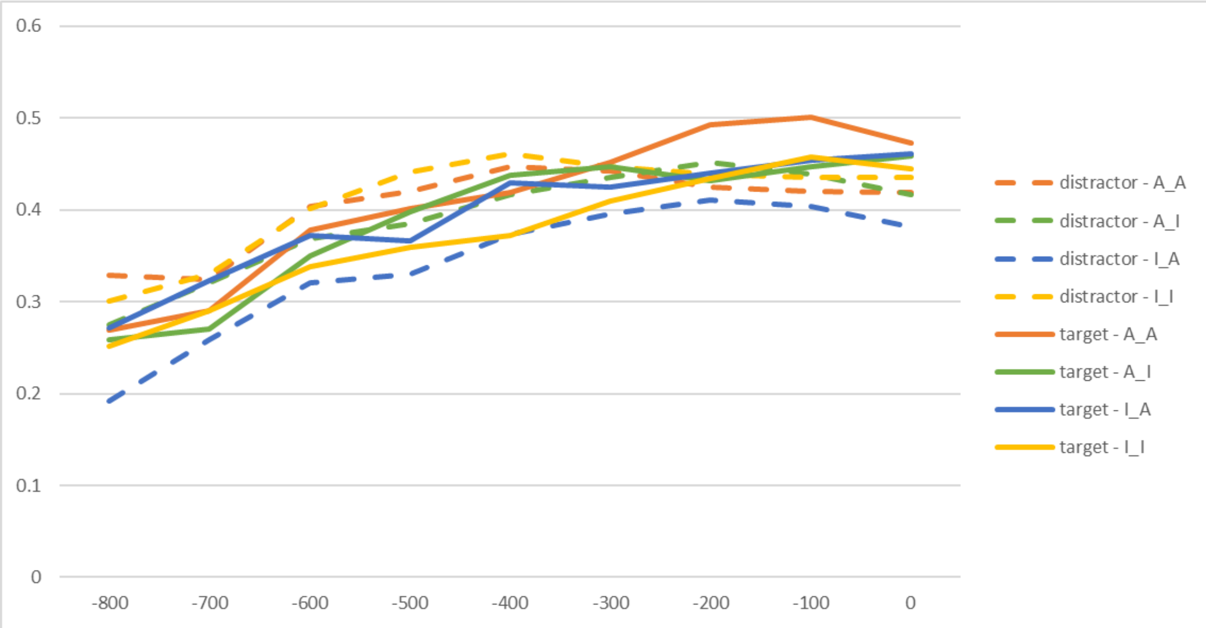


Figure 73. Eye-tracking: Native speakers and pre-verbal animacy effects and passive items

4.2.3.3 Native speakers and learner processing – a comparison

This section compares the NS data and the learner data at pre-test. This provided some insight into processing differences between NSs and learners. The NS data were not compared to the learner data post-intervention. This was because it was not an aim nor expectation of the current study that the interventions would make the learners' processing more "native-like". It was expected that the NSs and learners would behave differently for passive and active sentences.

Mixed-effects models were used to analyse the NS data and the learner data at pre-test. Fixed effects were group (native or learner) animacy, voice, reduced / full passives and random slopes were included for subject and item, as follows:

*Model > score ~ group * voice * animacy * reduced + (1 | sub) + (1 | item)*

Since reduced / full passives did not affect processing for the learners at pre-test or the NSs, this variable is not presented. Animacy is also not presented since there were no significant differences between the groups (see appendix 19 for model results). The effects of voice (passive / active) are the main focus of this section.

4.2.3.3 a) Native speakers and learner processing – the effects of voice

Passive sentences were processed much less accurately by the learners than the NSs – as shown by looks to target (see figure 74 below). Both groups appeared to show a preference for active sentences and interpreted the passive sentences as such (as shown by higher looks to distractor than target, for passives). The NSs and the learners looked to target more than distractor for actives, but the learners proportionally less so (see figure 75).

The model found a significant main effect of voice after the verb. This confirms the difference observed in the figures (estimate = 2.48e-01, SE = 1.10e-01, t = 2.25, p = .03).

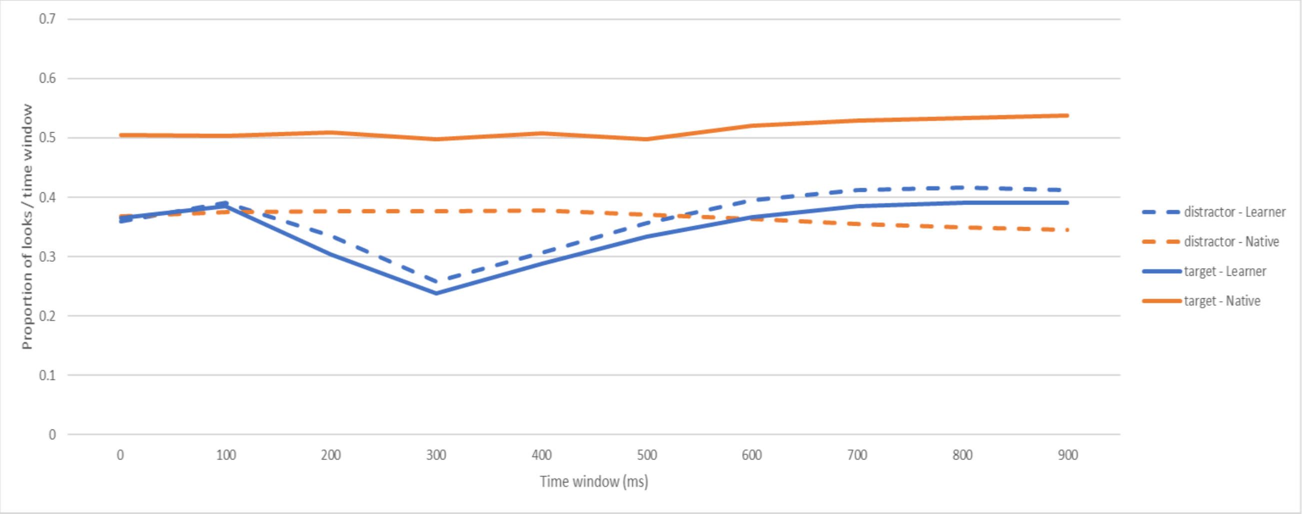


Figure 74. Eye-tracking: Native speakers and learners - passive items

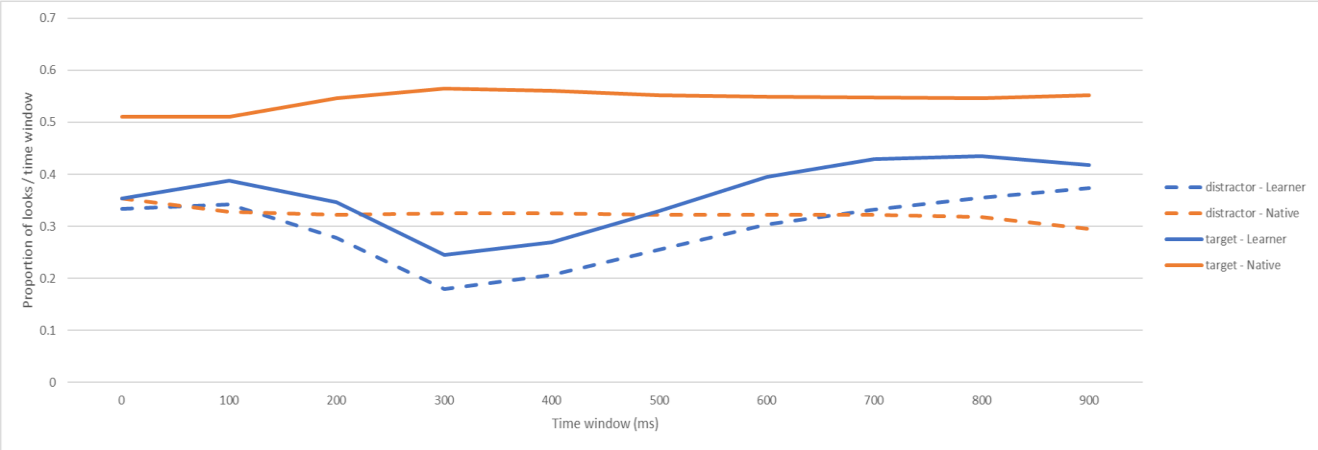


Figure 75. Eye-tracking: Native speakers and learners - active items

4.2.4 Visual world eye-tracking – summary of key findings

The effect of training on trained constructions

- Prior to training, all learners looked more to the distractor image than to the target for the passive items. The reverse was true for actives.
- At immediate post-test, the noun focused group seemed to process passives and actives more accurately than prior to training, and more accurately than the other groups (more looks to target image than distractor).
- At immediate post-test, the cue focused group also looked more to target for passives than at pre-test.
- At delayed post-test, for passives, the cue focused group looked more to target sooner than at immediate post-test, and sooner than the other groups (during the verb). This was followed by more looks to distractor in the remaining time windows.
- At delayed post-test, for passives, the noun focused group looked more to distractor than target, after the verb.

- At delayed post-test, for passives, only the test-only group looked to target more than distractor, after the verb.

The effects of post-verbal animacy and training

- At pre-test, only inanimate-animate passive and active sentences were processed significantly differently – looks to target were proportionally less than the other animacy combinations.
- Training did not appear to have a clear effect on the interpretation of the different animacy combinations.
- The intervention groups showed some significant differences for some animacy combinations at the post-tests, but these were not consistent between test times.
- The test-only group did not show any significant changes between test times.

The effects of pre-verbal animacy and training

- At pre-test, only animate-inanimate active sentences had more looks to target than distractor prior to the verb.
- At immediate post-test, the cue focused group looked more to target than distractor for inanimate-animate passive items than the noun focused group.
- At delayed post-test there were no significant changes between pre-test, or between groups.

The effect of training on untrained constructions

- At the post-tests, despite observable difference shown in the figures there were no significant differences between tests or between the groups.
- At delayed post-test, the figures showed that the cue focused group looked more to target than distractor after the verb ending and throughout the remaining time windows (compared to the other groups and test times).

Native speakers

- There was no significant difference between passive and active sentences for NSs, despite observable differences in the figures suggesting actives were processed more accurately (proportion of looks to target).
- Animacy did not have a significant effect on looks to target versus distractor, despite the figures suggesting a preference for the active interpretation of the input.
- Despite clear visual differences shown in the figures between the NSs and learners, there were no significant differences between the NSs and learners.

4.3 Passive and active morphosyntactic cue knowledge in learners: Production and grammaticality judgements

This section outlines the results of the offline tests used to investigate the second research question:

RQ2) To what extent do L1 Chinese learners of L2 English show knowledge of cues in a) offline grammaticality judgements and b) production of the passive voice in English?

Is production affected by:

> *a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?*
>
> *b) Trained and untrained constructions (present simple and present perfect)?*

Are grammaticality judgements affected by:

> *a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?*
>
> *b) Trained and untrained constructions (present simple and present perfect)?*
>
> *c) First and second noun animacy?*

The following analysis is presented for each of the above sections:

Analysis of pre-, immediate post- and delayed post-test scores and effect sizes for:

a) passive items compared to actives items (trained and untrained items).

b) present simple passive and present progressive active items compared to present perfect passive and present perfect progressive active items (trained and untrained items).
c) first and second noun animacy (for the JT)

Within-subjects effect sizes were considered large if exceeding 1.40, medium if 1.00 to 1.40, and small .60 to 1.0. Between-subjects effect sizes were considered large if exceeding 1.00, medium if .70 to 1.00, and small .40 to .70. This is in line with Plonsky & Oswald's (2014) field specific guide for interpreting effect sizes from their meta-analysis of effect sizes in L2 research.

R^2 was calculated for each model to determine the amount of variance explained by the fixed and random effects, both marginal R^2 and conditional R^2 are reported (marginal R^2 = fixed effects, and conditional R^2 = fixed + random effects) (Nakagawa & Schielzeth, 2013). R^2 was computed using the *MuMIn* package (Bartoń, 2018) in R (R Core Team, 2018). R^2 values range from 0–1. Values around .18, .32, and .51 are interpreted as small, medium, and large, respectively, in terms of the described variance they represent (Plonsky & Ghanbar, 2018).

Following Kasprowicz et al. (2019) and McManus and Marsden (2017) between-group effect sizes corrected for differences at pre-test were calculated. These give an indication of the change over time adjusting for baseline differences. Since 95% confidence intervals cannot be calculated for these effect sizes, the unadjusted effect sizes are presented in the main body of the thesis (Kasprowicz et al., 2019). The adjusted effect sizes are referred to if after correction the effect size is markedly different to the uncorrected original effect size. The corrected effect sizes can be found in appendices 22-26. The calculation for the corrected effect sizes was as follows:

effect size at immediate / delayed post-test − effect size at pre-test = corrected effect size

For sections in which the mixed-effects models or effect sizes did not indicate significant findings, tables containing means and effect sizes are in the appendices (22 – 26). The relevant appendices and tables are indicated in each section if they are not included in the main body of the chapter.

4.3.1 Oral production test

The oral production test elicited the production of four types of sentence: present progressive and present perfect progressive (active sentences e.g. 36a (trained) and 36c (untrained)) and present simple passive and present perfect passive (passive sentences e.g. 36b (trained) and 36d (untrained)).

36a. The dog is carrying the boy

36b. The dog is carried by the boy

36c. The man has been helping the woman

36d. The man has been helped by the woman

Scoring assessed the use of *be* + the correct verb inflection, and the use of *by*. These elements were focused on because they were the cues highlighted by the cue focused training. For example:

The picture has been drawn by the boy would score 2 (1 point for correct verb inflection (*be* + past participle) and 1 point for correct use of *by*)

The picture has been drawing by the boy would score 1 point for correct use of *by* but lose a point for incorrect use of the past participle; the correct form of ‘be’ was not scored).

A present progressive sentence would score 1 point for the absence of *by*, e.g.

The girl is paying the man would score 2 points (1 for the correct verb and 1 for not using *by*)

Prior to analysis, outliers were removed on the basis of 2.5 standard deviations above and below the mean for the group which the participant was in at each test (pre-test, immediate, and delayed post-test). These were removed from the data at all test times. This resulted in data loss of four participants from the cue focused group, one from the noun focused group, and four from the test only group. Mean percentage scores and corresponding Cohen’s *d* effect sizes are presented where appropriate.

A linear mixed-effects model, using the lmer function in R, was used to analyse the oral production data (R Core Team). This type of analysis was chosen because mixed-effects models deal with missing data and do not rely on assumed sphericity or a gaussian distribution (Howell, 2002). The oral production test data for the three groups were not normally distributed. This was determined using the Shapiro-Wilk test because it is most appropriate for smaller sample sizes (less than 100 cases) (p = .00). Mixed-effects models also produce more precise predictions than a using a non-parametric alternative to a traditional ANOVA (e.g. Friedman's) (Larsen-Hall, 2015). Random effects such as participants and test items are also taken into account.

The dependent variable was test score. Fixed effects were time, group, tense (trained / untrained), animacy, voice, and reversibility. The maximal random-effects structure supported by the data was used (Barr et al., 2013). Item and test version were random effects, with time as a random slope of version, and voice a random slope of item. Further specification of the random-effects structure led to a failure to converge. The predictor variables were sum coded so that the intercept moved to the middle of the categories. This is the conceptual version of "centering" for categorical predictors (following Winter, 2019). The resulting model was a follows:

*Model: lmer(score ~ group * time * voice * animacy * tense * reverse + (1 | subject) + (voice |item) + (time | version), data=new)*

The significance of each main effect was obtained using likelihood ratio tests comparing the full model (including the effect in question) with the model excluding the effect in question. *P* values for interactions were determined using the Satterthwaite approximation in the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team). *P* values were taken as significant when a likelihood ratio test had also shown that the fixed effect in question was significant (following Brown et al., 2014). The results of the likelihood ratio tests and corresponding *p* values are reported in the following sections (i.e. Chi-squared [$\chi 2$] and *p* from the model output).

R^2 indicated that the fixed effects described a relatively small proportion of variance (marginal $R^2 = .08$) and the fixed and random effects in combination described some of the variance (conditional $R^2 = .36$) (Nakagawa & Schielzeth, 2013). This is almost in line with previous research in language learning in which one fifth to half of the variance is explained by the model used (as found by Plonsky & Ghanbar, 2018).

Animacy and agent-patient reversibility were manipulated in the oral production test as they were in all tests. For this reason, they were included as predictor variables in the mixed-effects model. As expected, neither were significant main effects and neither significantly interacted with the other variables. Since, animacy and role reversibility were not expected to affect production and this was confirmed by the mixed-effects model, further results are not presented in this section (model results can be found in appendix 22).

4.3.1.1 Oral production test – production of passive items

This section presents the scores for passive items e.g.

37. The biscuit is baked by the woman

38. The cat is fed by the girl

Gains in mean percentage score for passive items were similar for the three groups across time. By delayed post-test, all the learners had improved their production accuracy (see table 6). Small-medium effect sizes between groups also indicated this (see table 8).

The mixed-effects model found a significant effect of voice on score ($\chi^2(85) = 17.39$, $p = .00$) with passives being scored .14 lower compared to the mean compared to actives (intercept = 1.64) ± .06 (standard errors). The difference between groups over time was not found to be significant by the model (see appendix 22 for interactions between group and time).

Table 6. Oral production test mean % score for passives across time (SDs) (*k*=16)

Group (*n*)	Pre	Immediate post	Delayed post
Cue (21)	70 (19)	83 (7)	85 (5)
Noun (24)	68 (15)	76 (13)	79 (13)
Test-only (19)	69 (18)	81 (11)	84 (7)

Effect sizes for the gains between pre-test and immediate post-test were small for all three groups (see table 8). Between pre-test and delayed post-test, the effect sizes for the cue focused and test-only group were very similar, suggesting no effect of instruction on passive scores (as shown by both within-subject and between-group effect sizes – see tables 7 and 8).

Table 7. Oral production test within-subject effect size (Cohen's* d*) between test times for passives

Group (*n*)	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (21)	.91 [.27, 1.54]	.33 [-.24, .90 [a]]	1.08 [.43, 1.73]
Noun (24)	.57 [-.01, 1.15 [a]]	.23 [-.34, .80 [a]]	.78 [.20, 1.37]
Test-only (19)	.80 [.14, 1.47]	.33 [-.31, .97 [a]]	1.10 [.42, 1.78]

[a] 95% confidence intervals pass through zero

Table 8. Oral production test between-groups effect sizes (Cohen's d) for passives

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.12 [-.47, .70 [a]]	.66 [.06, 1.26	.59 [.00, 1.19]
Cue vs. test-only	.05 [-.57, .67 [a]]	.22 [-.40, .84 [a]]	.17 [-.46, .79 [a]]
Noun vs. test-only	-.06 [-.66, .54 [a]]	-.41 [-1.02, .20 [a]]	-.46 [-1.07, .15 [a]]

[a] 95% confidence intervals pass through zero

4.3.1.2 Oral production test – production of active items

This section presents the analysis of the scores for active items only e.g.

39. The boy is asking the girl

40. The boy has been drawing a picture

Active items were produced more accurately than passive items at all three test times by all three groups (see tables 9 and 11). However, the model showed no significant effect of training on active items (see appendix 22 for model results).

Effect sizes (see table 10) showed that time had the biggest effect on the cue focused group between pre-test and delayed post-test. However, between-group effect sizes showed that the cue focused and noun focused groups did not differ significantly between pre and delayed post-test (see table 11). The cue focused training appeared to have a greater effect on score than no training (i.e. the test-only group) (see table 11), but when the baseline differences were taken into account this effect reduced (*d* = .45) (see appendix 23 for all corrected effect sizes).

Table 9. Oral production test mean % score (**SDs*****) for actives across time (*****k*****=16)***

Group (*n*)	Pre	Immediate post	Delayed post
Cue (21)	88 (7)	92 (6)	94 (5)
Noun (24)	88 (6)	90 (5)	92 (6)
Test-only (19)	85 (16)	93 (4)	89 (9)

Table 10. Oral production test within-subjects effect size (Cohen's* d*) between test times for actives

Group (*n*)	Pre- vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue focused (21)	.61 [-.01, 1.23 [a]]	.36 [-.25, -.97 [a]]	.99 [.35, 1.62]
Noun focused (24)	.36 [-.21, .93 [a]]	.36 [-.21, .93 [a]]	.67 [.09. 1.25]
Test-only (19)	.69 [.03, 1.34]	-.57 [-1.22, .07 [a]]	.31 [-.33, .95 [a]]

[a] 95% confidence intervals pass through zero

Table 11. Oral production test between-groups effect sizes (Cohen's* d*) for actives

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.00 [-.59, .59 [a]]	.36 [-.23, .95 [a]]	.36 [-.23, .95 [a]]
Cue vs. test-only	.25 [-.38, .87 [a]]	-.19 [-.82, .43 [a]]	.70 [.06, 1.34]
Noun vs. test-only	.26 [-.34, .86 [a]]	-.65 [-1.27, -.04]	.40 [-.21, 1.01 [a]]

[a] 95% confidence intervals pass through zero

The above analysis was on the total mean percentage score. This score was a combination of producing the correct verb inflection (*be* + past participle) and the correct use of *by*. The next section presents mean percentage scores for these two cues separately because they were focused on separately in the training.

4.3.1.3 Oral production test – production of be + verb inflection

For production of the auxiliary *be* plus the correctly inflected verb (*-ing* vs *-ed*), both the cue focused and noun focused groups improved similarly between each pair of tests (pre-and immediate, immediate and delayed, and pre and delayed post-test) as shown by medium-sized effects with overlapping 95% confidence intervals (see table 13).

The test-only group had small effect sizes, two of which passed through zero and were therefore unreliable. Between-groups effect sizes adjusted for baseline difference showed that the two intervention groups performed better than the test-only group at delayed post-test, this was particularly the case for the cue focused group (cue focused vs. test-only *d* = .88, noun focused vs. test-only *d* = .53).

Table 12. Oral production test mean % correct verb inflection across test times (SDs) (k=32).

Group (*n*)	Pre	Immediate post	Delayed post
Cue (21)	76 (13)	85 (8)	88 (6)
Noun (24)	71 (12)	80 (9)	81 (9)
Test-only (19)	79 (9)	77 (10)	83 (10)

Table 13. Oral production test within-subject effect size (Cohen's* d*) between test times for correct verb inflection

Group (*n*)	Pre v. Immediate *d* [CIs]	Immediate v. Delayed *d* [CIs]	Pre v. Delayed *d* [CIs]
Cue (21)	.83 [.22, 1.44]	.42 [-.16, 1.01 [a]]	1.19 [.56, 1.81]
Noun (24)	.85 [.27, 1.43]	.11 [-.44, .67 [a]]	.94 [.36, 1.53]
Test-only (19)	-.21 [-.79, .37 [a]]	.60 [.01, 1.19]	.42 [-.16, 1.00 [a]]

[a] 95% confidence intervals pass through zero

Table 14: Oral production test between-groups effect sizes (Cohen's* d*) for correct verb inflection

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.40 [-.19, .99 [a]]	.58 [-.01, 1.18 [a]]	.90 [.29, 1.52]
Cue vs. test-only	-.27 [-.89, .36 [a]]	.89 [.24, 1.54]	.61 [-.02, 1.25 [a]]
Noun vs. test-only	-.74 [-1.36, -.12]	.32 [-.29, .92 [a]]	-.21 [-.82, .39 [a]]

[a] 95% confidence intervals pass through zero

4.3.1.4 Oral production test – production of by

The oral production test elicited full passives, so correct production of *by* was needed to make up half the correct total score (i.e. *be* + verb inflection = 1, plus *by* = 1). In terms of mean score, all participants scored highly on correct production of *by* at pre-test and both post-tests (see table 15).

Effect sizes showed very little difference in the production accuracy of *by* across test times or between groups (see table 16 and 17).

Table 15. Oral production test mean % score for production of* by *across time (SDs) (*k *= 32)

Group (*n*)	Pre	Immediate post	Delayed post
Cue (21)	89% (5)	89% (7)	90% (10)
Noun (24)	89% (5)	90% (7)	90% (7)
Test-only (19)	90% (5)	86% (10)	91% (7)

***Table 16. Oral production test within-subject effect size (Cohen's* d*) between test times for use of* by**

Group (*n*)	Pre v. Immediate *d* [CIs]	Immediate v. delayed *d* [CIs]	Pre v. delayed *d* [CIs]
Cue (21)	.00 [-.59, .59 [a]]	.12 [-.52, .75 [a]]	.13 [-.51, .73 [a]]
Noun (24)	.16 [-.43, .76 [a]]	.00 [-.57, .57 [a]]	.16 [-.43, .76 [a]]
Test-only (19)	-.51 [-1.12, .11 [a]]	.58 [-.04, 1.20 [a]]	.16 [-.44, .77 [a]]

[a] 95% confidence intervals pass through zero

***Table 17. Oral production test between-groups effect sizes (Cohen's* d*) for use of* by**

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.00 [-.59, .59 [a]]	-.14 [-.73, .44 [a]]	.00 [-.59, .59 [a]]
Cue vs. test-only	-.20 [-.82, .42 [a]]	.35 [-.27, .98 [a]]	-.11 [-.74, .51 [a]]
Noun vs. test-only	-.20 [-.80, .40 [a]]	.47 [-.14, 1.08 [a]]	-.14 [-.75, .46 [a]]

[a] 95% confidence intervals pass through zero

4.3.1.5 Oral production test – production of untrained constructions

The present simple passive and present progressive active sentences were used in the training for both intervention groups. Present perfect passive and progressive sentences were included in the tests to see if the learners generalised correct verb inflection and the use of *by* from trained to untrained constructions.

41. The picture has been drawn by the boy

42. The boy has been driving the car

The untrained items (present perfect), both passives and actives, were produced less accurately (combined score for use of verb inflection and *by*) by all three groups at all test times. However trained / untrained was not found to be a significant main effect $\chi 2(212) = 111.74$, $p = .10$.

Within-subjects Cohen's *d* effect sizes showed that time did not have an effect on score. This was also shown by between-groups effect sizes even when adjusted for baseline difference (see appendix 23 for means and effect sizes). Effect sizes for production score of passives and actives separately showed that the cue focused group and test-only improved their score for passives, but between-groups effect sizes were small (see appendix 23). The noun focused did not improve their score.

Untrained actives were produced more accurately by the cue focused group at delayed post-test. The noun focused and test-only groups did not improve their score for actives.

4.3.1.6 Oral production test – summary of key findings

Overall, all three groups improved their oral production to some extent over time. The cue focused group improved in mean score for passive items, but the effect sizes indicated that this was unlikely to be due to training since the test-only group saw a similar improvement. Furthermore, between-groups effect sizes, between the cue focused and test-only groups, were small for both passives and actives at immediate and delayed post-test. The noun focused group showed similar improvements in passive and active items over time (within-subjects) and compared to the test-only and cue focused groups (between-groups).

For morphosyntactic cue production (*be* + verb inflection and *by*), training seemed to have a positive effect on use of correct verb inflection but not on use of *by*. Between tests, the two intervention groups improved in their use of correct verb inflection as indicated by medium Cohen's *d* effect sizes. Effect sizes were greatest for the cue focused group between pre and delayed post-test. Between-groups effect sizes showed that the cue focused group performed better than the noun focused group at delayed post-test compared to pre-test. Both intervention groups outperformed the test-only group at delayed post-test.

Untrained items, the present perfect passive and present perfect progressive (active), were produced less accurately by all three groups at all test times. Although this effect was not found to be significant by the model.

4.3.2 Writing test

The writing test elicited the same four constructions as the oral production test but with different lexical items (trained items: present simple passive and present perfect passive; untrained items: present progressive, present perfect progressive).

As in the oral production test, scoring took account of the correct production of *be* plus the correct verb inflection and correct use of *by* e.g.,

The picture has been drawn by the boy would score 2 (1 point for the correct verb inflection and 1 point for correct use of *by*).

Outliers were removed before analysis; these were scores that were 2.5 standard deviations above or below the mean at each test (pre-test, immediate, and delayed post-test). These participants' data were removed from all test times. This resulted in the removal of three participants from the cue focussed group, one from the noun focused group, and one from the test-only group.

The writing test data were analysed using a linear mixed-effects model using the lmer function from the lme4 package in R (Bates et al., 2015; R Core Team). Fixed effects were group, time (pre-, immediate post-, and delayed post-test), tense, animacy, voice, and reversibility. The maximal random-effects structure supported by the data was used (Barr et al., 2013). Random effects were intercepts for subjects, items, and test version, as well as by-subject random slopes for the effect of time, by-item random slopes for the effect of voice, and by-version random slopes for the effect of time. Further specification of the random-effects structure led to a failure to converge. The predictor variables were sum coded so that the intercept moved to the middle of the categories.

*Model: lmer(score ~ time * group * tense * animacy * voice * reverse + (1 | subject) + (1 + voice | item) + (1+ time | version)*

As in the oral production test, the significance of each main effect was determined by likelihood ratio tests. *P* values for interactions were determined using the Satterthwaite approximation in the lmerTest package (Kuznetsova et al., 2017). *P* values were counted as significant where a likelihood ratio test showed that the fixed effect in question was significant (following Brown et al., 2014). The result of the likelihood ratio test and corresponding *p* value is reported in the following sections.

R^2 indicated that fixed effects described a relatively small proportion of variance (marginal R^2 = .12) and the fixed and random effects in combination described some of variance (conditional R^2 = .37) (Nakagawa & Schielzeth, 2013). This is in line with previous research in language learning in which one fifth to half of the variance is explained by the model used (as found by Plonsky & Ghanbar, 2018). The full model results are in appendix 24.

Animacy and agent-patient reversibility were manipulated in the writing test as they were in all tests. For this reason, they were included as predictor variables in the mixed-effects model. As expected, neither were significant main effects and neither significantly interacted with the other variables. Since animacy and role reversibility were not expected to affect production and this was confirmed by the mixed-effects model, further results for animacy and role reversibility are not presented in here (model results can be found in appendix 24).

4.3.2.1 Writing test - Production of passive items

This section analyses the scores on passive items (*be* + correct verb inflection plus *by*) e.g., *43. The biscuit is baked by the woman*

44. The cat is fed by the girl

For passives, as shown in table 18 below, the two intervention groups improved their mean scores between pre-test and immediate post-test. All three groups had improved by delayed post-test.

Cohen's *d* effect sizes showed that the noun focused group improved the most over time, compared to changes between tests for the cue focused group, and the test-only group (see table 19 for within-subjects effect sizes). Between-groups effect sizes indicated that the effect seen between test times for the noun focused group was not indicative of differences between groups (see table 20). Between-groups effect sizes were small at all test times, and even when adjusted for small baseline differences (see appendix 25).

The mixed-effects model found a significant effect of voice on written production ($\chi^2(85)=17.39$, p = .00) with passives being scored .12 lower compared to the mean of all items (i.e. including active items) (intercept = 1.69) ± .04 (standard errors). There were no statistically significant effects or interactions of test time or group.

Table 18. Writing test mean % score for passives across time (SDs) (k=18)

Group (*n*)	Pre	Immediate post	Delayed post
Cue (22)	72 (13)	80 (12)	83 (13)
Noun (24)	67 (13)	77 (13)	79 (11)
Test-only (22)	73 (11)	70 (13)	80 (12)

Table 19. Writing test within-subject effect size (Cohen's* d*) between test times for passives

Group (*n*)	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (22)	.64 [.07, 1.21]	.24 [-.32, .80 [a]]	.85 [.27, 1.42]
Noun (24)	.80 [.22, 1.38]	.17 [-.39, .72 [a]]	1.00 [.41, 1.58]
Test-only (22)	-.27 [-.85, .31 [a]]	.80 [.20, 1.4]	.61 [.02, 1.20]

[a] 95% confidence intervals pass through zero

Table 20. Writing test between-groups effect sizes (Cohen's* d*) for passives

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun (22)	.38 [-.20, .97 [a]]	.24 [-.34, .82 [a]]	.33 [-.25, .92 [a]]
Cue vs. test-only (24)	-.08 [-.67, .51 [a]]	.80 [.19, 1.41]	.24 [-.34, .82 [a]]
Noun vs. test-only (22)	-.50 [-1.08, .09 [a]]	.54 [-.05, 1.13 [a]]	-.09 [-.67, .49 [a]]

[a] 95% confidence intervals pass through zero

4.3.2.2 Writing test - Production of active items

This section analyses the production of active items (*be* + correct verb inflection, plus *by*) e.g.

45. The boy is asking the girl

46. The boy has been drawing a picture

All three groups produced active sentences at a very high level of accuracy at pre-test, and therefore did not make large gains at post-tests (see table 21). All three groups performed similarly at all test times.

Within-subjects effect sizes showed that there was very little effect of time on score for active items (see table 22). Between-groups effect sizes showed very little, or no, difference at immediate or delayed post-test (see tables 23).

Table 21. Writing test mean % score for active items across time (*SDs*) (*k *=18)

Group (*n*)	Pre	Immediate post	Delayed post
Cue (22)	91 (6)	93 (5)	92 (8)
Noun (24)	91 (6)	93 (4)	93 (5)
Test-only (22)	91 (5)	93 (4)	93 (4)

Table 22. Writing test within-subject effect sizes (Cohen's* d*) between test times for actives

Group (*n*)	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (22)	.36 [-.20, .92 [a]]	-.15 [-.71, .41 [a]]	.14 [-.41, .70 [a]]
Noun (24)	.39 [-.17, .95 [a]]	.00 [-.57, .57[a]]	.36 [-.20, .92 [a]]
Test-only (22)	.44 [-.14, 1.03 [a]]	.00 [-.59, .59[a]]	.44 [-.14, 1.03 [a]]

[a] 95% confidence intervals pass through zero

Table 23. Writing test between-groups effect sizes (Cohen's d) for actives

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.00 [-.58, .58 [a]]	.00 [-.58, .58 [a]]	-.15 [-.73, .43 [a]]
Cue vs. test-only	.00 [-.59, .59 [a]]	.00 [-.59, .59 [a]]	-.16 [-.75, .43 [a]]
Noun vs. test-only	.00 [-.58, .58 [a]]	.00 [-.58, .58 [a]]	.00 [-.58, .58 [a]]

[a] 95% confidence intervals pass through zero

4.3.2.3 Writing test- Production of be *+ verb inflection*

In terms of mean percentage score, the cue focused group performed best at correctly producing the auxiliary *be* + correct verb inflection. The noun focused and test-only group performed similarly (see table 24).

The cue focused group showed a large effect size between pre-test and delayed post-test (see table 25). Between-groups effect sizes also showed that the cue focused group improved more in their use of *be* + verb inflection compared to the other groups at delayed post-test (see table 26). Groups differences between the noun focused and test-only group were small as indicated by small, unreliable between-group effect sizes (see table 26)

Table 24. Writing test mean % correct verb inflection across test times (**SDs*****) (*****k*****=32).***

Group (*n*)	Pre	Immediate post	Delayed post-
Cue (22)	71 (12)	83 (9)	86 (6)
Noun (24)	70 (10)	79 (9)	79 (9)
Test-only (22)	71 (11)	79 (9)	81 (9)

Table 25. Writing test within-subject effect size (Cohen's* d*) between test times for correct verb inflection

Group (*n*)	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (22)	1.13 [.53, 1.73]	.39 [-.17, .95 [a]]	1.58 [.95, 2.22]
Noun (24)	.95 [.35, 1.54]	.00 [-.57, .57 [a]]	.95 [.35, 1.54]
Test-only (22)	.80 [.18, 1.41]	.22 [-.37, .82 [a]]	1.00 [.37, 1.62]

[a] 95% confidence intervals pass through zero

Table 26. Writing test between-groups effect sizes (Cohen's d) across test times for correct verb inflection

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.09 [-.49, .67 [a]]	.44 [-.14, 1.0 3 [a]]	.91 [.30, 1.51]
Cue vs. test-only	.00 [-.59, .59 [a]]	.44 [-.15, 1.04 [a]]	.65 [.05, 1.26]
Noun vs. test-only	-.10 [-.67, .48 [a]]	.00 [-.58, .58 [a]]	-.22 [-.80, .36 [a]]

[a] 95% confidence intervals pass through zero

4.3.2.4 Writing test- Production of by

All three groups' mean percentage scores were high for production of *by* at pre-test so there was not much scope for improvement after the interventions (see table 27).

Within-subjects effect sizes showed that the cue focused group improved their score the most between pre-test and immediate post-test, and pre and delayed post-test compared to the other groups (as shown in table 28) with a medium effect size. The noun focused group also showed some improvement with a small-medium effect size. The test-only group showed no improvement.

Between-groups effect sizes adjusted for baseline differences showed that the intervention groups both improved their score to a greater extent than the test-only group, but the effect of either training type was not large when compared to the test-only group (cue focused vs. test-only *d* = .76, noun focused vs. test-only *d* = .54). Despite the cue focused group's within-subject effect sizes suggesting the most improvement over time, the between-groups effect sizes showed that the cue focused and noun focused training appeared to both result in similar improvements in using *by* (see table 29 and appendix 25).

Table 27. Writing test mean % correct use of* by *across test times (SDs) (*k*=32).

Group (*n*)	Pre	Immediate post	Delayed post
Cue (22)	89 (6)	93 (4)	94 (4)
Noun (24)	88 (9)	92 (4)	93 (5)
Test-only (22)	92 (5)	95 (4)	93 (5)

***Table 28. Writing test within-subject effect size (Cohen's* d*) between test times for correct use of* by**

Group (*n*)	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (22)	.78 [.18, 1.38]	.25 [-.33, .83 [a]]	.98 [.37, 1.59]
Noun (24)	.57 [.01, 1.14]	.22 [-.34, .78 [a]]	.69 [.12, 1.26]
Test-only (22)	.66 [.01, 1.32]	-.44 [-1.09, .20 [a]]	.20 [-.44, .84 [a]]

[a] 95% confidence intervals pass through zero

Table 29. Writing test between-group effect size (Cohen's d) between test times for correct use of by

Group (*n*)	Pre *d* [CIs]	Immediate d [CIs]	Delayed *d* [CIs]
Cue vs. noun	.13 [-.45, .71[a]]	.25 [-.33, .83[a]]	.22 [-.36, .80[a]]
Cue vs. test-only	-.54 [-1.15, .06[a]]	-.05 [-1.10, .10[a]]	.22 [-.37, .81[a]]
Noun vs. test-only	-.54 [-1.13, .05[a]]	-.75 [-1.35, -.15]	.00 [-.58, .58[a]]

[a] 95% confidence intervals pass through zero

4.3.2.5 Writing test- Production of untrained constructions

As in the oral production test, the present perfect passive and progressive sentences were included in the writing test to see if learners could generalise use of correct verb inflection, plus the use of *by*, from trained constructions to untrained constructions.

47. The picture has been drawn by the boy

48. The boy has been driving the car

Trained / untrained (tense) was not found to be a significant main effect by the model ($\chi^2(94) = 111.74$, $p = .10$).

There was very little difference in mean overall score for correct verb inflection plus production of *by* between trained (present simple) and untrained (present perfect) items at all test times for all groups (see appendix 25 table 5). This was also shown by similar effect sizes, with overlapping CIs, for trained and untrained items between test times for each group (see appendix 25 table 6).

Between-groups effect sizes showed that the cue focused group improved more at delayed post-test than the other two groups for both trained and untrained items (see appendix 25 table 7). However, since the effect size comparing the cue focused and test-only group for untrained items at delayed post-test passed through zero (d = .41, CIs = -.19, 1.01) the difference in groups may have been due to the noun focused group performing worse at delayed post-test, rather than because the cue focused training had a positive effect on score.

When analysed by voice (passive and active separately), effect sizes showed that all three groups improved for untrained passive items to a similar extent (appendix 25 table 9). Scores for active sentences did not improve (appendix 25 table 13).

4.3.2.6 Writing test– summary of key findings

All three groups improved in their overall production (verb inflection + *by*) of passives between pre-test and delayed post-test. Actives were produced almost at ceiling level at pre-test. There was little difference between groups at each test time, as shown by Cohen's d effect sizes.

Group differences were more marked when scores for each morphosyntactic cue were analysed separately (i.e. *be* + verb inflection and, separately, *by*). The cue focused group made the most improvement in written production of the correct verb inflection and use of *by* over time (between pre and immediate post-test, and pre and delayed post-test) according to the effect sizes. Between-groups effect sizes also showed that the cue focused training had a greater effect on production of the verb inflection than the noun focused training by delayed post-test, with a small-medium effect size and confidence intervals which did not overlap. However, the within-subjects effect sizes at delayed post-test all had overlapping confidence intervals for all three groups suggesting this difference may not have been reliable.

All groups improved production accuracy of untrained and trained items over time. The mixed-effects model did not find a significant effect of trained / untrained sentences on production.

4.3.3 Written untimed acceptability judgement test (JT)

The JT assessed the learners' ability to recognise and correct three types of error. These were missing auxiliary verb (k = 16) (49a), base form of the main verb (k = 16) (49b) and gerund form of the main verb (k = 16) (49c) (* indicates the error).

*49a. The child * being helped by the teacher.*

49b. The dog is being walk by the boy.*

49c. The person is being following by the police.*

The JT consisted of three steps for each item. For each item the participants were required to: (i) assess its grammatical acceptability, (ii) circle the incorrect word, and (iii) write a corrected sentence. In the following analyses, judgements were scored dichotomously (1 = correct, 0 = incorrect). Error identification was also scored in the same way. Corrections were also scored in the same way, but mistakes were coded to determine the nature of the incorrect corrections made. The full error identification and correction analysis (steps ii and iii) is included in appendix 27 (tables 21-50), the main findings are presented in section 4.3.3.4 and 4.3.3.5.

In the following sections data from step i (the grammaticality judgements) are analysed and described for trained passive and active items (i.e., present simple passive and present progressive respectively), followed by animacy effects on judgements, and finally untrained items (present perfect passive and present perfect progressive). After presenting the judgement data, then step ii) (the error identification scores) and step iii) (the corrections made), are briefly described.

4.3.3.1 Grammaticality judgements of trained constructions

Prior to analysis, outliers were removed. This was done on the basis of participants scoring 2.5 SDs above or below the mean for their group at each test (pre-test, immediate post, and delayed post-test). Outliers were calculated for acceptability judgements and identification of errors. Outlying participants were removed from the full dataset across all tests. The resulting number of participants in each group is presented in each table.

The judgement accuracy data were entered into a logistic mixed-effects model. The model was a logistic mixed-effects model (glmer function in R) because this is more appropriate with binomial data than a linear mixed-effects model using the lmer function in R (following de Wilde, Brysbaert, & Eyckmans, 2020; Lee & Révész, 2020; Kasprowicz et al., 2019; Morgan-Short et al., 2018; Winter, 2019). The acceptability judgements were scored as follows: 1 = correct judgement or 0 = incorrect judgement.

Fixed effects were group, animacy, voice (passive / active), tense (trained / untrained) and grammaticality. The maximal random-effects structure was used (Barr et al., 2013). Random effects were intercepts for subjects, and item, as well as by-item random slopes for the effect of voice. Further specification of the random-effects structure led to a failure to converge. The predictor variables were sum coded so that the intercept moved to the middle of the categories.

*Model <- glmer(score ~ group * voice * tense * animacy * grammatical + (1 | subject) + (1 + voice | item), data=new, family=binomial, control = glmer, Control(optimizer = "bobyqa"))*

The model would not converge when test time (pre-, immediate post-, and delayed post-test) was included as a predictor variable (despite step-wise removal of random slopes, followed by intercepts) therefore, the above model was run for each test time separately and the effects of time are discussed using figures and Cohen's *d* effect sizes.

P values for interactions were determined using the Satterthwaite approximation in the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team). *P* values were counted as significant where a likelihood ratio test showed that the fixed effect in question was significant (following Brown et al., 2014). The result of the likelihood ratio tests, and corresponding *p* values are reported in the following sections (i.e., Chi-squared and *p* from the model output). Since a logit model was used (for binary data) estimates are presented as logit coefficients.

Theoretical R^2 (used for binary data) indicated that fixed effects described a small proportion of the variance (pre-test marginal $R^2 = .04$, conditional $R^2 = .40$; immediate post-test marginal $R^2 = .07$, conditional $R^2 = .47$; delayed post-test marginal $R^2 = .07$, conditional $R^2 = .51$).

Mixed-effects models were used to analyse the majority of the JT data. If other inferential tests were used, these were parametric because the data for the three groups were normally distributed ($p>.05$). This was determined using the Shapiro-Wilk test because it is most appropriate for smaller sample sizes. Q-Q plots for each group also confirmed normality. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated ($\chi^2(2) = 2.18$, $p = .34$).

4.3.3.1 a) Acceptability judgements of passive sentences only

At pre-test, the log odds for judging passive items correctly were .61 higher than active items. This was confirmed by a significant effect of voice at pre-test ($\chi^2(48) = 79.42$, $p = .00$). This was also found at immediate post-test ($\chi^2(48) = 71.05$, $p = .02$) and delayed post-test ($\chi^2(48) = 156.79$, $p = .00$). This section presents the results for passives first, followed by actives.

All groups made gains in mean percentage score between pre-test and delayed post-test for passive sentences (as shown in table 30).

The mixed-effects model found a significant main effect of group ($\chi^2(64) = 96.21$, $p = .01$). Compared to the noun focused and test-only groups, the cue focused group made the most gains in score between pre and immediate, and pre and delayed post-test. This was shown by the medium within-subjects Cohen's *d* effect sizes (as shown in table 31).

Between-groups effect sizes between the cue focused and other groups were small-medium at immediate and delayed post-test, even when adjusted for baseline differences (see table 32 and appendix 27). This suggested that the difference seen between groups for within-subjects effect sizes may not have been reliable.

Table 30. JT - mean % accuracy score for passives across time (*SDs*) (*k=24*)

Group (*n*)	Pre-test	Immediate post	Delayed post
Cue focused (21)	79 (12)	87 (11)	90 (7)
Noun focused (22)	79 (16)	79 (10)	86 (11)
Test-only (22)	77 (12)	82 (9)	83 (14)

Table 31. JT - within-subject effect size (Cohen's d) between test times for judgements for passives.

Group (*n*)	Pre vs. Immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (21)	.70 [.07, 1.32]	.33 [-.28, .93 [a]]	1.12 [.47,1.77]
Noun (22)	.00 [-.59, .59 [a]]	.67 [.06, 1.27]	.51 [-.09, 1.11 [a]]
Test-only (22)	.47 [-.13, 1.07 [a]]	.09 [-.51, .68 [a]]	.46 [-.14,1.06 [a]]

[a] 95% confidence intervals pass through zero

Table 32. JT – between- groups effect sizes (Cohen's* d*) for passives

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.00 [-.60, .60 [a]]	.76 [.14, 1.38]	.43 [-.17, 1.04 [a]]
Cue vs. test-only	.17 [-.43, .77 [a]]	.50 [.11, 1.01]	.63 [-.02, 1.24 [a]]
Noun vs. test-only	.14 [-.45, .73 [a]]	-.32 [-.91, .28 [a]]	.24 [-.35, .83 [a]]

[a] 95% confidence intervals pass through zero

4.3.3.1 b) Acceptability judgements of active sentences

As previously mentioned, the mixed-effects model found a significant effect of voice for all test times, suggesting that active sentences were judged significantly less accurately at all test times. Effects sizes showed that the improvements for active sentences were not as great as for passive sentences for all groups.

Between tests, within-subject effect sizes (see table 34) showed that the cue focused and noun focused groups improved their score to a similar extent between pre-test and delayed post-test, with small-medium effect sizes and 95% confidence intervals that did not cross zero. The test-only group showed no, or little, change between the tests (evidenced either by mean score and effect size (*d*)). Between-groups effect sizes showed little difference between the groups at each test time (see table 35).

Table 33. JT - correct judgements - mean % score for actives across time (**SDs*****) (*****k*****=24)***

Group (*n*)	Pre	Immediate post	Delayed post
Cue (21)	72 (16)	82 (14)	82 (12)
Noun (22)	76 (15)	76 (16)	86 (11)
Test-only (22)	74 (18)	80 (14)	80 (15)

Table 34. JT within-subject effect size (Cohen's **d*****) between test times for judgements for actives.***

Group (*n*)	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (21)	.67 [.04, 1.29]	.00 [-.60, .60 [a]]	.71 [.08, 1.33]
Noun (22)	.00 [-.59, .59 [a]]	.72 [.12, 1.34]	.76 [.15, 1.37]
Test-only (22)	.37 [-.22, .97 [a]]	0 [-.59, .59 [a]]	.36 [-.23, .96 [a]]

[a] 95% confidence intervals pass through zero

Table 35: JT between-groups effect sizes (Cohen's* d*) for actives

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	-.26 [-.86, .34 [a]]	.40 [-.21, 1.00 [a]]	-.35 [-.95, .25 [a]]
Cue vs. test-only	-.12 [-.72, .48 [a]]	.14 [-.46, .74 [a]]	.15 [-.45, .75 [a]]
Noun vs. test-only	.12 [-.47, .71 [a]]	-.27 [-.86, .33 [a]]	.46 [-.14, 1.05 [a]]

[a] 95% confidence intervals pass through zero

4.3.3.1 c) Acceptability judgements of grammatical and ungrammatical items

Effect of grammaticality and voice on judgements

The mixed-effects model found a significant main effects of grammaticality at all test times (pre-test ($\chi^2(48) = 91.99$, $p = .00$); immediate post-test ($\chi^2(48) = 85.11$, $p = .00$); delayed post-test ($\chi^2(48) = 159.75$, $p = .00$). At pre-test, ungrammatical passives were scored .35 lower than grammatical actives (logit coefficient: -.35, SE = 2.66, $z = -.13$). The model results showed that passive ungrammatical items were harder to judge than the active grammatical and ungrammatical items, and harder to judge than passive grammatical items.

Group differences and the effects of grammaticality and voice on judgements

Ungrammatical passive items were judged least accurately (at all test times for noun focused group, delayed post-test for cue focused group, and immediate and delayed post-test for test-only group). Although gains were made for grammatical and ungrammatical items by all groups, the magnitude of the effect of time on score for grammatical and ungrammatical items varied between groups, as evidenced by Cohen's *d* effect sizes (see table 37).

The cue focused group was the only group to show much improvement for ungrammatical passive items. The cue focused group improved more than the other two groups in their judgements of ungrammatical passives at immediate post-test and maintained their improvement till delayed post-test. This was shown by medium between-groups effect sizes at immediate post-test (table 38) and bigger within-subjects effect sizes than the other groups between test times (table 37). This suggested that they were better able to recognise grammatical violations in the passive items than the other groups. The noun focused group improved the least for ungrammatical items.

A significant interaction for ungrammatical items and the noun focused group was found by the model at immediate post-test (logit coefficient = -.73, SE = .36, z = -2.05). However, as shown by between-group effect sizes (see table 38), differences between the other groups were small at immediate and delayed post-test for all item types (except ungrammatical passives as mentioned above).

Table 36. JT acceptability judgements - mean % score for grammatical and ungrammatical passives (k=24) and actives (k=24)* (SDs*)

Group (*n*)	Type	Grammatical	Pre	Immediate post	Delayed post
Cue (25)	Passives	Grammatical	82 (14)	89 (11)	92 (8)
		Ungrammatical	75 (16)	84 (13)	87 (11)
	Actives				
		Grammatical	66 (21)	80 (21)	79 (21)
		Ungrammatical	80 (18)	84 (20)	88 (17)
Noun (25)	Passives	Grammatical	85 (13)	90 (10)	92 (12)
		Ungrammatical	76 (22)	71 (20)	80 (17)
	Actives				
		Grammatical	76 (17)	79 (22)	88 (11)
		Ungrammatical	79 (22)	71 (26)	84 (17)
Test-only (23)	Passives	Grammatical	78 (21)	88 (11)	86 (22)
		Ungrammatical	74 (16)	75 (14)	79 (15)
	Actives				
		Grammatical	69 (24)	78 (16)	76 (25)
		Ungrammatical	81 (20)	83 (17)	85 (17)

Table 37. JT within-subject effect size (Cohen's* d*) between test times for grammatical and ungrammatical passives (k=24) and actives (k=24)

Group (*n*)	Type	Grammaticality	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (25)	Passives	Grammatical	.56 [-.00, 1.12 [a]]	.28 [-.28, .84 [a]]	.88 [.30, 1.46]
		Ungrammatical	.62 [.05, 1.18]	.25 [-.31, .81 [a]]	.87 [.29, 1.45]
	Actives	Grammatical	.67 [.10, 1.24]	-.05 [-.60, .51 [a]]	.62 [.05, 1.19]
		Ungrammatical	.21 [-.35, .77 [a]]	.22 [-.34, .77 [a]]	.46 [-.10, 1.02 [a]]
Noun (25)	Passives	Grammatical	.43 [-.13, .99 [a]]	.18 [-.37, .74 [a]]	.56 [-.01, 1.12[a]]
		Ungrammatical	-.24 [-.79, .32 [a]]	.48 [-.08, 1.05 [a]]	.20 [-.35, .76 [a]]
	Actives	Grammatical	.15 [-.40, .71 [a]]	.52 [-.05, 1.08 [a]]	.84 [.26, 1.42]
		Ungrammatical	-.33 [-.89, .23 [a]]	.33 [-.23, .89 [a]]	.25 [-.30, .81 [a]]
Test-only (23)	Passives	Grammatical	.60 [.01, 1.19]	-.12 [-.69, .46 [a]]	.37 [-.21, .95 [a]]
		Ungrammatical	.07 [-.51, .64 [a]]	-.28 [-.86, .31 [a]]	.32 [-.26, .90]
	Actives	Grammatical	.44 [-.14, 1.03 [a]]	-.10 [-.67, .48 [a]]	.29 [-.30, .87 [a]]
		Ungrammatical	.11 [-.47, .69 [a]]	.12 [-.46, .70 [a]]	.22 [-.36, .80 [a]]

[a] 95% confidence intervals pass through zero

Table 38. JT between-groups effect sizes (Cohen's d) for grammatical and ungrammatical passives (k=24) and actives (k=24)

Group (*n*)	Type	Grammaticality	Pre *d* [CIs]	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	Passives	Grammatical	-.22 [-.78, .33 [a]]	-.10 [-.65, .46 [a]]	.00 [-.55, .55 [a]]
		Ungrammatical	-.05 [-.61, .50 [a]]	.77 [.20, 1.35]	.49 [-.07, 1.05 [a]]
	Actives	Grammatical	-.52 [-1.09, .04 [a]]	.05 [-.51, .60 [a]]	-.54 [-1.10, .03 [a]]
		Ungrammatical	.05 [-.50, .60 [a]]	.56 [-.00, 1.13 [a]]	.24 [-.32, .79 [a]]
Cue vs. test-only	Passives	Grammatical	.23 [-.34, .79 [a]]	.09 [-.48, .66 [a]]	.37 [-.20, .94 [a]]
		Ungrammatical	.06 [-.50, .63 [a]]	.67 [-.09, 1.25 [a]]	.61 [.03, 1.19]
	Actives	Grammatical	.05 [-.70, .43 [a]]	.05 [-.52, .61 [a]]	.13 [-.44, .70 [a]]
		Ungrammatical	-.52 [-.52, .62 [a]]	.56 [-.01, 1.14 [a]]	.18 [-.39, .74 [a]]
Noun vs. test-only	Passives	Grammatical	.40 [-.17, .98 [a]]	.19 [-.38, .76 [a]]	.34 [-.23, .91 [a]]
		Ungrammatical	.10 [-.46, .67 [a]]	-.23 [-.80, .34 [a]]	.06 [-.50, .63 [a]]
	Actives	Grammatical	.34 [-.23, .91 [a]]	.05 [-.51, .62 [a]]	.63 [.05, 1.21]
		Ungrammatical	-.09 [-.66, .47 [a]]	-.54 [-1.12, .04 [a]]	-.06 [-.63, .51 [a]]

4.3.3.1 d) Acceptability judgements for each error type

The errors included in the design of the incorrect items of the JT were as follows: missing auxiliary verb (50a), base form (rather than past participle) of the main verb (50b), and gerund form of main verb (rather than past participle) (-*ing*) (50c).

*50a. The child * being helped by the teacher*

50b. The dog is being walk by the boy*

*50c. The person is being follow*ing by the police*

Error type only affected ungrammatical items, so a separate model was run on the ungrammatical data:

*glmer (score ~ time * group * passive * error + (1 | subject), data=new, family=binomial, control = glmerControl(optimizer ="bobyqa"))*

No significant main effects or interactions of error type were found (see appendix 26 for full model results). However, some descriptive differences could be seen in the group mean scores. The biggest gain in mean score at immediate post-test were for -*ing* errors by the cue focused and test-only group. The noun focused group did not make much improvement and generally scored slightly lower for -*ing* items (see table 39).

Within-subjects effect sizes showed that the cue focused group improved their score the most between pre-test and delayed post-test compared to the other groups.

The noun focused group made the least gains for all error types by delayed post-test (shown by mean scores and between-subject effect sizes – see tables 40 and 41). Medium-large within-subjects effect sizes showed that the cue focused group improved most in judging -*ing* errors and base form errors compared to other error types (see table 40).

At immediate post-test, medium effect sizes between groups showed that the cue focused group improved their score more than the other groups. Between-groups effect sizes showed that this difference between groups was not apparent at delayed post-test (see table 41).

Table 39. JT mean % score for error types across time (SDs)

Group (*n*)	Error type	Pre	Immediate post	Delayed post
Cue (25)	Base form (*k*=16)	80 (13)	87 (12)	91 (11)
	-ing (*k*=16)	72 (16)	84 (13)	86 (8)
	Missing be (*k*=16)	78 (16)	85 (14)	85 (11)
Noun (25)	Base form (*k*=16)	81 (15)	81 (13)	88 (20)
	-ing (*k*=16)	79 (14)	78 (15)	82 (22)
	Missing be (*k*=16)	78 (12)	79 (13)	79 (21)
Test-only (23)	Base form (*k*=16)	75 (23)	80 (12)	82 (15)
	-ing (*k*=16)	70 (21)	81 (11)	85 (12)
	Missing be (*k*=16)	77 (19)	83 (13)	84 (13)

Table 40. JT within-subject effect size (Cohen's* d*) between test times for error type

Group (*n*)	Error type	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (25)	Base form (*k*=16)	.56 [-.06, 1.13[a]]	.35 [-.21, .91[a]]	.88 [.30, 1.46]
	-ing (*k*=16)	.82 [.25, 1.40]	.19 [-.37, .74[a]]	1.11 [.51, 1.70]
	Missing be (*k*=16)	.47 [-.10, 1.03[a]]	.00 [-.55, .55[a]]	.51 [-.05, 1.07[a]]
Noun (25)	Base form (*k*=16)	.00 [-.55, .55[a]]	.42 [-.15, .98[a]]	.40 [-.16, .96[a]]
	-ing (*k*=16)	-.07 [-.62, .49[a]]	.21 [-.34, .77[a]]	.16 [-.39, .72[a]]
	Missing be (*k*=16)	.08 [-.48, .64[a]]	.00 [-.55, .55[a]]	.06 [-.50, .61[a]]
Test-only (23)	Base form (*k*=16)	.27 [-.28, .83[a]]	.15 [-.43, .73[a]]	.36 [-.22, .94[a]]
	-ing (*k*=16)	.66 [.06, 1.25]	.35 [-.24, .93[a]]	.88 [.27, 1.48]
	Missing be (*k*=16)	.37 [-.21, .95[a]]	.08 [-.50, .66[a]]	.43 [-.16, 1.0[a]]

[a] 95% confidence intervals pass through zero

Table 41. JT between-groups effect sizes (Cohen's* d*) for error type

Group (*n*)	Error type	Pre *d* [CIs]	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	Base form (*k*=16)	-.07 [-.64, .50 [a]]	.48 [-.08, 1.04 [a]]	.19 [-.37, .74 [a]]
	-ing (*k*=16)	-.38 [-.93, .18 [a]]	.43 [-.13, .99 [a]]	.24 [-.31, .80 [a]]
	Missing be (*k*=16)	.00 [-.55, .55 [a]]	.44 [-.12, 1.01 [a]]	.36 [-.20, .92 [a]]
Cue vs. test-only	Base form (*k*=16)	.27 [-.30, .84 [a]]	.58 [.02, 1.15]	.69 [.11, 1.27]
	-ing (*k*=16)	.11 [-.46, .67 [a]	.25 [-.32, .82 [a]]	.10 [-.47, .67 [a]]
	Missing be (*k*=16)	.06 [-.51, .62 [a]]	.15 [-.42, .71 [a]]	.09 [-.47, .66 [a]]
Noun vs. test-only	Base form (*k*=16)	.31 [-.26, .88 [a]]	.08 [-.49, .64 [a]]	.34 [-.23, .91 [a]]
	-ing (*k*=16)	.51 [-.07, 1.08 [a]]	-.23 [-.79, .34 [a]]	-.17 [-.73, .40 [a]]
	Missing be (*k*=16)	.06 [-.50, .63 [a]]	-.31 [-.88, .26 [a]]	-.28 [-.85, .29 [a]]

[a] 95% confidence intervals pass through zero

4.3.3.2 The effects of animacy of judgements

As in the visual world eye-tracking, the JT included four noun animacy combinations:

51a. Animate-animate: The boy is pushed by the girl

51b. Animate-inanimate: The boy has been hidden by the rock

51c. Inanimate-inanimate: The car is hit by the bike

51d. Inanimate-animate: The wheel has been chased by the cat

The main aim of the JT was to investigate learners' judgements of passive forms. Therefore, the analysis for the passive items is presented first. There was no expectation that judgements for the trained or untrained sentences would have been affected differently by animacy, so both trained (present simple) and untrained (present perfect) passive and active items were combined in the scores presented in the following sections.

4.3.3.2 a) The effects of animate first nouns: Passive and Active items

In this analysis, the two types of animate first noun items are analysed separately as in 52 and 53:

52. Animate first noun – animate second noun (A-A): The boy is pushed by the girl / The boy is pushing the girl

53. Animate first noun – inanimate second noun (A-I): The boy is hidden by the rock / the boy is hiding the rock

For items with animate first nouns, all three groups scored higher on the active items at all test times than the passive items (see descriptive statistics in appendix 27).

Most of the within-subjects and between-groups effect sizes were small and unreliable (confidence intervals crossed zero) for all groups after training (at immediate and delayed post-test) (see appendix 27 for means and effect sizes). Within-subjects effect sizes for the cue focused group suggested that their score had improved the most for passive sentences (d = 1.12, CIs = .46, 1.76), compared to actives and the other groups. However, between-groups effect sizes showed that groups differences were small. The tentative trends in the descriptive statistics and effect sizes were not confirmed by the model as no main effect of animacy was found ($\chi^2(72) = 56.91$, $p = .90$).

4.3.3.2 b) The effects of inanimate first nouns: Passive and Active items

This section presents the analysis for sentences with inanimate first nouns. Example of items with inanimate first nouns:

54. Inanimate first noun – inanimate second noun (I-I): The car is hit by the tree / The car is hitting the tree

55. Inanimate first noun – animate second noun (I-A): The ball is hit by the girl / The ball is hitting the girl

It was expected that passive items with inanimate first nouns would be more easily interpreted than actives with inanimate first nouns, and than passives with animate first nouns. Mean scores showed that this was not the case. At pre-test, actives with inanimate first nouns were interpreted with more accuracy than passives with animate first nouns (see appendix 27 table 9). Improvements were less for inanimate first noun items than the animate items described above.

Within-subjects effect sizes showed that none of the groups made much improvement in score by immediate post-test (between tests, effects sizes were small and unreliable – see appendix 27 table 10). The cue focused group made the most improvements by delayed post-test, but only for passive items.

Between-groups effect sizes indicated that the greater improvement seen by the cue focused group was only statistically different to the test-only group, and only for animate-inanimate passives (see appendix 27 table 11). All other effect sizes were small and unreliable (confidence intervals passed through zero) (see appendix 27 for full tables).

The model found no main effect of animacy on judgement accuracy ($\chi^2(72) = 56.91$, $p = .90$).

4.3.3.3 Acceptability judgements of untrained constructions

The untrained constructions, those not practised in the intervention sessions, were either present perfect passive (example 56a) or present perfect progressive (example 56b).

56a. The dog has been carried by the boy

56b. The dog has been carrying the boy

The following sections compares judgements for the trained and untrained items. The model showed that present perfect items (passive and active combined) were .53 less accurately judged than present simple items (passive and active combined) at pre-test (logit coefficient: -.53, SE = -.19, $z = 3.3$). The same effect was found at immediate post-test ($\chi^2(48) = 7.17$, $p = .02$) and delayed post-test ($\chi^2(48) = 136.96$, $p = .00$).

4.3.3.3 a) A comparison of judgements of untrained and trained passive items

The mean scores for all three groups were similar for untrained passives (present perfect) and trained passive (present simple) items at pre-test (as shown by in appendix 27 table 13).

Cohen's *d* effect sizes for each group, at each test time, confirmed that there was no significant difference between trained and untrained passives as the 95% confidence intervals overlapped. Within-subjects effect sizes suggested that the cue focused group improved their score for untrained passives more than the other groups (cue focused, *d* = .75, CIs = .12, 1.37; noun focused, *d* = .39, CIs = -.21, .98; test-only, *d* = .66, CIs = .06, 1.27). However, between-groups effect sizes did not confirm this (see table 15 in appendix 27). At delayed post-test between groups effect sizes were small and unreliable (95% confidence intervals passed through zero) (for full results see appendix 27).

4.3.3.3 b) A comparison of judgements of untrained and trained active items

All three groups were more accurate in their judgements for trained (present simple) active items at than untrained (present perfect) active items at all test times. This was shown by the model by a significant main effect of trained v untrained at pre-test ($\chi^2(48)=87.79$, p = .00), immediate post-test ($\chi^2(48) = 7.17$, p = .02) and delayed post-test ($\chi^2(48) = 136.96$, p = .00)

Within-subjects effect sizes showed that the cue focused group, between pre and both post-tests, improved more for trained (present simple) active items than untrained (present perfect) active items, although the confidence intervals for trained and untrained overlapped, so this trend may not have been reliable.

The noun focused group, between pre and both post-tests, made the most gains for untrained (perfect) items compared to the trained items (as evidenced by small-medium within-subjects effect sizes – see appendix 27 table 18). However, overlapping confidence intervals for trained and untrained effect sizes indicate this difference was not reliable. The test-only group showed little difference in improvements between trained and untrained items at any test time.

Between-groups effect sizes were small and unreliable for both trained and untrained actives at both post-tests (see appendix 27 for full results).

4.3.3.4 Error identification for ungrammatical items

For items judged to be ungrammatical, the learners were asked to circle the word that they judged to be incorrect. 64 items (across pre, post and delayed-post tests in all groups) did not have an issue circled so these were excluded from the analysis (1.7% of items [64/3,787]).

The data were binomial because the scoring was either 1 = correct error identified, or 0 = incorrect error identified. Therefore, a logistic mixed-effects model was used to analyse this data. Error identification was not affected by grammaticality, since error identification was only performed on items the participants judged to be ungrammatical. Error identification was also not expected to be influenced by animacy. Therefore, fixed effects were group, time (test time), voice (passive or active) and error type. The maximal random-effects structure was used (Barr, Levy, Scheepers, & Tily, 2013). Random intercepts were subject and item. Further specification of the random-effects structure led to a failure to converge. The predictor variables were sum coded.

*Glmer (score ~ group * time * voice * error + (1 | subject) + (1 | item), data=new, family=binomial, control = glmerControl(optimizer = "bobyqa"))*

4.3.3.4 a) Error identification for passives and actives

For passives, mean scores increased slightly across the test times for the cue focused and test-only group, but not the noun focused (see appendix 27 table 21).

This was indicated by a significant main effect of immediate post-test found by the mixed-effects model (logit coefficient = 2.29, SE = 1.19, z = 1.92, p = .05) and an interaction, approaching significance, of the cue focused training and passives (logit coefficient = 1.75, SE = .99, z = 1.72, p = .08) There were no other significant effects or interactions found by the model (for full model results see appendix 26).

For passives, within-subjects effect sizes showed that the test-only group improved their score more than the other two groups between each test times. However, between-groups effect sizes did not suggest significant differences between the groups (see appendix 27 tables 21 – 24 for means and effect sizes).

For actives, all three groups improved their mean percentage error identification score between pre-test and both immediate and delayed post-tests. Between-group effect sizes showed very little difference between each groups' scores at each post-test (see appendix 27 tables 25 – 28).

4.3.3.5 Corrections of ungrammatical items – passives and actives

The final task the learners had to complete was correcting the sentences judged to be ungrammatical. The incorrect corrections were analysed descriptively since they were so few.

For passives, the test-only group showed the biggest change in score between pre and immediate post-test, and pre and delayed post-test, compared to the other groups, indicated by a large effect size between pre and delayed post-test (see appendix 27 table 30). However, between-groups effect sizes adjusted for baseline differences showed that the difference between the intervention groups and the test-only group at each post-test was small (appendix 27 table 32).

For actives, the cue focused group made the biggest gains in mean score between test times. Effect sizes adjusted for baseline differences showed that the cue focused group performed better than the other groups at immediate post-test (cue focused vs. noun focused d = .96, cue focused vs. test-only d = 1.03). This group difference was smaller by delayed post-test (see appendix 27 tables 29 – 36 for all means and effect sizes). A summary of the types of correction errors made is included in appendix 27 tables 37-50).

4.3.3.6 JT – summary of key findings

In terms of mean percentage acceptability judgements, all three groups improved between the pre-test and delayed post-tests. For passives and actives, the cue focused group improved earlier than the other groups showing most increase in score by immediate post-test. They maintained these gains till delayed post-test. The test-only group improved the least between test times.

Cohen's *d* effect sizes did not show substantial differences between each group at delayed post-test for passives or actives. For passives, the cue focused group improved most, and soonest, compared to the other groups. For actives, both intervention groups improved their judgement accuracy to a similar extent. The test-only group showed very little change across test times for actives. This suggests that training positively affected judgement scores on passive and active items and that the cue focused training had quicker, larger and longer-lasting effects.

Agent and patient animacy did not have any effects on judgement accuracy at any test time for any group. This was true for passive and active sentences.

Judgement accuracy for untrained and trained items improved in all groups and to a similar extent across the groups. Only the cue focused group showed any within-subject differences, by improving their judgement accuracy on untrained passives more so than the other sentence types. They also improved their score for these items more than the other groups.

In terms of error identification and correction, all three groups were very accurate overall at correcting errors, even at pre-test, and changes over time, and differences between groups, were small to negligible.

4.3.4 Summary of the results of the offline tests

Overall findings

- In the production tests, written and oral, the cue focused group improved the most over time in terms of morphosyntactic cue use.
- In the JT, the cue focused group improved in accuracy of acceptability judgements soonest compared to the other two groups (by immediate post-test) and maintained the gains till delayed post-test.
- All three groups showed some improvements in production (written and oral) and grammaticality judgement over time, but overall, the cue focused training appeared to result in greater improvements in accuracy compared to the other groups.

Voice

Results were mixed for the production of passives versus actives:

- In oral production:
  - Both the cue focused and test-only groups improved more in oral production of the passive voice than the noun focused group.
  - The cue focused group showed more improvement in oral production of actives compared to the noun focused and test-only group.
- In written production:
  - Both intervention groups only improved their written production of passives with little to no change in their written production of actives

- On the JT, voice did not affect score as both passives and actives were judged more accurately over time by all groups.

Morphosyntactic cues

When looking at morphosyntactic cue production (*be* + verb inflection and *by*), both types of training seemed to have benefitted correct verb inflection and use of *by*.

- In oral production, both intervention groups improved the accuracy of *be* + verb inflection over time.
- In written production, both groups improved the use of the verb and *by*. The cue focused group gained the most over time compared to the noun focused group.

Animacy

- Animacy of the agents and patients did not have an effect on production (written or oral).
- Animacy did not have a significant effect on acceptability judgments.

Generalisability (untrained items)

- In written production, the cue focused group improved the most for untrained items over time compared to the other two groups.
- Oral production of untrained passives and actives did not significantly improve for any group.
- In the JT, the cue focused group improved in judgement accuracy for untrained items over time more than the other groups.

Chapter 5: Discussion

5.1 Summary of the aims of the study

This study aimed to investigate how English NSs and Chinese learners of English process the passive voice, and whether teaching learners about morphosyntactic cues can aid their online processing, comprehension, and production. Previous research is mixed as to the extent to which NSs and learners use morphosyntactic cues to process input. This study aimed to teach learners to use these cues to facilitate processing and production. This builds on the few studies which attempted to train learners in cue use in order to facilitate processing (Andringa & Curcic, 2015; Hopp, 2016). Previous studies have not designed training activities to explicitly focus on cues and have not investigated online processing and comprehension along with offline production and grammaticality judgements. In our study, the noun focused training is, to some extent, comparable with the exposure element of the intervention used in Andringa and Curcic (2015) and Hopp (2016). These studies provided exemplars of the target structure with (Andringa & Curcic, 2015) or without (Hopp, 2016) rule explanation. Training (i.e., practice) that is designed to force attention onto morphosyntactic cues has not previously been investigated.

The following sections will show how the findings relate to each RQ. The first section of the discussion focuses on the online processing of the NSs (RQ3). This is followed by a discussion of the learners' data in comparison to the NSs' data (RQ3). The effect of training on the learners' use of the passive is then discussed, as are the effects of animacy (RQs 1 and 2).

5.2 Online processing of the passive and active voice by native speakers

This section discusses the findings relating to the main part of the third research question:

RQ3) To what extent do native English speakers show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

Overall, the NSs processed the active voice with greater ease than the passive. This was expected since the passive has been found to be acquired late by English NSs, potentially because of its complexity (see Armon-Lotem et al., 2016; Marinis & Saddy, 2013; Stromswold et al., 2002,). It also a less frequent structure in English than the active (Quirk et al., 1972). The findings for the passive and active sentences will be discussed in two parts: before the verb and after the verb. Eye-movements before the verb explored the effects of noun animacy on online role assignment. After the verb, eye-movements indicated if NSs showed sensitivity to the morphosyntactic cues i.e. the verb inflection and *by*.

5.2.1 Before the verb – effects of first noun animacy

During sentence processing, NSs of English have been found to rely on word order as a cue more so than noun animacy (Liu et al. 2012). The first noun in a sentence would commonly be assumed to be the agent (Bever, 2013; Cook et al., 2003; Keenan & Comrie, 1977). Noun animacy is a competing cue which has been found to be second in importance to word order in English (Bates & MacWhinney, 1993; Lui et al., 2012) and animate nouns are prototypical agents (Primus, 2010). In the current study, eye movements when the NSs heard the first noun indicated the effects of these competing cues (word order and animacy). If neither cue had an influence, and the NSs reserved role assignment until they heard the verb inflection and / or second noun, then looks to the target and distractor pictures would not have diverged prior to the verb phrase. If word order was the strongest cue for role interpretation, a bias towards the image depicting the first noun as the agent might have been expected. If noun animacy was a strong competing cue, animate first nouns might have been assumed to be agents more often than inanimate first nouns, and so we would have seen a bias towards the image depicting an animate first noun in trials where one noun was animate and the other inanimate.

The NSs appeared to use word order as a cue upon hearing the first noun. For active items, fixations on the target image diverged from the distractor at this point. This suggested that the NSs assumed an active interpretation of the sentences around hearing the first noun. For passives, around the first noun phrase, looks to target (i.e., the image depicting the first noun as a patient) were lower compared to active items. The comparatively lower proportion of looks to target for passives versus actives suggested a preference for first noun agency for some of the items (but not all) and a tendency to use word order to interpret the upcoming sentence.

The animacy of the first noun also appeared to have some influence over its interpretation. The eye movement figures visualising the eye-tracking data suggested that anticipation of the agent-patient roles varied slightly depending on first noun animacy, but the mixed effects model did not find the effect of animacy to be significant (see section 4.2.3.2). The variation seen in the eye-tracking data may have been due to the influence of a combination of noun animacy and real-world plausibility (as found by Kamide et al., 2003). Non-syntactic information, such as animacy and plausibility, have been found to influence the early stages of processing (Jackson & Roberts, 2010). In other words, interpretation of the first noun's role may have been dependent on its animacy *and* its real-world likelihood, an interaction which was not specifically manipulated in the current study. For example, an inanimate first noun is more likely in a passive construction than an active. However, this is not always the case, as the semantics of the nouns and the plausibility of a sentence is not always perfectly correlated with animacy. For these sentences, it could have been that noun animacy and real-world likelihood outweighed word order as a cue to agency since an inanimate agent is generally less common. Since the NSs were given time to preview the two images before hearing the audio, they had opportunity to anticipate the upcoming sentence (Ferreira & Lowder, 2016). Therefore, the scenarios depicted by the stimuli may have had some effect on first noun interpretation. Some of the scenarios may be more plausible than others, for example in the trial below (57).

57. The boy is filming the camera

In this example (57), the more likely scenario in the real world is that the *camera is filming the boy*, rather than *the boy is filming the camera*. Therefore, the NSs may have been more likely to interpret this as a passive construction on hearing '*boy*' as the first noun. However, the animacy of the first noun, and prior entrenchment of canonical word order, would have been expected to result in an active interpretation upon hearing *the boy*. In fact, for this particular item (*the boy is filming the camera*) there were more looks to the picture depicting the passive interpretation than the active.

Of the 24 active sentences with animate first nouns and inanimate second nouns 10 were assumed to be passive sentences. Examples of these sentences include: *the cat is transporting the box* (tended to be interpreted as *the cat is transported by the box*), the *boy is hitting the snow* (tended to be interpreted as *the boy is hit by the snow*) and *the bird is holding the cage* (tended to be interpreted as *the bird is held by the cage*). These verbs are more likely to occur in the passive because they result in the "patient to come to be in a relatively permanent end state" (Gries & Stefanowitsch, 2004, p.110). That is to say, the verb *hit* is more likely to be used to describe the patient having been hit rather than to describe an ongoing action.

In sum, the apparent differences in interpretation of some of the scenarios provided some evidence contrary to the idea that the first noun is always assumed to be the agent and that its interpretation is reassessed upon hearing the verb ending (contrary to findings by Ferreira, 2003). It seemed that the NSs assumed some of the first nouns to be agents, and some patients, potentially depending on the scenario – though this is an area requiring further analysis of the data as it was not intentionally manipulated. It is possible that real-world likelihood played a part in forming early judgements of upcoming roles (as found by Kamide et al., 2003), and that first noun animacy also had some effect, albeit potentially weakened by the effect of plausibility.

5.2.2 After the verb – morphosyntactic cue use

Evidence from research into online processing of relative clauses has shown that grammatical roles are more easily assigned when an animate first noun phrase and agency coincide (Jackson & Roberts, 2010; Mak et al., 2002). When these do not coincide (i.e., in the passives with animate first nouns or inanimate-inanimate sentences) this has been found to result in processing difficulty in the disambiguating region, due to morphosyntactic and semantic information in the verb phrase causing a reanalysis of the initial interpretation (Mak et al., 2002; Traxler et al., 2002; Traxler et al., 2005). If this influence of noun animacy also applied to passive sentences, it would be expected that the NSs would have experienced processing difficulty around the verb (disambiguating region) for passives with animate subjects or actives with inanimate subjects. This was not found to be the case. Animacy did not interact significantly with voice (passive or active), and passives with animate subjects or actives with inanimate subjects did not cause more processing difficulty after hearing the verb.

After hearing the verb, the NSs looked more to the target image than the distractor for all items – active and passive. For actives, looks to target increased most upon hearing the verb inflection. This suggested some sensitivity to the verb ending -*ing*. The roles in the active sentences may have been resolved sooner than in the passives due to facilitation, because the active interpretation was anticipated upon hearing the first noun (as noted above and found by Bar, 2009; den Ouden et al., 2012; Friston, 2010; Lupyan & Clark, 2015; Rao & Ballard, 1999). For the passive sentences, looks to the target image increased most upon hearing *by* (rather than on hearing the verb inflection -*ed*).

The increase in looks to the target image seen around the verb inflection in the actives, and *by* (or its absence) in the passives, suggested some use of these morphosyntactic cues to facilitate processing. These morphosyntactic cues may have been used in combination with the cues of word order and first noun animacy (see section 5.2.1 above) to resolve agent-patient roles. This fits with the Given-New concept of processing (proposed by Haviland & Clark, 1974 and developed by Ferreira & Lowder, 2016). This approach combines good enough processing and prediction as a facilitator for the integration of new information. This approach suggests that information that is already known by the parser (given information) is processed in a shallow, good-enough way, and that new information is processed using prediction as a mechanism for integration. In other words, successful comprehension is a combination of processing *given* information quickly and successfully integrating *new* information, facilitated by prediction mechanisms. In both the active and passive sentences in our study, the nature of the action was known (i.e., the likely verb stem) since images were provided prior to listening to the sentences. In the example below (58), the verb could have been predicted based on the given information depicted in the images.

58. The dog is carried by the bag

On hearing the first noun phrase, the most likely interpretation of the sentence in example 58 would be *the dog is carrying the bag*, based on a combination of the cues of word order and animacy (and potentially, real-world likelihood). So, for active sentences, the noun phrase and verb (e.g. *the dog* and *carrying*) would have been integrated with ease because the information was both predictable (due to word order and animacy cues) and given (due to the images). This was indicated by increased looks to target after hearing the first noun and straight after the verb inflection. For passives, on hearing the first noun (e.g. *the dog*), the same (active) interpretation was likely to be assumed, resulting in good enough processing up until the verb ending. In trials such as example 58 above, the verb ending (past participle, carri-*ed*) would have been less predictable than the *-ing* ending so processing would have been more difficult. This was indicated by no increase in looks to target at the offset of the verb. Upon encountering a past participle, the preposition *by* would have been expected to follow, therefore *by* would have been processed with greater ease than the verb ending. This was suggested by a slight increase in looks to target around *by*. So, the sentences appeared to be processed word-by-word, with expectations of the upcoming input being reassessed incrementally (as found by Altmann & Kamide, 1999; Andringa & Curcic, 2015; Boland et al., 1995; Kamide et al., 2003).

To summarise, the NSs processed active sentences with greater ease than passive sentences regardless of noun animacy, or type of passive (reduced or full). This indicated that the assumption that passive sentences would be more difficult for learners to process was likely to be a valid one. The sentences also appeared to be processed incrementally, and their interpretation was reassessed as the input progressed. The NSs did not always assume the first noun to be the agent. Various cues seemed to influence role judgement before hearing the verb. Whilst the NSs appeared to use word order as a cue, animacy and real-world knowledge also appeared to play a role. A number of studies have investigated the probabilities of certain verbs occurring in the passive and active voice (e.g. Altmann & Kemper, 2006; Bock, 1986, 1989; Bock & Loebell, 1990; Bock et al., 1992; Ferreira 1994; Ferreira, 1996) and the role these likelihoods play in online processing is an area for future investigation. Research has also found that NSs tend to misparse passive sentences considered to be implausible (Ferreira & Stacey, 2000). The effect of item plausibility on the NSs' online processing is also an area of potential future analysis.

In terms of morphosyntactic cue use, the NSs showed some sensitivity to the verb inflection and *by*. In actives, the NSs resolved the agent-patient roles after hearing the verb inflection. In passives, this occurred after hearing *by*. If the learners were also sensitive to these cues, similar evidence of agent-patient role resolution would be expected after hearing the verb inflection and *by*.

5.3 Online processing of the passive and active voice by second language learners

This section discusses the findings relating to the first research question:

RQ1) To what extent do L1 Chinese learners of L2 English show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

NS and learners are also compared in the following sections to address the second part of RQ3:

To what extent is native processing of the passive different to learners?

The eye-tracking pre-test data (i.e., prior to the experimental intervention), for the three learner groups combined, were analysed to determine how L1 Chinese learners of English processed the passive and active voice in English. A number of studies have found that L2 learners are able to process language in native-like ways (e.g., Dussias et al., 2013; Osterhout et al., 2006, 2008; Sabourin & Stowe, 2008; Tokowicz

& Warren, 2010), whereas other studies have found that this is not the case (e.g. Hopp, 2017; Kamide et al., 2003; Martin et al., 2013; Lew-Williams and Fernald, 2010). To add to this body of research the learner data are contrasted with the NS data in this section to investigate processing differences.

Active sentences were processed with more ease than passives by both NSs and learners. The NSs looked more to the target image than the distractor for passives, but proportionally this was a less strong tendency than for actives. The learners looked more to the distractor image than target for passives. This suggests that both the learners and NSs found passives more difficult to processes. However, the NSs were able to resolve the difficulty, whereas the learners were less able to do so. The difficulty for learners posed by passive sentences was expected since previous research has found that the English passive takes a long time to master (Hinkel, 2002; Izumi & Lakshmanan, 1998; Larsen-Freeman, 1997; Quinn, 2014; Williams & Evans, 1998), is processed with more difficulty by L2 learners compared to NSs (Marinis & Saddy, 2013), and is processed differently to the Chinese passive (Feng et al., 2015). Furthermore, noun animacy is a stronger cue in Chinese than word order, so competing cues may have caused greater difficulty for the learners than for the English NSs (who in the current study appeared to be only mildly influenced by animacy on the interpretation of voice, and these effects were not statistically significant) (Bates et al., 1991; Li et al., 1992, 1993; Miao et al., 1986).

The NSs showed sensitivity to the verb ending in actives (-*ing*) and *by* in passives. The learners also showed some sensitivity to the verb ending. For active items, upon hearing the verb inflection (-*ing*), looks to target surpassed looks to distractor, but this was not the case for passives (on hearing -*ed*). So, although the learners appeared to have noticed the verb ending (-*ing*) in active sentences, its absence in passives (and the presence of a past participle) did not seem to help them process the passive sentences. The learners also did not show any sensitivity to *by*. So, for passives, the learners exhibited uncertainty, throughout the input, as to which interpretation of the agent-patient roles was correct.

The learner data at pre-test presented a mixed picture with regard to comparing learner and NS processing. The data suggested that for the active voice in English, Chinese L1 learners were sensitive to verb inflection cues (-*ing*) in a similar way to NSs, but for passives this was not the case. This may have been because the active (present progressive) is a familiar construction, with canonical word order, and therefore the learners may be able to process it in a way approaching native-like. Research has found that low proficiency, or lack of language experience, can limit learners from processing in the same way as NSs (Jackson, 2008; Jackson & van Hell, 2011). Since the passive was a structure hypothesised to be less familiar to the learners in the current study this difference between the voices might have been expected. The findings at pre-test confirmed that teaching the learners to use morphosyntactic cues to process the passive voice with more ease was a worthwhile pursuit.

5.4 Morphosyntactic cue use after training for L2 learners

The following section focuses on the first and second RQs. The effect of training on online cue use is discussed first, followed by the effect of training on production and offline comprehension.

RQ1) To what extent do L1 Chinese learners of L2 English show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

Is cue sensitivity affected by:

a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?

b) First and second noun animacy?

c) Trained (present simple) and untrained (present perfect) constructions?

RQ2) To what extent do L1 Chinese learners of L2 English show knowledge of cues in a) offline grammaticality judgements and b) production of the passive voice in English?

Is production affected by:

a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?

b) Trained and untrained constructions (present simple and present perfect)?

Are grammaticality judgements affected by:

a) Training (either morphosyntactic-cue-focused or noun-focused) as compared to a test-only control group?

b) Trained and untrained constructions (present simple and present perfect)?

c) First and second noun animacy?

5.4.1 The effect of training on passive and active morphosyntactic cue sensitivity as evidenced by eye-tracking

The aim of the interventions was to investigate how different types of exposure to morphosyntactic cues could affect their processing and production. The cue focused training provided learners with EI and instruction with explicit focus on grammatical form, prior to and during sentence processing, *with* focus on morphosyntactic cues during practice (verb inflection and *by*). The noun focused training provided the learners with the same EI, but with practice in processing the passive and active voice *without* explicitly focusing on the morphosyntactic cues (i.e. without focus on grammatical form).

The next sections discuss the eye-tracking test results for the two intervention groups, compared to a test-only group, with regard to the processing of passive and active sentences as a whole, and the morphosyntactic cues individually, i.e. the verb inflection (gerund or past participle) and *by*.

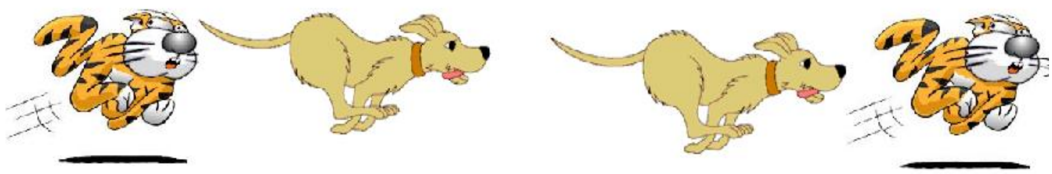
5.4.1.1 The effect of training on online processing of passive morphosyntactic cues

For passives, both intervention groups showed more sensitivity to the verb ending and *by* immediately after training than prior to training (i.e. immediate post-test vs. pre-test). Immediately after training, the noun focused group showed the greatest improvement, in terms of overall proportion and speed of looks to the target image. This group showed sensitivity to the verb ending cue (past participle or -*ing*), but not *by*. Compared to the other groups, the cue focused group showed the most change in sensitivity to *by* – the proportion of looks to target was greater around the offset of *by* compared to at pre-test. At immediate post-test, the test-only group did not look to target more than distractor at any point in the input, in other words, their interpretation of passives was unreliable. Therefore, both types of training appeared to have had an immediate positive effect on the processing of the passive voice. The noun focused training seemed to have increased sensitivity to the verb inflection and the cue focused training seemed to have resulted in increased sensitivity to *by*.

The cue focused group's reduced response (in terms of increased looks to target) to the *verb ending* compared to *by* was similar to the patterns of eye movements displayed by the NSs. This may be because the training focused on the morphosyntactic cues. For example, upon hearing the first noun in the item below (59), *the cat*, the learners were most likely to assume an active interpretation due to word order, and the fact that the first noun was animate. For passives, upon hearing the verb, if they were sensitive to the verb ending (-*ed*), then the 'surprisal' caused by the unexpected past participle may have resulted in processing difficulty which was resolved upon hearing the next (now expected) cue *by* (Ferreira & Lowder, 2016).

Their training focused on use of the verb inflection and *by*, so may have caused them to expect *by* upon hearing a past participle. However, they may also have been expecting the ending -*ing* most of the time due the likelihood of the active interpretation.

59. The cat is chased by the dog



The noun focused group's training did not draw attention to the verb inflection or *by* because the learners only had to focus on the first and second nouns to correctly interpret the agent-patient roles (though the EI before the practice *did* draw their attention to the form and meaning of the passive *be* + verb inflection + by). In example 60 below, the noun focused group only needed to listen to the first noun to resolve the sentence roles and decide which picture the sentence was compatible with. Their training may have resulted in the noun focused group learning to disregard *by*. Their training appeared to result in more sensitivity to the verb ending compared to the cue focused group, but only at immediate post-test (this was not the case at delayed post-test).

60. The cat is chased by the dog

For the training effects discussed above to be meaningful they would need to be apparent longer after training than the immediate post-test, since time is needed for the consolidation of newly acquired information (as suggested by Issa & Morgan-Short, 2019). Therefore, the findings at delayed post-test were arguably more important when determining the effects of the two training types. Six weeks after the training, at delayed post-test, the two intervention groups showed some sensitivity to *by*. Compared to immediate post-test, the noun focused group were less reliable at resolving the roles of the passives until they had heard the final noun (or its absence in the reduced passives). Previous research (e.g. Akakura, 2012; Ellis, 2009; Mackey, 1999) has found that instruction may have a delayed effect on implicit knowledge, due to the time required for processed input to be converted into implicit knowledge (Garcia & Gass, 2000; Nassaji & Fotos, 2004; VanPatten, 1996). If the effects of training are delayed, this would suggest that the training had resulted in the noun focused group relying on hearing the second noun phrase before making role judgements. This suggests that the noun focused training resulted in the learners becoming less sensitive to the morphosyntactic cues.

In contrast, by delayed post-test, the cue focused training appeared to have increased sensitivity to the verb inflection – looks to the target image increased greatly at the offset of the verb. After showing sensitivity to the verb ending, the cue focused group displayed some indecision about the correct interpretation of the passive sentences. Between the verb inflection and *by* they looked more to the distractor image. So, their training appeared to facilitate them responding to the verb ending cue but then resulted in some uncertainty. This could be a delayed effect of the training resulting in increased awareness and accessing of explicit knowledge. Similar was found for learners who received explicit instruction in Andringa and Curcic's (2015) study. Learners who received training with EI appeared to be uncertain about some agent-patient roles after training, specifically for sentences corresponding to images with matching first and second noun animacy. In the current study this was also found for matching animacy combinations (inanimate-inanimate passives, animate-animate and inanimate-inanimate actives). Another possible explanation for the low proportion of looks to target after the verb, might be that the cue focused training resulted in increased sensitivity to the verb inflection and so the learners interpreted the sentence at this point (and looked to target). Then having already interpreted the sentence, they might have looked to either picture somewhat randomly i.e. not in response the aural input. This explanation would mean that rather than being uncertain about the correct interpretation, the learners were so sure that they stopped processing the agent-patient roles immediately after the verb inflection.

Six weeks after the immediate post-test, the test-only group were the only group to look more to target than distractor from the offset of the verb onwards. Therefore, in terms of mean proportion of looks to target the test-only group outperformed the two intervention groups at delayed post-test. This may be because the repeated tests (pre, immediate and delayed) resulted in learning, and by delayed post-test the exposure to many examples resulted in increased sensitivity to the passive cues. This might suggest that repeated exposure to examples *without* any EI about cues could facilitate online processing. Studies have found that EI and explicit instruction which focuses on form helps learners to process input accurately sooner (using trials to criterion) than exposure without EI (e.g. Henry et al., 2009, Fernández, 2008). It might be that by the delayed post-test, after exposure to three tests-worth of examples, the test-only group had managed to accumulate sufficient practice items to show improvements in their processing.

To summarise, in the current study, for passives, it is not possible to say that either intervention had a positive effect on processing, but processing changes that lasted until delayed post-test were observable. These changes seemed to consist of: an increase in use of the verb inflection for the group trained to focus on the verb inflection (but this did not seem to 'last' in terms of interpretation during the rest of the sentence); little evidence of increased sensitivity to cues in the group that were simply exposed to the structures after having received EI; some evidence of increased sensitivity to cues in the group that only did the tests.

5.4.1.2 The effect of training on online processing of active morphosyntactic cues

As seen for the passives, directly after training, the noun focused group showed the most improvements in processing speed and accuracy for active sentences - they looked to the target image sooner, and proportionally more, than the other groups (and compared to pre-test). At immediate post-test, compared to the noun focused group, the cue focused group looked less to target overall, but they appeared more sensitive to the verb ending (-*ing*) (as the noun focused group's divergent looks were later in the sentence). This was similar to the NS behaviour for actives. So, straight after training, the cue focused group behaved similarly to the NSs for both passive and active sentences.

As noted for the passives, the delayed post-test findings likely provided more of a reliable indication of instructional effects. Six weeks after training, as seen for passives, the noun focused group appeared to be more reliant on the second noun than the verb inflection (looks to the target image increased most around the second noun phrase). The cue focused group's looks to target were lower than at the other test times (looks to target and distractor were equal throughout the input i.e. at all time windows). In other words, for actives, the cue focused group could not reliably interpret the agent-patient roles. It is difficult to say why the cue focused training would result in greater processing difficulty for active sentences, particularly since sensitivity to the verb ending was observed for passives. The test-only group looked more to target around the end of the verb and throughout the rest of the input, and so as for passives, processed the actives with greater ease than at the other test times.

So, as with the passive items, both types of training may have had a somewhat negative effect on online decision-making. Both intervention groups appeared to process passive and active sentences with less speed and accuracy after training, this may have been due to EI being counterproductive, at least within the timescale of the current project, when the L1 and L2 are syntactically quite different. In a study investigating the effects of EI on offline grammaticality judgements and written production, Andringa et al. (2011) found EI to be unhelpful for certain linguistic features when the learners' L1 and L2 express things syntactically differently, and only beneficial when the L1 and L2 are similar. Although Andringa et al. (2011) is not directly comparable to the current study, because online processing was not investigated, it may be that the limiting effect of L1-L2 difference on the effectiveness of EI on offline measures might also affect online processing. This may have been the case in the current study since the passive in Chinese and English is expressed syntactically differently.

Since both training types resulted in different processing behaviour at delayed post-test this suggested that the two types of training might have affected processing differently. The cue focused training, which focused on the form-meaning connection of passive and active morphosyntactic cues, appeared to result in sensitivity to the passive verb ending but not the active. This provided evidence that focus on some, but not all, morphosyntactic cues can result in increased sensitivity to them. The noun focused training focused on the nouns' role and appeared to result in overreliance on the nouns for interpreting the sentences. This suggested that focus on the nouns in training resulted in greater sensitivity to them during processing. Both training types provided some evidence that focus on form, whether it be morphosyntactic or lexical, could result in changes during online processing. This is somewhat in line with research finding that focus on form, in particular PI, results in online processing changes (Henry, 2015; Lee & Doherty, 2019) but contrary to research finding that form focused instruction, specifically PI, does not result in online changes (Dracos, 2012). The findings from the current study did not provide evidence that PI-like training (as in the cue focused group) had a more positive effect on online processing than other training types, but did suggest that focus on form can modify processing.

5.4.2 The effect of training on passive and active morphosyntactic cue production and interpretation as evidenced by offline tests

The learners completed two production tests, one oral and one written, and a written untimed grammaticality JT. The next section will discuss the findings of these tests, first for the passive and active voice as whole constructions, and then the individual cues (verb inflection and *by*).

5.4.2.1 Oral production test– sentence completion task

In the oral production test, taking within-subjects effects sizes for both passives and actives into account, the cue focused group made the most gains immediately after training compared to the other groups. These improvements in production were maintained until delayed post-test. However, by delayed post-test between-groups effect sizes suggested that group differences were small. The EI in cues plus the explicit nature of the practice and the feedback given in the cue focused training may have improved production sooner than just being presented with exemplars, as in the noun focused training. This is in line with previous research showing that explicit instruction that focuses on form plus explicit feedback results in higher accuracy in constrained production tasks than training that is less explicit in nature (as found by Carroll & Swain, 1993; Marsden, 2006; Marsden & Chen, 2011; Muranoi, 2000; Nagata, 1993; Sanz & Morgan-Short, 2005). Tasks that require learners to produce target language in restricted contexts, without time pressure, have been thought to measure explicit knowledge more than implicit knowledge (Ellis, 2004). The constrained nature of oral sentence completion tasks, like the one in the current study, means that they most likely measure explicit knowledge to a greater extent than implicit knowledge. Greater gains on measures which tap more into explicit knowledge than implicit knowledge are typically seen as a result of explicit training (for reviews see Andringa, 2011; DeKeyser, 2003; Norris & Ortega, 2000; Spada & Tomita, 2010). Since the cue focused group were likely to be highly aware of the cues being focused on (due to their training focusing explicitly on grammatical form), they may have been better able to access explicit knowledge, resulting in improvements in production accuracy sooner than the other groups.

To explore the oral production of passives and actives in more detail, the scoring was broken down into use of *be* + the correct verb inflection, and correct use of *by*. After training, the cue focused group showed the most improvements in their oral production of the correct verb inflection (past participle or gerund). The test-only group made the least gains. Based on effect sizes, both training types had a similar positive effect on accurate production of the verb inflection at immediate post-test, but the cue focused training resulted in longer lasting effects as evidenced by further gains at delayed post-test. It was expected that the cue focused group would improve their production accuracy of the verb inflection the most since the practice activities in their training explicitly drew attention to its morphology. The noun focused training practice activities did not focus on morphosyntax, but this group improved to a similar extent as the cue focused group at immediate post-test. This could be because the EI prior to practice was given to both groups and drew attention to the verb ending. So, it may have been the presence of EI that resulted in the improvements seen by the noun focused group. EI, with and without explicit instruction, has been found to result in learning sooner than training without EI (as found by Henry et al., 2009, Fernández, 2008). However, a number of studies have found that EI in itself does not appear to affect learning (e.g. Benati, 2004; Sanz, 2003; Sanz & Morgan-Short, 2005; Stafford et al., 2012). It is only possible to tentatively explain the noun focused groups' gains as potentially resulting from being given EI because the noun focused training tasks were not tested without EI. An extension to the current study would be needed to unpick the effects of EI in combination with two intervention types, in which both sets of training tasks were given without EI. This would allow for the importance of EI to be better assessed, the current study did not allow for the role of EI in itself to be investigated.

As previously mentioned, delayed post-test scores are perhaps most useful as a window into L2 development. Therefore, the further improvement seen for the cue focused group at delayed post-test indicated an advantage of the cue focused over the noun focused training for production accuracy of the cues *be* + verb inflection. This suggested that EI along with instruction that focuses on the morphology of grammatical cues might be more beneficial for the oral production of those cues than EI with instruction without focus on morphosyntactic form.

All three groups' oral production of *by* was very accurate at pre-test and did not improve in any significant way after training. However, since the EI given to the intervention groups focused on *by,* some change in score might have been expected for the intervention groups. Perhaps this may have been because *by* is a more salient cue than the verb ending and the learners were already aware of it, hence their high pre-test scores, and therefore did not pay as much attention to its usage in the training. This also meant that there was a strong ceiling effect for production of *by* resulting in less room for improvement.

5.4.2.2 Writing test

For written production of passives, the noun focused group improved the most in the test immediately after training. This finding was in contrast to the findings from the oral production test in which the noun focused group improved the least for passive sentences. The cue focused group improved their written production score, to a similar extent as they did in the oral production test. The test-only group improved their score by delayed post-test but less so compared to the other groups (i.e., with smaller effect sizes).

For active sentences, the three groups all made small improvements by immediate and delayed post-test, but all three groups had scored very high at pre-test (mean = 91%) so there was very little room for improvement due to ceiling effects.

The increased improvement by the noun focused group in the written production of passives compared to oral production might be explained by the different modality of the tests. Previous studies have found that instruction effects are not always seen in oral production tests but are more apparent in written production tests (as found by Day & Shapson, 2001; Marsden, 2006; Marsden & Chen, 2011; VanPatten & Sanz, 1995). Therefore, if the noun focused training did have an effect on production, it would likely be more evident in a writing test.

As in the oral production test, the writing test's overall score constituted use of *be* + correct verb inflection and use of *by*. The cue focused training increased production accuracy of *be* + verb inflection more than the other groups (i.e., effect sizes were large between pre- and delayed post-test). In terms of use of *by*, the three groups' use of *by* was not as accurate at pre-test as it was in the oral production test. Therefore, there was some scope for improvements. For written production of *by*, the cue focused group showed the biggest gains by immediate and delayed post-test compared to the other groups. As in the oral production test, it appeared that the cue focused training improved cue use more than the noun focused training. This is likely to be because the training forced attention to the verb ending and *by* whereas the noun focused did not. The smaller gains seen by the test-only group suggested that the cue focused groups' gains in both written and oral production were due to their training.

The greater improvement seen in cue production by the cue focused group showed that explicit focus on form can result in improvements in accurate language use in oral production tasks as well as written sentence level tasks, as has been found by various studies investigating PI (e.g. Benati, 2001, 2004, 2005; Cadierno, 1995; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Wong, 2004). As mentioned above, improvements in oral production of cues were not as marked for the noun focused training group, which did not focus on form, compared to in the written production task, and compared to the cue focused group.

5.4.2.3 Written untimed acceptability judgement test (JT)

The JT was the final test in the battery. The analysis was divided into judgment rating, error identification, and error correction. Since the more explicit nature of the cue focused training seemed to have resulted in the greatest gains in cue use for the production tests, the same would be expected to be the case for the judgement measure, the JT, since explicit instruction has consistently been found to have a positive effect on grammaticality judgements (Bialystok, 1979; Ellis, 1991; Hedgcock, 1993; Suzuki, 2017).

For passives, by delayed post-test, the cue focused training resulted in increased judgement accuracy (medium within-subjects effect sizes and larger between-groups effect sizes compared to the other groups). The other groups did not improve significantly (small, unreliable within-subjects effects sizes). For active items, the change in score over time was smaller compared to passives for all groups. Both intervention groups improved their accuracy in judging active sentences to a similar extent by delayed post-test, and the test-only group did not improve. So, the cue focused training had a slightly greater positive effect on judgement accuracy for both passives and actives.

All three groups were more accurate at identifying ungrammatical sentences if they were actives; for passives, accuracy was lower. The cue focused group showed the greatest improvement across test times for the passive ungrammatical items compared to the other groups (medium and reliable effect size versus small unreliable effects sizes). Previous research has shown that ungrammatical items in untimed JTs tap into explicit knowledge more so than grammatical items (R. Ellis, 2005; Godfroid et al., 2015). For example, Gutiérrez (2013) found that grammaticality of items had the greatest effect on the type of knowledge used in JTs (implicit or explicit knowledge), even more so than time pressure, and argued that grammatical items measured implicit knowledge whereas ungrammatical items measured explicit knowledge. This might be because ungrammatical sentences have a clear critical area (i.e., the error) that may encourage access of explicit knowledge (see Bialystok, 1979; Ellis, 1991; Hedgcock, 1993). The additional focus on cues and feedback in the cue focused training, combined with the ungrammatical nature of the sentences, may have resulted in greater access to the explicit knowledge required to correctly identify ungrammatical sentences. Furthermore, the explicit focus on grammatical form in the cue focused training appeared to result in a better ability to identify errors around that form. This indicated that instruction that forces learners to make form-meaning connections might be beneficial for noticing and correcting errors. To investigate whether instruction that explicitly focuses on form results in learners being able to assess grammaticality better, more studies comparing instructional types need to be done. Most studies investigating instructional methods like PI have not used a JT in their test battery and have not compared multiple instructional methods and a control group.

For the sentences judged to be unacceptable, the learners were asked to identify the error by circling the incorrect word. Accuracy for error identification was high for all three groups. The cue focused group and test-only group both increased their score by immediate and delayed post-test. The noun focused group did not, their mean group score was the same at pre-test and delayed post-test. This was likely due to their error identification score for passive items being very high at pre-test (94%) so there was very little room for improvement. The sentences judged to be unacceptable were also corrected by the learners. The learners were asked to do this to investigate if they knew why the word they had identified as erroneous was incorrect. Scores at pre-test were not as high for corrections as they were for error identification. This suggested that it was easier for the learners to recognise errors than it was to correct them. All three groups improved their scores by delayed post-test. Between-groups effect sizes showed that group differences were small. Therefore, as for error identification, training did not have an effect on correction accuracy.

5.4.3 Untrained constructions – the present perfect passive and present perfect progressive

To investigate the learners' ability to generalise cue use to untrained constructions, two constructions were included in the outcome measures that were not included in the training – the present perfect simple passive and the present perfect progressive (active). In this section, the learners' online processing of the untrained sentences after training is discussed. This is followed by a discussion of the effects of training on production and interpretation (i.e. grammaticality judgement) of the untrained constructions.

5.4.3.1 Generalisability from trained to untrained constructions in online processing

In the eye-tracking test, prior to training, untrained items (both active and passive) and trained passive (present simple) items were processed in a similar way - looks to target did not diverge much from looks to distractor. However, active untrained items (present perfect continuos) were judged less accuratey than the trained items (both passive and active), suggesting that these were more complex and / or unfamiliar than the other structures.

For the untrained sentences (passive and active combined), after training, the noun focused group and the cue focused group looked proportionally more to the target image than at pre-test. The test-only group looked less to the target than distractor at the post-tests. This showed that the two intervention groups appeared to be able to apply the EI they had received on the trained constructions (present simple passives and present simple progressive) to the untrained perfect tense. This suggested that the explicit knowledge about cues for the trained constructions was accessible to the learners during online processing and could be implicitly transferred to the untrained constructions.

To investigate the untrained items further, actives and passives were analysed separately (present perfect passives and present perfect progressive actives). At pre-test, the learners processed the untrained passives (present perfect passive) with more ease than the untrained actives (present perfect progressive) – they looked to target proportionally more for untrained passives than actives. Perhaps because the present perfect progressive is less common that the present perfect passive. In a corpus study into NSs academic writing, the present perfect progressive was used only 0.21% of the time. In contrast, the present perfect passive was used was 4.65% (Alzuhairy, 2012).

After training (at immediate and delayed post-test), for untrained passives (present perfect simple), the cue focused group looked proportionally more to the target image than the other two groups. The cue focused training seemed to result in more sensitivity to the verb inflection as looks to target increased upon hearing the verb and remained higher than distractor throughout the rest of the input. At delayed post-test, the noun focused group did not look more to target than distractor until after the offset of *by* – slightly earlier in the input than at immediate post-test (in which most looks to target were around the second noun), but later than the cue focused group. After showing no changes in processing at immediate post-test, at delayed post-test, the test-only group looked to target around the verb but then immediately exhibited uncertainty (looks to target were less than to distractor for the remaining time windows). This was quite different to the trained passives, for which the test-only group looked to the target image more than distractor throughout the input.

For untrained actives (present perfect progressive), at immediate post-test, only the cue focused group appeared sensitive to the verb ending. At delayed post-test, all three groups showed sensitivity to the verb ending (more looks to target than distractor around the end of the verb), but then for the remainder of the time windows the proportion of looks to target did not clearly diverge from looks to distractor. Since the untrained active sentences were most likely to be the least familiar to the learners it might have been expected that they would continue to have problems interpreting them after training. Their behaviour after training might have reflected a need for more time to be able to transfer recently trained morphosyntactic cues to the present perfect progressive construction. The need for time to assimilate recently learned language has been found to result in a U-shaped trajectory of learning (Kellerman, 1985; Long, 2010; Long & Robinson, 1998). That is, mistakes would be made earlier in the learning process and more time and practice would be needed to see gains. This might be one explanation for the lack of ability to transfer from trained to untrained actives at delayed post-test.

It must be acknowledged that difficulty in processing the present perfect, in particular the present perfect progressive, may not only be evidence of an absence of ability to generalise from the trained constructions. The present perfect's complexity and infrequency may have also been a cause of the observed difficulties in its processing. The present perfect has been found to be learned late by NSs (Bardovi-Harlig 1997; Gathercole 1986) and learners (Dulay, Burt & Krashen 1982; Ellis, 1994; Fuchs, Götz & Werner, 2016). It is also more common for telic achievement verbs i.e., verbs describing a completed action (as observed by Collins, 2002). Since verbs of this type were used in both the passive and active sentences in order to create stimuli with reversible animate and inanimate agent-patient roles, the present perfect sentences may have been unusual in some of the depicted scenarios, this was likely to be particularly true for some of the present perfect progressive items.

In summary, it can tentatively be said that the cue focused group appeared to be able to generalise their knowledge of passive cues during online processing better than the other groups at both post-tests (this was not the case for actives). The cue focused training included training tasks and metalinguistic corrective feedback explicitly focusing on the morphosyntactic cues needed to generalise from the present simple passive (trained) to the present perfect passive (untrained) (i.e. the auxiliary *be* + past participle). The explicit nature of the training (EI + focus on cues) may explain why this group appeared to be better at generalising passive cue use during online processing. Previous research has found that awareness, and access to explicit knowledge, is a predictor of the ability to generalise in offline measures (e.g. in elicited oral production as found by Brooks et al., 2006; Brooks & Kempe, 2013; Kempe et al., 2010; Kempe & Brooks, 2008). The current study provides evidence for this in online processing, as evidenced by eye-tracking.

5.4.3.2 Generalisability from trained to untrained construction in production and grammaticality judgements

As in the eye-tracking test, the present perfect simple passive and the present perfect progressive (active) were included in the production tests and JT to investigate the learners' ability to generalise cue use to untrained constructions.

In the oral production test, there was no significant difference found by the mixed-effects model in oral production score between trained and untrained sentences. For untrained passives, the cue focused and test-only group both improved their production score by delayed post-test. The difference between the two groups was small. The noun focused group did not improve their accuracy on untrained passive items. Scores for untrained actives were high at pre-test and did not change much over time. In the writing test, the untrained items (perfect tense) were produced with similar accuracy to the trained items by all three groups at all test times. Slight improvements were seen for passive items only, since for active items pre-test scores were high. There was no difference between the groups' post-test scores.

In terms of oral and written production, it is not possible to say that either training resulted in a better ability to apply rules to untrained constructions. This may have been due to the formulaic nature of the test and the use of the cue "*since this morning*" to prompt production of the present perfect. The learners may have been used to responding to *since* with the present perfect as it is a prompt typically used in the EFL classroom and in instructional materials. The learners may have struggled to produce the untrained constructions in a freer production task. However, research has found that explicit instruction and more implicit instruction can both positively impact free production, so differences between the intervention groups would not necessarily be expected (e.g., Andringa et al., 2011; Sanz & Morgan-Short, 2004; Williams & Evans, 1998).

The findings for the production of the present perfect progressive items differed from those of the eye-tracking test. Untrained actives caused the most online processing problems for the learners. This showed that online processing of these untrained sentences was more difficult than their production. This would be expected because in production tasks, such as the ones in the current study, learners have time for reflection and planning before responding. This gives them opportunity to use explicit knowledge to make conscious and controlled responses (Marinis, 2012). Tests like this can sometimes overestimate language abilities (ibid), this might explain the difference between the findings for the untrained sentences in the production tests and the eye-tracking tests. Therefore, evidence of group differences in ability to generalise knowledge of the passive and active cues might have been more reliable in the eye-tracking task.

In the JT, untrained sentences (both passive and active combined) were judged significantly less accurately than trained sentences by all three groups at all test times. This showed that these sentences were harder to judge than the trained sentences. All of the learners showed some improvements in judgement accuracy for both passive and active untrained items but group differences at each post-test were small. Therefore, it was not possible to say that either training had a more positive effect on the judgement of untrained items (passives and actives combined).

For untrained passives, the cue focused group showed the biggest improvement overall in accuracy of acceptability judgements at both post-test times (as shown by within-subjects effect sizes). However, between-group effect sizes showed that the difference between groups at delayed post-test was small. Improvements were much smaller for untrained actives across time for all groups. So, as in the eye-tracking task, the untrained actives caused the learners the most problems. For the passives, it seemed that for all of the learners (to a slightly greater extent the cue focused group) abstract knowledge of morphosyntax related to the passive voice (i.e. the auxiliary *be* + verb inflection) was available to be transferred from the trained constructions to the untrained.

To summarise, after training, the cue focused group were slightly better able to generalise rules learnt for the trained passives to the untrained passives in the JT than the other two groups. This slightly greater improvement in judgement accuracy might be due to explicit training and feedback resulting in a better ability to generalise learnt grammar, in this case when judging grammaticality. This was not the case for written or oral production (contrary to findings found for elicited oral production by Brooks et al., 2006; Brooks & Kempe, 2013; Kempe et al., 2010; Kempe & Brooks, 2008). It is not possible to definitively say that the cue focused training resulted in a better ability to generalise the rules learned in the training in the offline tests (JTs and production). Although some benefit of the cue focused training was seen for the untrained passives in the JT, and also in the eye-tracking test.

The findings for the cue focused group also built on previous research into the effects of form focused instruction and explicit information (e.g. PI) by investigating untrained items. The results from the JT suggested that focus on form might increase learners' abilities to generalise from a previously learned construction to a more novel one. Since the noun focused group also improved their acceptability judgement accuracy it may also be that EI influenced learners' ability to generalise when recognising grammatical errors. This study did not investigate the effects of EI without training. Therefore, this study does not add much to evidence of EI's importance, however, as previously discussed EI might have played a part in the improvements seen in the noun focused group who did not received instruction focusing on morphosyntactic form.

5.5 The effect of first and second noun animacy on learner interpretation of the passive voice - online (eye-tracking) and offline (judgement test).

Animacy is a cue to agent-patient roles. In the current study, animacy was manipulated to investigate if different agent-patient animacy combinations affected the online processing and error judgement of the passive and active voice. Animacy was manipulated in the interventions and the outcome measures (eye-tracking test and JT) so that its effects before and after training could be investigated (see RQ1b and 2b).

RQ1) To what extent do L1 Chinese learners of L2 English show sensitivity during online processing to morphosyntactic cues in the English passive voice as evidenced by visual world eye-tracking?

Is cue sensitivity affected by:

b) First and second noun animacy?

RQ2) To what extent do L1 Chinese learners of L2 English show knowledge of cues in a) offline grammaticality judgements and b) production of the passive voice in English?

Are grammaticality judgements affected by:

b) First and second noun animacy?

5.5.1 Animacy effects and online processing

Research has shown that animacy affects online processing (e.g. Andringa & Curcic, 2015; Hinkel, 2002; Jackson & Roberts, 2010; Mak et al., 2002; Traxler, Morris, & Seeley, 2002; Traxler et al., 2005; Weckerly & Kutas, 1999). This was expected to be the case for the learners in the current study. This was partly because Chinese speakers are more sensitive to the animacy of nouns than word order when determining agent-patient roles (Bates et al., 1991; Li et al., 1992, 1993; Miao et al., 1986; Su, 2001), and partly because in English and Chinese, animacy is a cue to agency. For example, animate nouns are usually agents, and inanimate nouns are usually patients (Bates & MacWhinney, 1989; Primus, 2010; Su, 2001). So, transfer of L1 interpretation strategies to the L2 may occur and result in inefficient processing (e.g. as found in Chinese by Jiang, 2004, 2007; Zhang, 2015, and other languages by Hopp, 2010; Isabelli, 2008; Jackson, 2007; Jackson & Dussias; 2010, Tokowicz & MacWhinney; for a review Frenck-Mestre, 2005). The differing strength of the competing cues of animacy in Chinese (L1) and word order in English (L2) was expected to potentially cause issues for the learners.

The eye-tracking test contained the following animacy combinations:

61a. Animate-animate: The boy is pushed by the girl

61b. Animate-inanimate: The boy is hidden by the rock

61c. Inanimate-inanimate: The car is hit by the bike

61d. Inanimate-animate: The wheel is chased by the cat

Prior to training, the animacy of the first noun seemed to affect the processing of both passive and active sentences. For active sentences, animate-inanimate and inanimate-inanimate sentences were often misinterpreted as passive sentences (reflected by looks to the picture conveying the passive voice). So, for example upon hearing *The boy…*, the sentence was assumed to be *The boy is hidden by the rock* rather than *The boy is hiding the rock.* The same was true for sentences such as, *The car is hitting the bike*, upon hearing *The car…*, the learners expected the passive construction *The car is hit by the bike.* For these sentences, the data provided some evidence that the learners were making judgements about agency prior to hearing the verb. This finding is contrary to that of previous research into the processing of relative clauses which found that learners did not incrementally assign roles prior to the point of disambiguation. In other words, learners did not assign roles based on the animacy of the first noun (as found for German L2 learners of Dutch, Havik et al. 2009). It is also contrary to the idea that L2 learners tend to assume first nouns to be agents i.e. the First Noun Principle (VanPatten et al., 2013).

The learners did not judge roles prior to hearing the verb for all animacy combinations. For animate-animate and inanimate-animate sentences (in both passives and actives), the learners did not seem to make any judgements as to the correct interpretation before hearing the verb (looks to target and distractor were similar prior to the verb). If the learners did assume the first noun to be the agent, upon hearing the first noun in the animate-animate sentences, looks to the image depicting the active interpretation would have been expected. However, this was not always the case, and the learners appeared unsure about a possible interpretation of roles. It was expected that the learners would interpret inanimate-animate sentences as passives more often than actives, because the passive would be the most likely interpretation since animate nouns are prototypical agents (Bates & MacWhinney, 1989; Su, 2001), however this was not the case. This might be because learners have been found to postpone assignment of roles when agency and animacy do not coincide (Jackson & Roberts, 2010). In the current study, this would apply to sentences with inanimate subjects and animate objects. For these sentences, the animacy of the subject might not facilitate interpretation of agent-patient roles. As a result, the learners appeared to wait until after the verb to assign roles and look more to one of the images (the passive or active interpretation).

Another potential explanation for the uncertainty exhibited by the learners for the inanimate-animate sentences were the competing cues of word order and animacy. As previously mentioned, advanced Chinese learners of English have been found to be influenced by word order more than lower proficiency Chinese learners of English (Su, 2001). Since the learners' proficiency varied to some extent in the current study (mean IELTS = 6.00, SD = .42, range = 5.5-8.5), this may explain why some of the learners in the current study tended to interpret inanimate-animate sentences as active. This finding seemed somewhat contrary to research showing that late bilinguals tend to forward transfer Chinese L1 cues to L2 English (as found by Liu et al., 1992). In Liu et al., Chinese learners of English rarely interpreted inanimate first nouns as agents.

To summarise, the pre-verbal eye-tracking data prior to training suggested that learners made some judgements about agency based on first noun animacy. However, this was only apparent for animate-inanimate and inanimate-inanimate sentences. Animate-animate sentences would be expected to result in looks to both active and passive depictions since both were equally likely. Lexical–semantic expectations, such as animacy and word order, may have resulted in the learners postponing role assignment for inanimate-animate sentences until after hearing the verb (as also found by Jackson & Roberts, 2010).

5.5.1.1 First noun animacy and voice after training

After training, by delayed post-test, the cue focused group tended to assume the active interpretation most of the time, except for animate-animate sentences for which they were most unsure. These sentences were just as likely to be interpreted as active or passive upon hearing the first noun. This makes sense since either noun would be equally likely to be the agent of the sentence. The tendency to look to the image depicting the active interpretation, regardless of first noun animacy, suggested reliance on word order, rather than animacy, as the main cue to agency. This might have been due to the learners' proficiency increasing as a result of training, causing animacy effects to have been weakened and word order to have more influence, as found by Su (2001) for high proficiency Chinese L1 learners of L2 English.

By delayed post-test, the noun focused group seemed to be the least affected by animacy. This was indicated by an equal proportion of looks to target and distractor for most animacy combinations (except inanimate-animate). This suggested that the learners were considering both the passive and active interpretation of the sentences on hearing the first noun. In other words, any bias due to first noun animacy seemed to have lost its effect. This may have been because the cues of word order and animacy were competing, or that some of the learners tended to use animacy more than word order to interpret roles, and vice versa. The apparent lack of first noun animacy influence may have been an indication that the noun focused training had less of an effect on proficiency than the cue focused training. The cue focused group appeared to rely on word order more than the noun focused group. Another explanation for this might be that because the noun focused training focused on the noun semantics, rather than the morphosyntactic cues, the learners reserved role judgement until after hearing the second noun (as shown by the post-verbal eye-tracking data for this group at delayed post-test).

At delayed post-test, the test-only group tended to interpret sentences with animate first nouns as active and inanimate first nouns as passives. This suggested that they were relying on animacy as a cue to interpret the first noun's role and were making judgements about agency prior to the verb, as was the case at pre-test. There was little difference between pre-test and delayed post-test for the test-only group which suggested that the differences in animacy effects exhibited by the intervention groups were likely to be as a result of their training. The findings for the three groups, though different, provided evidence that learners were able to make role judgements prior to encountering the verb (as also found by Jackson & Roberts, 2010 Havik et al., 2009; Hopp, 2006; Jackson, 2008).

5.5.1.2 Post-verbal animacy effects and voice after training

Noun animacy has been found to affect processing of the main verb in a sentence (indicated by increased RTs) so it was expected that animacy effects would be apparent after the verb (as found by Mak et al., 2002; Traxler et al., 2002; Traxler et al., 2005). After training, the cue focused group judged inanimate-animate passives as actives the majority of the time (higher looks to distractor than target for the time windows after the verb). This was unexpected because these sentences would be more unusual in an active form than the passive form as they would result in sentences such as: *The car is stopping the girl*; *The ball is hitting the girl*; *The donut is eating the man*. This active voice bias may have been a result of the higher frequency of active sentences compared to passive sentences that the learners will likely have encountered in their English learning experience both in the English classroom and in a UK university context. In NS academic writing, the passive voice has been found to be used less often than the active voice (median rate of frequency: passive voice = 1.32; present simple active = 9.72; past simple active = 1.82) (Hinkel, 2004). The passive voice has also been found to be used less in spoken British English than written British English (Roland et al., 2007). English language textbooks also tend not to address tense, voice, or aspect in much detail and so exposure to the passive voice is minimal (Hinkel, 2004). The processing difficulty seen for inanimate-animate passives might also have been due to speakers of Asian languages having been found to experience difficulty comprehending sentences with inanimate subjects and active verbs (Master, 1991) e.g., *The ball is hit by the girl.*

At delayed post-test, inanimate-inanimate and animate-animate sentences were often misinterpreted as passives (the mean proportion of looks to distractor was proportionally lower than to target). The increase in animacy effects for matched animacy sentences might have been due to the cue focused training resulting in the learners 'over-thinking' the agent-patient roles to some extent. Using a visual world eye-tracking task in which learners listened to sentences and saw images of two objects on a screen (half the sentences had inanimate direct objects and half animate). Andringa and Curcic (2015) found a similar period of indecision for sentences corresponding to images with matching first and second noun animacy after training that included EI (compared to training without EI). The authors suggested this might have been due to increased thinking about roles as a result of the EI. In the current study both intervention groups received EI, so this potential period of indecision might have been expected for both groups. The more explicit nature of the cue focused training and feedback might have had a greater effect on accessing explicit knowledge than the noun focused training, in line with previous research showing that explicit instruction results in access to explicit knowledge (see Andringa, 2011; DeKeyser, 2003; Norris & Ortega, 2000; Spada & Tomita, 2010).

The noun focused group, for passives, seemed to be the least affected by animacy at immediate post-test (after hearing the verb, looks to target were higher than distractor for all animacy combinations). By delayed post-test, the noun focused group misinterpreted passives and actives with inanimate first nouns. So, if the noun focused training had reduced the strength of animacy effects on whole sentence interpretation, for the sentences with inanimate first nouns the effects were shortlived. For inanimate-animate passives and inanimate-inanimate actives, the noun focused training appeared to result in reliance on the second noun (looks to the target images surpass distractor only at this point). For these sentences, the focus on the noun semantics in the training seemed to result in the learners waiting to judge the sentences' roles until after hearing the second noun.

The test-only group were the least affected by animacy at delayed post-test (this was also the case for first noun interpretation). For passives, only inanimate-inanimate sentences were processed inaccurately (this was the case for all three groups). For actives, animate-inanimate sentences were processed more accurately than at the other test times. There was little visible difference between the other animacy combinations. It may have been that the repeated exposure to passive and active sentences and all animacy combinations (without the influence of any training) resulted in a reduction in animacy effects for the test-only group.

To summarise, training seemed to have had some effect on the strength of noun animacy as a cue to agent-patient roles. The cue focused training may have resulted in some indecision, in particular for matched animacy sentences (specifically for inanimate-inanimate passives, and for animate-animate and inanimate-inanimate actives). The noun focused training appeared to reduce animacy effects straight after training, but these effects did not last till delayed post-test (six weeks after training). Their training, which focused on noun semantics, appeared to result in reliance on the second noun for interpreting agency.

5.5.2 Animacy effects on written grammaticality judgements

In order to complete the JT, the learners first had to decide if each sentence matched the image provided (as in example 60a below). Therefore, interpretation of the roles in the sentence had to occur before a judgement of its grammaticality could be made. Interpretation of roles has been found to be affected by animacy offline (in a picture matching task by Liu et al., 1992) and online (Andringa & Curcic, 2015; Jackson & Roberts, 2010; Mak et al., 2002; Traxler et al., 2002; Traxler et al., 2005; Weckerly & Kutas, 1999). A number of these studies also found that first noun animacy affects the processing of the main verb during reading (indicated by increased RTs in SPR and eye-tracking) (as found by Mak et al., 2002; Traxler et al., 2002; Traxler et al., 2005). Since the errors in the JT involved the verb, noun animacy might have affected its judgement. Few studies have investigated animacy affects using both online and offline measures. In a study investigating noun animacy effects and gender agreement, animacy was found to have an effect on both online (SPR) and offline (JT) sentence interpretation (Sagarra, & Herschensohn, 2011). In the current study, animacy was also manipulated online (eye-tracking) and offline (JT) in order to investigate if noun animacy affected online and offline interpretation of passive sentences.

As in the visual world eye-tracking test, the JT included four noun animacy combinations. Each item was presented alongside one matching image (see 62 and image matching 62a).

62a. Animate-animate: The boy is pushed by the girl

62b. Animate-inanimate: The boy is hidden by the rock

62c. Inanimate-inanimate: The car is hit by the bike

62d. Inanimate-animate: The wheel is chased by the cat

Prior to training, animacy appeared to have some effect on acceptability judgements. Animacy was found to be a significant predictor of judgement accuracy at pre-test (as evidence by the mixed-effects model). For sentences with animate first nouns, active items were judged more accurately at pre-test than the passive items. For passives and actives with inanimate first nouns (e.g. 62c and d above), judgement accuracy was similar. This was contrary to what might have been expected – that passive items with inanimate first nouns would be more easily interpreted than actives with inanimate first nouns since agents tend to be placed in the subject position and prototypical agents are animate (Barker & Dowty, 1993; Primus, 2010). This may have been evidence of the learners focusing purely on the morphosyntactic cues in order to judge some of the sentences, and as a result disregarding the matching image. In other words, interpretation of the entire sentence and its roles may not have always been taking place. Therefore, noun animacy would not interfere with all of the grammaticality judgements. Furthermore, since the JT was untimed, it may have been that the learners reassessed their original interpretation prior to recording their response. A timed JT would have better explored whether noun animacy consistently affected grammaticality judgements.

After training, animacy was not found to have a significant effect on judgements. Effect sizes (within and between subjects) were small for all three groups and animacy combinations. The effects seen at pre-test for actives with animate first nouns were no longer apparent after training and / or testing. It may have been that the repeated testing resulted in the learners ignoring the images and focusing only on the region potentially containing an error (around the verb). Therefore, as previously suggested, the animacy of the nouns, and corresponding images, may not have had any effect on judgements, and, interpretation of the sentence as a whole may not have been required to complete the task.

Chapter 6: Conclusion

This chapter draws together the main findings of the current study and discusses them in terms of their possible contributions to language processing research and language teaching research. The study's limitations are also acknowledged, and future research is suggested.

6.1 Summary of the study

This thesis presented the findings of an intervention study investigating morphosyntactic cue use by learners of English, specifically in order to aid processing, comprehension, and production of the passive voice in English. Participants were 73 Chinese learners of English and 29 native English speakers. The learners were all studying pre-master's courses at the University of York and were upper intermediate English learners. The passive voice was the target feature because it is frequently used in academic English, so is likely to cause problems for this group of learners. Furthermore, English and Chinese differ in their use of the passive voice and in its construction.

This study compared the effectiveness of two types of instruction on passive voice cue use. A group of L1 English speakers were also tested to compare native speaker online processing of the passive voice with that of the learners. The two types of instruction differed in their focus. The cue focused training drew learners' attention to the morphosyntactic cues of the present simple passive and contrasted them with those of the present progressive. The noun focused training focused the learners' attention on the semantics of the nouns. The present simple passive was contrasted with the present progressive in the training to provide a comparison between the morphosyntactic cues in each structure i.e., *The cat is chas**ed** **by** the dog* versus *The cat is chas**ing** the dog*. Two computer-based, intervention sessions of around 20 minutes were administered to the individual learners, one week apart.

One week prior to training, the learners were given a series of tests to assess their comprehension and usage of the passive, which were: a visual-world eye-tracking test; an oral production test; a written production test; and a written untimed grammaticality JT. The test battery was also administered immediately after the second (i.e. final), training session, and six weeks after the final training session. A test-only group was included to investigate the effects of the outcome measures. The tests assessed the accuracy of comprehension, production, and grammaticality judgement of the passive voice by the learners. Present perfect constructions were also included to investigate the learners' ability to transfer rules learned in the training to untrained constructions. In order to test the learners' cue use online, the eye-tracking test used sentences containing reversible agent-patient roles of varying animacy combinations e.g., *The cat is chased by the dog* has two animate nouns which could be role-reversed. The oral production test and written production test were designed to investigate controlled production of the passive cues. Finally, the judgement test was used to investigate the effect of training on the learners' ability to identify and correct errors.

6.2 Summary of the findings

The NS eye-tracking data showed that passive voice sentences were harder to process than active voice sentences, which was also true for the learners. The learners also interpreted both active and passive sentences with less accuracy than the NSs. This showed that the passive voice is indeed a structure that causes online processing problems for Chinese learners of English.

The eye-tracking data were analysed to investigate processing prior to hearing the verb in each sentence to explore the effects of first noun animacy. The NSs appeared to use various cues to interpret agent-patient roles prior to the verb. Word order appeared to play a part in determining roles; the active interpretation appeared to be preferred which suggested that the first noun was assumed to be the agent. This would be expected since SVO is the most common word order in English (Bever, 2013; Cook et al., 2003; Keenan & Comrie, 1977). Animacy and real-world knowledge may also have had an effect on first noun role interpretation. Some sentences seemed to be interpreted as passives potentially because the verb may have been more likely to occur in a passive construction, regardless of the first noun's animacy. This is a tentative observation which requires further investigation.

The eye-tracking data suggested that the NSs processed the audio input word-by-word and made predictions about the upcoming input incrementally rather than waiting until the end of each sentence. Active sentences were resolved sooner than passives, after the verb ending, whereas looks to the target image in the passives were highest after *by*. This was evidence of the NSs using the morphosyntactic cues to facilitate processing and was in line with the Given-New theory of input parsing proposed by Haviland and Clark (1974) and developed by Ferreira and Lowder, (2016). Since most sentences were assumed to be active upon encountering the first noun, the present progressive verb ending, *-ing*, would be expected, and therefore would facilitate interpretation of the sentence. The passive past participle verb ending, e.g. *-ed*, would be unexpected so would not result in role resolution. Then, after hearing the past participle, *by* would be expected, so would facilitate interpretation. This was tentatively evidenced by a slight increase in eye movements to the correct image around *by*.

The learners' eye-tracking test data prior to training (at pre-test) were compared with that of the NSs. Unlike the NSs, the learners did not assume an active interpretation of the upcoming sentence. Additionally, first noun animacy did not have a strong effect prior to the verb, but first and second noun animacy appeared to effect processing later in the input (i.e. after the verb). There was a great deal of variation between learners, animacy combinations, and items. Research has found that high proficiency Chinese learners of English use word order as the main cue to roles, and that lower proficiency learners tend to rely on animacy (Su, 2001). The fact that the learners in the current study were approaching advanced level may have resulted in differing cue strengths for different learners and account for the variation seen between items and participants.

Both training types had an effect on the processing and comprehension of the passive. In terms of online processing, training appeared to have a delayed effect as there were different findings at immediate post-test and delayed post-test for both training groups. The delayed post-test findings were likely to be more indicative of training effects than the immediate post-test because time is needed for the consolidation of information acquired during the experiment (Issa & Morgan-Short, 2019). At delayed post-test, the cue focused training appeared to result in indecision after the verb (particularly for sentences with nouns of the same animacy). Sensitivity to the verb ending increased but was then followed by more or less equal looks to target and distractor until the second noun. This may have been due to the cue focused training providing EI + explicit instruction + metalinguistic feedback, and the activation of metalinguistic knowledge causing over-thinking and doubt (as found by Andringa & Curcic, 2015 and Cornillie et al., 2017). Another explanation could be that the cue focused group resolved the agent-patient roles after hearing the verb ending and reverted to random looks to each image based on non-linguistic information. This might suggest increased sensitivity to the verb inflection after training. The noun focused training reduced sensitivity to the verb ending over time and seemed to result in reliance on the second noun in order to interpret voice. Their training focused on the meaning of nouns not the morphosyntactic cues, so this reliance on the nouns could be explained by their training.

The observed training effects were further evidenced by the test-only group. After having already received the test twice (i.e. at delayed post-test) the test-only group looked proportionally more to the target images than the intervention groups. They appeared to be more sensitive to the verb ending at delayed post-test than the other groups. Eye-movements showed that they chose the correct image after hearing the verb and did not exhibit the indecision seen by the intervention groups. The difference in behaviour between the intervention groups and the test-only group suggest that the interventions had some effect on processing.

Training on passive cue use was found to have a positive effect on the production of those cues. The cue focused training resulted in greater improvements in oral and written production of the verb ending. This is in line with previous research which found that training of an explicit nature, in this case explicit training on cue use, results in production gains on constrained production tasks (Carroll & Swain, 1993; Muranoi, 2000; Nagata, 1993; Sanz & Morgan-Short, 2005). The cue focused training also resulted in improvements seen sooner than the other two groups (i.e. at immediate post-test) in the written production test. The explicit nature of the cue focused training seemed to result in improvements in cue production sooner (with less time for consolidation, perhaps) than training focusing on the nouns. These gains were maintained until delayed post-test. This may be because instruction effects are often more pronounced in written tests compared to oral production tests (Andringa et al., 2011; Day & Shapson, 2001; Marsden, 2006; Marsden & Chen, 2011; VanPatten & Sanz, 1995). This may also explain why the noun focused group improved more in the written test than they did in the oral production test, albeit to an overall lesser extent than the cue focused group.

Focusing on morphosyntactic cues during training also appeared to have a greater positive effect on grammaticality judgements compared to focus on the noun semantics. The cue focused training had a particularly positive effect on judgement accuracy for passive sentences and ungrammatical items. Ungrammatical items have been found to tap into explicit knowledge more so than grammatical items on JTs (R.

Ellis, 2005; Godfroid et al). The cue focused training explicitly focused on the passive and active morphosyntactic cues and provided extra feedback, both of which have been found to be linked to increased awareness and explicit knowledge (Doughty, 2008; R. Ellis, 2005; Han & Ellis, 1998; Norris & Ortega, 2000). So, focusing on morphosyntactic cues may have helped the learners better recognise errors and correct them. The noun focused group and test-only group also improved their scores over time, although to a lesser extent, which suggested that exposing learners to examples, with or without EI, may also be of some benefit when it comes to judging grammaticality.

6.3 Limitations and future research

The present study had a number of limitations which reduced the generalisability of its findings, but that also suggest avenues for future research.

The first issue was the artificial nature of the stimuli in the training and outcome measures used. As pointed out by Ferreira and Lowder (2016), an issue with research into sentence processing and comprehension is that “subjects are typically shown lists of single, unrelated sentences, or occasionally they might be presented with sentence pairs” (p.236). This was the case with the stimuli in the current study. Stimuli of this type were used to control for a number of variables and to allow for counterbalancing across conditions. A lack of authentic stimuli is problematic because the learners may not be able to apply what they learnt to real-world discourse.

A further issue with the current study, and other experimental research like it, was that of the artificial, and restricted, nature of the outcome measures. Furthermore, research suggests that the effects of explicit instruction are more observable on explicit knowledge measures (such as JTs, especially untimed JTs as found by Plonsky et al., 2019) and restricted production tasks (such as sentence completion tasks) than freer production measures (Andringa et al 2011, DeKeyser 2003). Therefore, the results might have been different in a more naturalistic language context. However, research has also found explicit knowledge to be more flexible than implicit knowledge, and therefore more applicable to tasks differing from those used in training (De Jong, 2005; Reber et al., 1999). This has been suggested to be because explicit knowledge can be generalized and used flexibly across different contexts (Ullman & Lovelett, 2016). Skill acquisition theory suggests that declarative knowledge learnt about language can be proceduralised through repeated use (DeKeyser, 2007). That is, explicit knowledge plays a role in the development of implicit knowledge over much practice. Once fully automatized, knowledge which is more implicit in nature, is not so readily applied to other skills, for example from comprehension to production (as found by De Jong, 2005; DeKeyser, 1997; DeKeyser & Sokalski, 2001; Shintani, Li & Ellis, 2013). In order to test explicit instruction's influence on freer production, an oral and written production test such as story narration without prompts could have been used. Due to the nature of the tests used, it was not possible to say that the gains made in the post-tests after training would result in improved production and comprehension outside the experimental context.

Since the stimuli used were unusual in some cases, real-world knowledge and plausibility may have had an effect on processing and on the effectiveness of the training. Plausibility has been found to influence processing, particularly in the early stages of language learning (Jackson & Roberts, 2010). The preview of the images, when certain scenarios were less plausible than others, might have resulted in processing difficulties later in the input since role assignment may have been influenced by the likelihood of those roles in the real-world. The issue with plausibility extends to the verbs as well as the agent-patient roles. The relative likelihoods of certain verbs appearing in the active or passive voice has been investigated (e.g., Gries & Stefanowitsch, 2004), and the role these likelihoods play in online processing is an area for future investigation.

Building on previous research into explicit instruction, this study investigated the effects of two types of explicit training. The noun focused group received the same EI as the cue focused group and also received feedback as to whether their responses were correct or not (yes / no feedback). The reason for not comparing the cue focused training with a more implicit training was twofold (i.e., training without EI or feedback). The aims of this study were to investigate different types of focus on form i.e., focus on morphosyntactic cues versus focus on the meaning of nouns, as might be done in a less grammar-focused (more meaning-focused) classroom. And, secondly, to ensure parity of exposure to the target forms between the groups compared, avoiding a situation in which EI would be included in training as an 'extra activity', which is a weakness of research in this area observed by Andringa et al. (2011). Nevertheless, if amount of exposure to tokens of the target feature and time on task could have been controlled, it may have been useful to explore the effect of mere exposure alone to the passive (i.e., noun focused without EI), in order to determine the effects of EI + cue focused training to a more incidental learning condition where learners would have had to induce the rules.

It was outside the scope of the current study to explore the separate effects of EI, feedback and the cue focused training tasks. Since the cue focused tasks were not provided with and without EI, or with and without feedback, it is not possible to determine the strength of the effect of cue focused tasks in isolation. The results of the noun focused group give some insight into the role of feedback, since they only received yes/no feedback on the meaning of the nouns. For this study to be more conclusive, the cue focused training would need to be provided with and without EI and feedback. Since research has found that explicit instruction and feedback improves scores on EK measures, it may be that the EI + metalinguistic feedback resulted in the gains seen by the cue focused group, rather than being as a result of the practice in using cues during sentence processing (Doughty, 2008; R. Ellis, 2005; Han & Ellis, 1998; Norris & Ortega, 2000). Future research could explore this by comparing cue focused tasks + EI + feedback with cue focused tasks + EI - feedback and with cue focused tasks - EI - feedback and with a test-only group.

Future research into the benefits of cue focused instruction in various L1s and L2s would give insight into the effects of crosslinguistic difference. Since cue use varies across languages, effectiveness of cue focused instruction may also vary. Research has found that if a construction is not present, or common, in the learners' L1, then its morphosyntactic cues will not be used during processing (e.g. Tokowicz & McWhinney, 2005 and Tokowicz & Warren, 2010). Cue focused training may be more effective when the learners' L1 has similar cues, such as the Spanish passive voice which has a similar construction to English. To investigate this, a replication of the current study could be done with Spanish learners of English.

The learners in the current study were of a higher proficiency than learners in most intervention studies which tend to sample intermediate learners (Plonsky, 2013). This was because training on the target feature would be of particular benefit to them, and because the number of overseas students increased in UK universities in the past few years, researching problems that this group may encounter with English grammar was a worthwhile pursuit. Replicating the current study with a group of lower proficiency learners may find more pronounced results. The effect of focusing on grammatical cues when the structure is novel to learners would provide more insight as to the effectiveness of these types of training. However, the EI and feedback provided may have to be administered in the learners' L1 for this to be effective with lower proficiency levels

6.4 Contributions to:

6.4.1 Research into language processing

Previous research that has found that NSs anticipate roles based on word order cues and animacy (e.g. Bar, 2009; Den Ouden et al., 2012; Friston, 2010; Liu et al., 1992; Lupyan & Clark, 2015; Rao & Ballard, 1999) this study adds to this body of research as the findings suggested that NSs do use these cues to process input in an anticipatory way. Research is mixed as to whether NSs also use morphosyntactic cues to facilitate processing. Some research suggests that NSs make superficial, and as a result inaccurate, interpretations of language input, and therefore are not sensitive to morphosyntactic cues (Ferreira et al., 2002) as evidenced by research into garden-path sentence processing (e.g. Christianson et al., 2001; Ferreira & Stacey, 2000; Patson et al., 2009). Other research has found that NSs use morphosyntactic cues, such as case marking, to anticipate meaning (e.g. DeLong et al., 2005; Hopp, 2017; Kamide et al., 2003). The current study contributes evidence to suggest that NSs use morphosyntactic cues, as well as word order and animacy cues, to facilitate processing when those cues are expected, and that they revaluate expectations as input unfolds.

In terms of L2 processing, the current study provides evidence that Chinese L1 – English L2 learners approaching advanced level proficiency tend not to be sensitive to English passive voice morphosyntactic cues. Previous research has found that learners incrementally process input (e.g., Jackson, 2008; Jackson & Roberts, 2010; Juffs & Harrington, 1995, 1996; Roberts & Felser, 2011) possibly through morphosyntactic cue use (Hopp, 2017) or using other strategies such as the first noun principle (VanPatten et al., 2013). Prior to training, the current study did not provide evidence to support either of these potential processing mechanisms. It did provide evidence that instruction can affect processing, and that learners may use cues to process sentences incrementally after repeated exposure to cues (with or without explicit training in their use). The effect that instruction in the current study had on processing was mixed. Compared to prior to training: the cue focused training increased verb inflection sensitivity; the noun focused training increased sensitivity to the second noun phrase, and no training (test-only) resulted in increased sensitivity to the verb inflection.

This study also indicates the importance of using a broad test battery when investigating language use. While, the passive voice was found to cause processing problems for the learners in the current study, these processing difficulties did not predict production problems. The passive was produced much more accurately compared to its online interpretation. Many processing studies do not use offline tests alongside online methods whereas the current study used a battery of online and offline tests. This demonstrated the need for various measures of language ability since performance in online and offline tests may differ.

6.4.2 Research into second language instruction

A great deal of previous research has investigated the effects of explicit instruction on language processing and use. The current study was the first, to my knowledge, to design training that explicitly focuses on morphosyntactic cues with the intention of increasing learners' cue sensitivity as they encounter them in an unfolding sentence. This study builds on recent research which aimed to encourage learners to (or investigate if they could) use cues to aid processing (Andringa & Curcic, 2015; Hopp, 2016), it also builds on a wide body of research investigating the effects of instruction which focuses on form-meaning connections, such as PI. The current study did not aim to teach prediction per se, as in the previous studies, but aimed to teach cue sensitivity in order to facilitate processing (role assignment) and production. Whilst the findings of the current study were not definitive, it appeared that instruction focusing on the morphology of cues might have a beneficial effect on offline language use, and that online cue sensitivity might be affected by training focusing on cues.

Further research into the nature of the relationship between EI, instruction focusing on cues, and online processing needs to be done in order to gain more insight into the way in which this type of instruction might benefit learners.

In the current study, the test-only group showed most improvement in terms of accuracy in online processing. So, there is some indication that the tests themselves may have resulted in online processing gains (or, in another interpretation, the two treatments caused disruption to processing). However, turning to the offline tasks, instruction that focused on morphosyntactic cues appeared to have a positive effect on the production of those cues and improved learners' ability to deal with errors (recognition and correction). However, focusing on cues did not necessarily have a positive effect on online processing; it may have resulted in indecision, which indicates increased awareness and activation of explicit knowledge. Further studies into instruction that focuses on morphosyntactic cues need to be done to determine whether the observed effects in the current study can be generalised to other languages and grammatical structures. Instruction that focused on noun semantics, rather than morphosyntax, also appeared to be beneficial for production. In addition, testing (i.e. the test-only group) also had some benefits on processing and production. Therefore, it seemed that repeated examples of a structure (with or without meaningful practice activities) may result in improvements in production and online comprehension.

The current study including EI alongside both training types. EI, with and without explicit instruction, has been found to result in learning sooner than training without EI (as found by Henry et al., 2009, Fernández, 2008). The presence of EI alongside the noun focused training might explain the noun focused group's gains in some of the tests, such as the production measures. On the other hand, a number of studies have found that EI in itself does not appear to affect learning (e.g. Benati, 2004; Sanz, 2003; Sanz & Morgan-Short, 2005; Stafford et al., 2012). The current study did not allow for a conclusion about the role of EI to be made because the two training types were not administered without EI. This would be an important future extension of this study.

In sum, the current study showed the importance of a test-only group and of the use of a variety of outcome measures in instruction research, as (a) no training, but undertaking tests can reveal useful findings, and (b) different effects can be seen between online tests and offline tests. The current study also suggested that instruction that explicitly focuses on form (cue focused) might be more beneficial in terms of production of those cues compared to instruction that does not focus on form. It also indicated the potential importance of EI alongside instruction, although the effect of EI needs to be unpicked further to be sure of the magnitude of its influence.

Appendix 1 - Types of Chinese passive

Table 1 - passive constructions in Chinese

Syntactic	Example
Bèi	Règǒu bèi nánhái chī le Hotdog by the boy eat [ASP] *The hot dog was eaten by the boy.*
Wei….suo	ta wei ta de ai suo gandong, ta jueding quanli zhichi ta de shiye she PSV he GEN love PRT move, she decide full support he GEN career *She was moved by his love and decided to support his career fully.*
Gei…	Wo mama ye gei ci-le I mother also PSV fire-ASP *My mother was also fired.*
Jiao….	Zhe-xia bu jiao wo cai-zhun-le? This-CL not PSV I guess-right-ASP *Haven't I guessed right this time?*
Rang…	Wo rang ta tou-le liang-kuai qian I PSV/ask/allow he steal-ASP two-dollar money *I had two dollars stolen by him/I asked (allowed) him to steal two dollars.*
Lexical	
Ai…	youde haizi zai jia ai-le da, chu jiamen jiu zhao ren faxie

	some children at home suffer-ASP beat, out house-gate then look-for other give-vent-to *Having been beaten up at home, some children let off their anger on others when they go out.*
Shou…	Yi-ge shou-le hechi de xiaoxuesheng One-CL suffer-ASP berate GEN schoolchild *A schoolchild who has been berated*
Zao...	Youzhiyuan suishi you zao pohuai de weixian Kindergarten any-time have suffer destroy GEN risk *The kindergarten risked being destroyed at any time.*

Appendix 2 – Consent forms

Consent form – learners (intervention groups)

PARTICIPANT CONSENT FORM

Project title: Processing of syntax in L2 English

Researchers: Sophie Thompson and Emma Marsden

Come and experience an English learning experiment! Improve your English and experience an Education/TESOL experiment!

If you take part in this study, you will:
get **<u>£20 for taking part</u> in all sessions**
<u>improve your English</u>
get **<u>insight into Education/TESOL related experiments</u>**

What this involves and benefits

Taking part will involve 4 sessions:

Session 1) you will complete an eye-tracking task, a speaking task, a writing task and a grammar judgement task. (1 hour 15 mins)

Session 2) instruction on a feature of English. (30 minutes)

Session 3) instruction on a feature of English followed by an eye-tracking task, a speaking task, a writing task and a grammar judgement task. (1.5-2 hours)

Session 4) you will complete an eye-tracking task, a speaking task, a writing task and a grammar judgement task. (1 hour 15 mins)

The exact dates of these sessions will be arranged at times convenient for you.

The sessions will take place approximately as follows:

1st session (in July).

2nd session will be about 2 weeks after the 1st (in July).

3rd session will be about 1 week after the 2nd (in August).

4th session will be about 4 weeks after the 3rd (in September).

Anonymity

The data you provide will be stored by code number. Any information that identifies you will be stored separately from the data and will be destroyed 2 weeks after the end of data collection. Before this time you will be able to withdraw your data at any time. After the end of data collection, and the identifiable data is destroyed, it will no longer be possible to withdraw your data.

Storing and using your data

Data will be stored in secure filing cabinets and on a password protected computer. The anonymous data will be kept for an indefinite period. The data may be used for future analysis, and it may be made available via the internet for other researchers to see and use for research or training purposes, but participants will not be identified individually – only the results of the language tasks will be available (i.e. numbers and reaction times).

Information about confidentiality

The data that we collect (your responses) may be used *anonymously* in different ways. Please indicate on the consent form enclosed/attached with a ☑ if you are happy for this anonymised data to be used in the ways listed.

We hope that you will agree to take part!

If you have any questions about the study that you would like to ask before giving consent or after the data collection, please feel free to contact Sophie Thompson by email sophie.thompson@york.ac.uk or the Chair of Ethics Committee via email education-research-administrator@york.ac.uk

If you are happy to participate, please complete and sign the form attached and leave on the desk in front of your computer.

Please keep this information sheet for your own records.

Thank you for taking the time to read this information.

Yours sincerely

Sophie Thompson (Department of Education)

Dr. Emma Marsden (Department of Education)

English grammar training study

Consent Form

Please tick each box if you are happy to take part in this research.

I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve me taking part as described above. ☐

I understand that the purpose of the research is to investigate the effectiveness of syntax processing instruction ☐

I understand that data will be stored securely on a password protected computer and only Sophie Thompson and Emma Marsden will have access to any identifiable data. I understand that my identity will be protected by use of a code and no others will be able to recognise that I participated in the study. ☐

I understand that my data will not be identifiable after the final session ☐

This anonymous data may be used:

in publications that are mainly read by university academics and language teachers ☐

in presentations that are mainly attended by university academics and language teacher ☐

in publications that are mainly read by teachers ☐

I understand that Sophie Thompson will keep the file linking my name to the data for up to two weeks after data collection, after which it will be destroyed ☐

I understand that the anonymised data will be kept indefinitely as it may be made freely available online and it could be used for future analysis or research training ☐

I understand that I can withdraw my data at any point during data collection and up to two weeks after my final session ☐

Name: ________________________

Signature: _____________________

Date: _________________________

Consent form – learners (control group)

PARTICIPANT CONSENT FORM

Project title: Processing of syntax in L2 English

Researchers: Sophie Thompson and Emma Marsden

Come and experience an English learning experiment! Improve your English and experience an Education/TESOL experiment!

If you take part in this study, you will:

get £20 for taking part in all sessions
get a free (optional) grammar session which will improve your English
get insight into Education/TESOL related experiments

The purpose of the project is to investigate how native speakers of English and learners process syntax.

What this involves

Taking part will involve 3 sessions, each last one hour.

In each session you will complete an eye-tracking task, a speaking task, a writing task and a grammar judgement task. (1 hour)

The exact dates of these sessions will be arranged at times convenient for you.

The sessions will take place approximately as follows:

1st session (in November).

2nd session will be about 3 weeks later (in December).

3rd session will be about 6 weeks after the 2nd session (in January).

If you wish, you can come back for a fourth session that will provide a ***free grammar training session*** *that will help you will the English you use in your studies and every day. Please just let the researcher know if you would like to do this. This session would be in September / October at a time convenient to you.*

Anonymity

The data you provide will be stored by code number. Any information that identifies you will be stored separately from the data and will be destroyed 2 weeks after the end of data collection. Before this time you will be able to withdraw your data at any time. After the end of data collection, and the identifiable data is destroyed, it will no longer be possible to withdraw your data.

Storing and using your data

Data will be stored in secure filing cabinets and on a password protected computer. The anonymous data will be kept for an indefinite period. The data may be used for future analysis, and it may be made available via the internet for other researchers to see and use for research or training purposes, but participants will not be identified individually – only the results of the language tasks will be available (i.e. numbers and reaction times).

Information about confidentiality

The data that we collect (your responses) may be used *anonymously* in different ways. Please indicate on the consent form enclosed/attached with a ☑ if you are happy for this anonymised data to be used in the ways listed.

We hope that you will agree to take part!

If you have any questions about the study that you would like to ask before giving consent or after the data collection, please feel free to contact Sophie Thompson by email sophie.thompson@york.ac.uk or the Chair of Ethics Committee via email education-research-administrator@york.ac.uk

If you are happy to participate, please complete and sign the form attached and leave on the desk in front of your computer.

Please keep this information sheet for your own records.

Thank you for taking the time to read this information.

Yours sincerely

Sophie Thompson (Department of Education)

Dr. Emma Marsden (Department of Education)

English grammar training study consent Form

Please tick each box if you are happy to take part in this research.

I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve me taking part as described above. ☐

I understand that the purpose of the research is to investigate the effectiveness of syntax processing instruction ☐

I understand that data will be stored securely on a password protected computer and only Sophie Thompson and Emma Marsden will have access to any identifiable data. I understand that my identity will be protected by use of a code and no others will be able to recognise that I participated in the study. ☐

I understand that my data will not be identifiable after the final session ☐

This anonymous data may be used:

in publications that are mainly read by university academics and language teachers ☐

in presentations that are mainly attended by university academics and language teacher ☐

in publications that are mainly read by teachers ☐

I understand that Sophie Thompson will keep the file linking my name to the data for up to two weeks after data collection, after which it will be destroyed ☐

I understand that the anonymised data will be kept indefinitely as it may be made freely available online and it could be used for future analysis or research training ☐

I understand that I can withdraw my data at any point during data collection and up to two weeks after my final session ☐

Name: ______________________

Signature: ____________________

Date: _______________________

Consent form – native speakers

PARTICIPANT CONSENT FORM

Project title: Processing syntax in English

Researchers: Sophie Thompson and Emma Marsden

The purpose of the project is to 1) investigate how native speakers of English and learners process syntax, and 2) investigate the extent which instruction can facilitate processing of syntax.

What this involves and benefits

Taking part will involve 1 session of about 1 hour 15 minutes in total in which you will complete an eye-tracking task, two speaking tasks, a writing task, and a grammar judgement task. The exact dates of the session will be arranged at a time convenient for you.

Anonymity

The data you provide will be stored by code number. Any information that identifies you will be stored separately from the data and will be destroyed 4 weeks after the end of data collection. Before this time you will be able to withdraw your data at any time. After the end of data collection, and the identifiable data is destroyed, it will no longer be possible to withdraw your data.

Storing and using your data

Data will be stored in secure filing cabinets and on a password protected computer. The anonymous data will be kept for an indefinite period. The data may be used for future analysis, and it may be made publicly available via the internet for other researchers to see and use for research or training purposes, but participants will not be identified individually – only the results of the language tasks will be available (i.e., numbers and reaction times).

Information about confidentiality

The data that we collect (your responses) may be used *anonymously* in different ways. Please indicate on the consent form enclosed/attached with a ☑ if you are happy for this anonymised data to be used in the ways listed.

We hope that you will agree to take part!

If you have any questions about the study that you would like to ask before giving consent or after the data collection, please feel free to contact Sophie Thompson by email sophie.thompson@york.ac.uk or the Chair of Ethics Committee via email education-research-administrator@york.ac.uk

If you are happy to participate, please complete and sign the form attached and leave on the desk in front of your computer. Thank you for taking the time to read this information.

Yours sincerely

Sophie Thompson (Department of Education)

Dr. Emma Marsden (Department of Education)

English grammar training study

Consent Form

Please tick each box if you are happy to take part in this research.

I confirm that I have read and understood the information given to me about the above named research project and I understand that this will involve me taking part as described above. ☐

I understand that the purpose of the research is to investigate the effectiveness of syntax processing instruction ☐

I understand that data will be stored securely on a password protected computer and only Sophie Thompson and Emma Marsden will have access to any identifiable data. I understand that my identity will be protected by use of a code and no others will be able to recognise that I participated in the study. ☐

I understand that my data will not be identifiable after the final session ☐

This anonymous data may be used:

in publications that are mainly read by university academics and language teachers ☐

in presentations that are mainly attended by university academics and language teacher ☐

in publications that are mainly read by teachers ☐

I understand that Sophie Thompson will keep the file linking my name to the data for up to two weeks after data collection, after which it will be destroyed ☐

I understand that the anonymised data will be kept indefinitely as it may be made freely available online and it could be used for future analysis or research training ☐

I understand that I can withdraw my data at any point during data collection and up to two weeks after my final session ☐

Name: ________________________

Signature: _____________________

Date: _________________________

Appendix 3 – Learner information and proficiency questionnaire

Q1 The following questions are about your **personal details** and **educational background**.
Please answer all of the questions.

Q2 How old are you?

- ○ 18-25
- ○ 26-35
- ○ 36-45
- ○ 46-55
- ○ 56-65
- ○ 66-75
- ○ 76 +

Q3 Which of the following qualifications do you have? (you can select multiple)

- ☐ Bachelor's degree
- ☐ Diploma
- ☐ Master's degree
- ☐ PhD
- ☐ Other ____

Q4 What subject(s) did you study at Bachelor's degree?

Q5 What subject(s) did you study at Diploma?

Q6 What subject(s) did you study at Master's degree?

Q7 What subject(s) did you study at PhD?Q168 What subject(s) did you study in your other qualification(s)?

Q8 How long have you been in the UK?

- ○ less than 1 month
- ○ less than 3 months
- ○ less than 6 months
- ○ 6 months - 1 year
- ○ 1 - 2 years
- ○ 2 - 3 years
- ○ 3 - 4 years
- ○ 4 - 5 years
- ○ more than 5 years

Q9 Did you grow up in a bilingual household (i.e. more than one language used at home)?

- ○ Yes
- ○ No

Q10 Do you speak any languages other than Chinese and English?

- ○ Yes, I am fluent in another language
- ○ Yes, but not fluently
- ○ No

Q11 Which languages?

____

Q12 How many years have you studied English?

____

Q13 How would you describe your English level?

- ○ Beginner
- ○ Pre-intermediate
- ○ Intermediate
- ○ Upper intermediate
- ○ Advanced
- ○ Proficient

Q14 How would you rate each of these skills in English from 1 - 4? (1 = your best skill, 4 = your worst skill)

Type a number from 1 - 4 in the box

_______ Speaking
_______ Writing
_______ Listening
_______ Reading

Q15 Do you have an IELTS certificate?

○ Yes

○ No

Q16 What is your IELTS score?

________ Reading
________ Writing
________ Listening
________ Speaking

Q17 What English qualification do you have?

Please provide the name of the certificate and the grade e.g. TOEFL score 90

____

____

____

____

____

Q18 Have you ever taught English (e.g. language or linguistics, or as a foreign/second language)?

- Yes
- No

Q19 Do you have a formal English teaching qualification? (e.g. CELTA, DELTA, PGCE, PGDE)

- If yes, what qualification? ____
- No

Q20 What is/was your parent(s) job?

- Enter job(s) or sector(s) ____
- Unknown

Q21 Did either of your parents go to university?

- Yes
- No

Appendix 4 – C-test

Passage 1: Test passage "Public Alert"

(Based on the reading material "Police Description" in *Meanings into Words* by Dough, Jones & Mitchell. Cambridge University Press, 1984, p. 16)

Police are looking for a man in connection with this morning's bank robbery in Hong Kong. It is known that the sus____[1] is a man in his ea____[2] thirties, is lightly built, and i____[3] about five feet eight inches ta____[4]. He has small eyes a____[5] a pale complexion with shoulder len____[6] brown hair. He is well dre____[7], wears a gold ring on h____[8] left hand, and speaks wi____[9] a British accent. Police believe h____[10] is still carrying the gun us____[11] in the robbery, and members o____[12] the public are warned not t____[13] approach him but instead to not____[14] the police immediately if he is sig____[15]. Extreme caution is urged in approaching the suspect.

Passage 2: Test passage "Advertisement"

(Based on the reading material "The Ultimate Advertising Medium" in *Academically Speaking* by Kayfetz & Spice, Hineley & Hineley, 1987, p. 109)

Radio remains a vital force in advertising, but television dominates the media world today. It is only natural that television has bec____[16] the dominant advertising medium as we____[17]. An important lesson that was fi____[18] learned about advertising on radio w____[19] applicable to television also; in a mar____[20] flooded with numerous products, the fo____[21] of the ad was a____[22] least as important as the con____[23]. When advertising on television began, i____[24] was a challenge since adver____[25] could now picture the product a____[26] well as describe it in wo____[27]. Cigarette commercials in the m____[28]-1950s showed scene after scene o____[29] spring fields. Clearly the mes____[30] was that smoking is a healthy, fresh and clean experience. How times have changed!

Passage 3: Test passage “Space Shuttle”

(Based on the reading material “The Shuttle and Beyond” in *Meanings into Words* by Dough, Jones & Mitchell, Cambridge University Press, 1984, p. 140)

The development of the space shuttle has dramatically reduced the cost of sending loads into space. The shu____[31] is a reusable type o____[32] space craft which takes o____[33] from the earth like a roc[34], and lands like an air___[35]. It can transport not on____[36] its own crew, but al____[37] passengers, and has a hu____[38] cargo-hold which is cap____[39] of carrying large satellites o__[40] a space laboratory. It i____[41] difficult to imagine the imm____[42] opportunities that have been created by the shu____[43]. One of the great advan__[44] of having a reusable sp____[45] vehicle is that it c____[46] take one load after ano____[47] into orbit. Very large sp____[48] stations cannot be laun____[49] in their complete form dire____[50] from the earth, but they can be built piece by piece in space. The space shuttle has been used as a general workhorse for the past thirty years and it is scheduled to be retired from service in 2011 after 135 launches.

Appendix 5 – Exploratory study vocabulary test items

Nouns	**Verbs**
child	help
process	message
woman	walk
computer	make
noise	follow
shop	cut
people	give
grass	hit
student	throw
company	show
boy	eat
paper	open
film	draw
fish	deliver
window	fire
house	do
man	pour
bottles	transport
baby	carry
box	tell
carrot	take
story	leave
camp	send
girl	change
phone	teach
stone	pay
taxes	provide
walls	wash
building	hunt
car	knock off
rabbit	present
vase	frighten
children	push
space shuttle	chase
teacher	hug
internet	kiss
alarm	destroy
machine	ruin
police	cover

mower
state
photocopier
projector
wind
van
boss
garage
plane
radio
satellite
heat
country
foundations
trees
carwash
wolf
ball
professor
email
rocket
horse
crocodile
dog
bird
bear
boat
bike
tank
machine gun
wine
cake
chocolate
nuts
cigarette
fire
train
water
plants
homework
cat
chicken
kittens
news
parasol
shade

light
stop
repair
ride
break
improve
annoy
find

Appendix 6 – Exploratory picture decision task items

Critical items	Distractors	Fillers
The man is told by the woman	The woman tells the man	The boy does the homework
The dog is frightened by the horse	The horse frightens the dog	The woman is building the house
The crocodile is eaten by the man	The man eats the crocodile	The fox ate the chicken
The man is paid by the woman	The woman pays the man	The students have finished the exam
The boy is pushed by the girl	The girl pushes the boy	The dog chases the cat
The dog is chased by the bird	The bird chases the dog	The woman is driving the car
The boy is hugged by the bear	The bear hugs the boy	The boy made the cake
The boy is hit by the ball	The ball hits the boy	The fish have swum a long way
The baby is kissed by the man	The man kisses the baby	The man puts on a hat
The child is carried by the dog	The dog carries the child	The woman is planting some flowers
The computer is messaged by the phone	The phone messages the computer	The child drank the milk
The boat is transported by the car	The car transports the boat	The people have climbed the mountain
The car is hit by the bike	The bike hits the car	The oven bakes the bread
The tree is hit by the car	The car hits the tree	The boat is transporting the bananas
The tank is destroyed by the machine gun	The machine gun destroys the tank	The film scared the audience
The wine is ruined by the cake	The cake ruins the wine	The company makes shoes
The chocolate is covered by the nuts	The nuts cover the chocolate	The newspaper reports the news
The cigarette is lit by the fire	The fire lights the cigarette	The radio is playing music
The train is stopped by the car	The car stops the train	The shop sold books
The water is covered by the plants	The plants cover the water	The bank has made a profit

Appendix 7 – Exploratory study JT items

Critical items

The child is being helped by the teacher
The process being helped by the internet
The woman being sent a message by the man
The computer is being send a message by the phone
The dog is being walked by the boy
The noise is being make by the alarm
The woman is being make dinner by the man
The shop is being made chocolate by the machine
The person is being followed by the police
The grass is being cutting by the mower
The student is being giving homework by the teacher
The company is being given money by the state
The boy has been hit by the girl
The paper has eaten by the photocopier
The boy has thrown the ball by the girl
The film has been showed to the audience by the projector
The fish has been caught by the children
The window has been opened by the wind
The child has been draw a cartoon by the man
The building has been provided protection by the trees
The man has been fired by the boss
The repair has been doing by the garage.
The child has been pouring milk by the woman
The house has been delivered groceries by the van
The baby was carried by the man
The box carried by the plane
The baby taken to the cot by the woman
The camp was left supplies by the plane
The man was change by the woman
The stone was changed by the heat
The woman was shown the picture by the man
The wall was give strength by the foundations
The dog was washed by the boy
The car was washing by the carwash
The student was presenting an award by the professor
The computer was given a virus by the email
The dog eaten the rabbit
The radio has broken the news
The boy has given the girl flowers
The parasol given the table shade
The woman has teach the man
The boy has sent the girl a letter
The satellite has send the phone a message
The wolf has been hunt the rabbit
The washing machine has been washing the clothes
The teacher has been giving the students tests
The rocket has been carry the space shuttle into space

Appendix 8 – Intervention items

Noun focused – session one

Task 1	**Task 2**	**Task 3**	**Task 4**
The koala is watching	The monkey is seen	The ghost is scared by the fly	The mouse is frightened by the cat
The child is transported	The student is showing	The bird is asking the cat	The child is helping the woman
The man is eating	The girl is sent	The crocodile is watched by the man	The boy is asked by the girl
The glass is filled	The man is feeding	The cat is paying the dog	The child is kissing the father
The dog is following	The match is lit	The woman is cooked for by the man	The gun is destroyed by the plane
The bag is hitting	The book is holding	The satellite is messaging the watch	The pizza is ruining the ice cream
The boat is transporting	The robot is made	The tree is hit by the car	The yoghurt is covered by the fruit
The ball is chasing	The fire is burning	The paper is lighting the candle	The phone is asked by the man
The boy is filmed	The plant is hidden	The phone is shown by the girl	The boy is following the cheese
The bear is hugged	The man is watching	The machine is fixing the woman	The boat is stopped by the woman
The phone is messaged	The cart is pulled	The baby is carried by the pram	The man is pushing the bike
The bag is carried	The train is taking	The camera is seeing the cat	The person is cleaned by the bath
The car is chased	The cat is asked	The cat is held by the basket	The boy is finding the plane
The boy is asked	The cat is following	The boy is taking the boat	The man is hit by the wardrobe
The lion is asking	The man is stopped	The boy is hidden by the box	The fish is seeing the camera
The koala is watched	The cat is showing	The sheep is pulling the truck	The mouse is frightening the cat
The child is transporting	The monkey is seeing	The ghost is scaring the fly	The child is helped by the woman
The man is eaten	The student is shown	The bird is asked by the cat	The boy is asking the girl
The glass is filling	The girl is sending	The crocodile is watching the man	The child is kissed by the father
The dog is followed	The man is fed	The cat is paid by the dog	The gun is destroying the plane
The bag is hit	The match is lighting	The woman is cooking for the man	The pizza is ruined by the ice cream
The boat is tranported	The book is held	The satellite is messaged by the watch	The yoghurt is covering the fruit
The ball is chased	The robot is making	The tree is hitting the car	The phone is asking the man
The boy is filming	The fire is burnt	The paper is lit by the candle	The boy is followed by the cheese
The bear is hugging	The plant is hiding	The phone is showing the girl	The boat is stopping the woman
The phone is messaging	The man is watched	The machine is fixed by the woman	The man is pushed by the bike

The bag is carrying
The car is chasing
The boy is asking
The lion is asked

The cart is pulling
The train is taken
The cat is asking
The cat is followed
The man is stopping
The cat is shown

The baby is carrying the pram
The camera is seen by the cat
The cat is holding the basket
The boy is taken by the boat
The boy is hiding the box
The sheep is pulled by the truck

The person is cleaning the bath
The boy is found by the plane
The man is hitting the wardrobe
The fish is seen by the camera
The train is stopped by the bike
The train is stopping the bike

Noun focused – session two

Task 1
The elephant is asked
The boy is hugged
The lion is watched
The girl is asked
The bottle is filled
The satellite is messaged
The bike is transported
The bike is chased
The woman is hit
The boy is followed
The dog is chased
The cat is carried
The plant is eaten
The tablet is filmed
The basket is covered
The plane is transported
The elephant is asking

Task 2
The elephant is seen
The teacher is shown
The boy is sent
The child is fed
The candle is lit
The shelf is held
The machine is made
The boy is burnt
The man is hidden
The satellite is watched
The horse is pulled
The child is taken
The phone is asked
The toy is followed
The rock is stopped
The picture is shown
The elephant is seeing

Task 3
The ghost is scaring the fly
The bird is asked by the cat
The crocodile is watching the man
The cat is paid by the dog
The woman is cooking for the man
The tree is hit by the car
The paper is lit by the candle
The robot is fixed by the woman
The child is carrying the pram
The camera is seen by the cat
The cat is holding the basket
The boy is taken by the boat
The boy is hiding the box
The satellite is messaged by the watch
The phone is showing the girl
The sheep is pulled by the truck
The fly is scared by the ghost

Task 4
The mouse is frightened by the cat
The child is helping the woman
The boy is asked by the girl
The child is kissing the father
The gun is destroyed by the plane
The pizza is ruining the ice cream
The yoghurt is covered by the fruit
The phone is asked by the man
The boy is following the cheese
The boat is stopped by the woman
The man is pushing the bike
The person is cleaned by the bath
The boy is finding the plane
The man is hit by the wardrobe
The fish is seeing the camera
The cat is frightening the mouse
The woman is helped by the child

The boy is hugging
The lion is watching
The girl is asking
The bottle is filling
The satellite is messaging
The bike is transporting
The bike is chasing
The woman is hitting
The boy is following
The dog is chasing
The cat is carrying
The plant is eating
The tablet is filming
The basket is covering
The plane is transporting

The teacher is showing
The boy is sending
The child is feeding
The candle is lighting
The shelf is holding
The machine is making
The boy is burning
The man is hiding
The satellite is watching
The horse is pulling
The child is taking
The phone is asking
The toy is following
The rock is stopping
The picture is showing

The cat is asking the bird
The man is watched by crocodile
The dog is paying the cat
The man is cooked for by the woman
The car is hitting the tree
The candle is lighting the paper
The woman is fixing the robot
The pram is carried by the child
The cat is seeing the camera
The basket is held by the cat
The boat is taking the boy
The box is hidden by the boy
The watch is messaging the satellite
The girl is shown by the phone
The truck is pulling the sheep

The girl is asking the boy
The father is kissed by the child
The plane is destroying the gun
The ice cream is ruined by the pizza
The fruit is covering the yoghurt
The man is asking the phone
The cheese is followed by the boy
The woman is stopping the boat
The bike is pushed by the man
The bath is cleaning the person
The plane is found by the boy
The wardrobe is hitting the man
The camera is seen by the fish
The train is stopped by the bike
The bike is stopping the train

Cue focused – session one

Task 1
The koala is watching
The child is transported
The man is eating
The boy is filmed
The koala is watched
The child is transporting
The man is eaten
The boy is filming
The ball is chasing
The bear is hugged
The phone is messaged
The cat is covered

Task 2
The monkey is seen
The student is showing
The girl is sent
The man is feeding
The match is lit
The book is holding
The robot is made
The fire is burning
The plant is hidden
The man is watching
The cart is pulled
The train is taking

Task 3
The ghost is scared by the fly
The bird is asking the cat
The crocodile is watched by the man
The cat is paying the dog
The woman is cooked for by the man
The satellite is messaging the watch
The tree is hit by the car
The paper is lighting the candle
The phone is shown by the girl
The machine is fixing the woman
The baby is carried by the pram
The camera is seeing the cat

Task 4
The mouse is frightened by the cat
The child is helping the woman
The boy is asked by the girl
The child is kissing the father
The gun is destroyed by the plane
The pizza is ruining the ice cream
The yoghurt is covered by the fruit
The phone is asked by the man
The cheese is following the boy
The boat is stopped by the woman
The bike is pushing the man
The person is cleaned by the bath

The ball is chased
The bear is hugging
The phone is messaging
The cat is covering
The glass is filling
The dog is followed
The boat is transporting
The bag is carried
The glass is filled
The dog is following
The boat is transported
The bag is carrying
The bag is hitting
The car is chased
The boy is asked
The lion is asking
The bag is hit
The car is chasing
The boy is asking
The lion is asked

The cat is asked
The cat is following
The man is stopped
The cat is showing
The monkey is seeing
The student is shown
The girl is sending
The man is fed
The match is lighting
The book is held
The robot is making
The fire is burnt
The plant is hiding
The man is watched
The cart is pulling
The train is taken
The cat is asking
The cat is followed
The man is stopping
The cat is shown

The cat is held by the basket
The boy is taking the boat
The boy is hidden by the box
The sheep is pulling the truck
The ghost is scaring the fly
The bird is asked by the cat
The crocodile is watching the man
The cat is paid by the dog
The woman is cooking for the man
The satellite is messaged by the watch
The tree is hitting the car
The paper is lit by the candle
The phone is showing the girl
The machine is fixed by the woman
The baby is carrying the pram
The camera is seen by the cat
The cat is holding the basket
The boy is taken by the boat
The boy is hiding the box
The sheep is pulled by the truck

The boy is finding the plane
The man is hit by the wardrobe
The fish is seeing the camera
The mouse is frightening the cat
The child is helped by the woman
The boy is asking the girl
The child is kissed by the father
The gun is destroying the plane
The pizza is ruined by the ice cream
The yoghurt is covering the fruit
The phone is asking the man
The cheese is followed by the boy
The boat is stopping the woman
The bike is pushed by the man
The person is cleaning the bath
The boy is found by the plane
The man is hitting the wardrobe
The fish is seen by the camera
The train is stopped by the bike
The train is stopping the bike

Cue focused – session two

Task 1
The elephant is seen
The teacher is shown
The boy is sent
The child is fed
The elephant is seeing
The teacher is showing
The boy is sending
The child is feeding
The phone is asked

Task 2
The elephant is asked
The boy is hugged
The lion is watched
The girl is asked
The bottle is filled
The satellite is messaged
The bike is transported
The bike is chased
The woman is hit

Task 3
The woman is helped by the child
The cat is frightened by the mouse
The girl is asked by the boy
The father is kissed by the child
The plane is destroyed by the gun
The ice cream is ruined by the pizza
The fruit is covered by the yoghurt
The bike is stopped by the train
The man is asked by the phone

Task 4
The fly is scared by the ghost
The cat is asked by the bird
The man is watched by crocodile
The dog is paid by the cat
The man is cooked by the woman
The watch is messaged by the satellite
The car is hit by the tree
The candle is lit by the paper
The girl is shown by the phone

The toy is followed
The rock is stopped
The picture is shown
The phone is asking
The toy is following
The rock is stopping
The picture is showing
The candle is lit
The shelf is held
The machine is made
The boy is burnt
The candle is lighting
The shelf is holding
The machine is making
The boy is burning
The man is hidden
The satellite is watched
The horse is pulled
The child is taken
The man is hiding
The satellite is watching
The horse is pulling
The child is taking

The boy is followed
The dog is chased
The cat is carried
The plant is eaten
The tablet is filmed
The basket is covered
The plane is transported
The elephant is asking
The boy is hugging
The lion is watching
The girl is asking
The bottle is filling
The satellite is messaging
The bike is transporting
The bike is chasing
The woman is hitting
The boy is following
The dog is chasing
The cat is carrying
The plant is eating
The tablet is filming
The basket is covering
The plane is transporting

The cheese is followed by the boy
The woman is stopped by the boat
The bike is pushed by the man
The bath is cleaned by the person
The plane is found by the boy
The wardrobe is hit by the man
The camera is seen by the fish
The child is helping the woman
The mouse is frightening the cat
The girl is asking the boy
The child is kissing the father
The gun is destroying the plane
The pizza is ruining the ice cream
The yoghurt is covering the fruit
The train is stopping the bike
The phone is asking the man
The boy is following the cheese
The boat is stopping the woman
The man is pushing the bike
The person is cleaning the bath
The boy is finding the plane
The man is hitting the wardrobe
The fish is seeing the camera

The woman is fixed by the robot
The pram is carried by the child
The cat is seen by the camera
The basket is held by the cat
The boat is taken by the boy
The box is hidden by the boy
The truck is pulled by the sheep
The fly is scaring the ghost
The cat is asking the bird
The man is watching the crocodile
The dog is paying the cat
The man is cooking for the woman
The watch is messaging the satellite
The car is hitting the tree
The candle is lighting the paper
The girl is showing the phone
The woman is fixing the robot
The pram is carrying the child
The cat is seeing the camera
The basket is holding the cat
The boat is taking the boy
The box is hiding the boy
The truck is pulling the sheep

Appendix 9 – Eye-tracking list logic

Item	V1	V2	V3	V4	V5	V6	Key
1	a	b	c	d	e	f	a
2	b	c	d	e	f	a	b
3	c	d	e	f	a	b	c
4	d	e	f	a	b	c	d
5	e	f	a	b	c	d	e
6	f	a	b	c	d	e	f
7	a	b	c	d	e	f	
8	b	c	d	e	f	a	
9	c	d	e	f	a	b	
10	d	e	f	a	b	c	
11	e	f	a	b	c	d	
12	f	a	b	c	d	e	
13	a	b	c	d	e	f	
14	b	c	d	e	f	a	
15	c	d	e	f	a	b	
16	d	e	f	a	b	c	
17	e	f	a	b	c	d	
18	f	a	b	c	d	e	
19	a	b	c	d	e	f	
20	b	c	d	e	f	a	
21	c	d	e	f	a	b	
22	d	e	f	a	b	c	
23	e	f	a	b	c	d	
24	f	a	b	c	d	e	
25	a	b	c	d	e	f	
26	b	c	d	e	f	a	
27	c	d	e	f	a	b	
28	d	e	f	a	b	c	
29	e	f	a	b	c	d	
30	f	a	b	c	d	e	
31	a	b	c	d	e	f	
32	b	c	d	e	f	a	
33	c	d	e	f	a	b	
34	d	e	f	a	b	c	
35	e	f	a	b	c	d	
36	f	a	b	c	d	e	
37	a	b	c	d	e	f	
38	b	c	d	e	f	a	
39	c	d	e	f	a	b	
40	d	e	f	a	b	c	
41	e	f	a	b	c	d	
42	f	a	b	c	d	e	
43	a	b	c	d	e	f	
44	b	c	d	e	f	a	
45	c	d	e	f	a	b	

46	d	e	f	a	b	c
47	e	f	a	b	c	d
48	f	a	b	c	d	e

Appendix 10 – Eye-tracking items

Item number	A - full passive	B - reduced passive	C - active
1	The boy has been hugged by the bear for a while	The boy has been hugged for a while	The boy has been hugging the bear for a while
2	The man has been seen by the panda over there	The man has been seen over there	The man has been seeing the panda over there
3	The baby has been kissed by the man for a while	The baby has been kissed for a while	The baby has been kissing the man for a while
4	The dog has been frightened by the horse for a while	The dog has been frightened for a while	The dog has been frightening the horse for a while
5	The bird has been heard by the monkey over there	The bird has been heard over there	The bird has been hearing the monkey over there
6	The child has been fed by the woman over there	The child has been fed over there	The child has been feeding the woman over there
7	The wine has been ruined by the cake over there	The wine has been ruined over there	The wine has been ruining the cake over there
8	The cigarette has been lit by the fire over there	The cigarette has been lit over there	The cigarette has been lighting the fire over there
9	The train has been stopped by the car for a while	The train has been stopped for a while	The train has been stopping the car for a while
10	The water has been covered by the plant over there	The water has been covered over there	The plant has been covering the water over there
11	The bus has been chased by the car over there	The bus has been chased over there	The bus has been chasing the car over there
12	The plane has been followed by the helicopter over there	The plane has been followed over there	The plane has been following the helicopter over there
13	The boy has been hidden by the rock for a while	The boy has been hidden for a while	The boy has been hiding the rock for a while
14	The child has been pulled by the car over there	The child has been pulled over there	The child has been pulling the car over there
15	The person has been cleaned by the shower over there	The person has been cleaned over there	The person has been cleaning the shower over there
16	The man has been fixed by the robot over there	The man has been fixed over there	The man has been fixing the robot over there
17	The child has been carried by the bike over there	The child has been carried over there	The child has been carrying the bike over there
18	The bird has been held by the cage for a while	The bird has been held for a while	The bird has been holding the cage for a while
19	The picture has been shown by the girl over there	The picture has been shown over there	The picture has been showing the girl over there
20	The wheel has been chased by the cat over there	The wheel has been chased over there	The wheel has been chasing the cat over there
21	The camera has been seen by the thief over there	The camera has been seen over there	The camera has been seeing the thief over there
22	The camera has been filmed by the boy for a while	The camera has been filmed for a while	The camera has been filming the boy for a while
23	The paper has been burnt by the girl over there	The paper has been burnt over there	The paper has been burning the girl over there
24	The box has been transported by the cat over there	The box has been transported over there	The box has been transporting the cat over there
25	The child is carried by the dog over there	The child is carried over there	The child is carrying the dog over there
26	The boy is pushed by the girl over there	The boy is pushed over there	The boy is pushing the girl over there

27	The man is asked by the woman over there	The man is asked over there	The man is asking the woman over there
28	The crocodile is eaten by the man over there	The crocodile is eaten over there	The crocodile is eating the horse over there
29	The man is helped by the woman for a while	The man is helped for a while	The man is helping the woman for a while
30	The boy is beaten by the girl over there	The boy is beaten over there	The boy is beating the girl over there
31	The computer is messaged by the phone over there	The computer is messaged over there	The computer is messaging the phone over there
32	The boat is transported by the car over there	The boat is transported over there	The boat is transporting the car over there
33	The computer is shown by the TV over there	The computer is shown over there	The computer is showing the TV over there
34	The tree is hit by the car over there	The tree is hit over there	The tree is hitting the car over there
35	The tank is destroyed by the gun over there	The tank is destroyed over there	The tank is destroying the gun over there
36	The car is carried by the bag over there	The car is carried over there	The bag is carrying the car over there
37	The woman is pushed by the door over there	The woman is pushed over there	The woman is pushing the door over there
38	The boy is followed by the truck for a while	The boy is followed for a while	The boy is following the truck for a while
39	The man is found by the helicopter over there	The man is found over there	The man is finding the helicopter over there
40	The boy is hit by the snow over there	The boy is hit over there	The boy is hitting the snow over there
41	The girl is watched by the camera for a while	The girl is watched for a while	The girl is watching the camera for a while
42	The child is taken by the toy over there	The child is taken over there	The child is taking the toy over there
43	The computer is told by the woman over there	The computer is told over there	The computer is telling the woman over there
44	The helicopter is followed by the man over there	The helicopter is followed over there	The helicopter is following the man over there
45	The car is stopped by the girl for a while	The car is stopped for a while	The car is stopping the girl for a while
46	The bag is taken by the dog over there	The bag is taken over there	The bag is taking the dog over there
47	The ball is hit by the girl over there	The ball is hit over there	The ball is hitting the girl over there
48	The donut is eaten by the man over there	The donut is eaten over there	The donut is eating the man over there

Appendix 11 – Eye-tracking fillers and distractors

Distractors

The boy that the girl kisses is over there.
The cat that the dog is beating is over there
The boy that is greeting the girl is happy.
The book that the boy has been putting the box on is over there.
The coffee that the girl pours into the milk is bitter.
The book that the boy is pressing a bag on is over there.
The napkin that the girl has been placing on the plate is clean.
The chair that the boy pushes the table towards is over there.
The butterfly that is chasing the cat is red
The fish that the frog has been scaring is over there.
The elephant that greets the panda is happy
The bear that the gorilla greets is over there
The man that is hearing the boy is tall
The giraffe that the monkey asks is over there.
The son that calls the mother is blond
The boy that is calling the grandpa is over there
The chicken that the cow has been seeing is brown
The man that chases the computer is over there.
The dog that the bird calls is over there.
The fish that is chasing the dolphin is green.
The cat that the dog has been chasing is small.
The dog that the cat beats is over there.
The man that is eating the whale is hungry.
The girl that the boy is kissing is over there.
The girl that the boy has been greeting is happy.
The box that the boy puts on the book is over there.
The milk that the girl is pouring into the coffee is sweet.
The bag that the boy has been pressing on a book is over there.
The plate that the girl places on the napkin is clean.
The table that the boy has been pushing the chair towards is over there.
The cat that has been chasing the butterfly is pink
The frog that scares the fish is over there.
The panda that greets the elephant is big
The gorilla that the bear is greeting is over there
The dolphin that has been chasing the fish is blue.
The boy that the man hears is short
The monkey that is asking the giraffe is over there.
The mother that has been calling the son is blond
The grandpa that has been calling the boy is over there
The cow that sees the chicken is brown

The computer that the man is chasing is over there.
The bird that the dog is calling is over there.

Fillers

The boy does the homework over there
The woman is building the house for a while
The fox ate the chicken over there
The students have finished the exam
The dog chases the cat for a while
The shelves are holding the books over there
The washing machine cleaned the clothes
The boy has finished his dinner over there
The woman is driving the car over there
The boy made the cake over there
The oven is cooking the chicken for a while
The van transported the bananas over there
The film has entertained the audience
The fish have swum a long way
The man puts on a hat over there
The company makes shoes
The woman is planting some flowers for a while
The radio reported the news for a while
The child drank the milk
The people have climbed the mountain
The shop sold books
The bank has made a profit
The radio is playing music for a while
The book tells a story

Appendix 12 – Eye-tracking comprehension questions

Target	CQ	Correct response
The dog that chases the cat is big.	Was there a dog?	Y
The boy does the homework over there	Was there some homework?	Y
The fox ate the chicken over there	Was there a cat?	N
The students have finished the exam	Was there a teacher?	N
The coffee that the girl pours into the milk is bitter.	Was there some wine?	N
The dog chases the cat for a while	Was the dog sleeping?	N
The shelves are holding the books over there	Were the shelves over there?	Y
The butterfly that is chasing the cat is red	Was the butterfly pink?	N
The washing machine cleaned the clothes	Was the washing machine cleaning the clothes?	Y
The elephant that greets the panda is happy	Was there an elephant?	Y
The man that is hearing the boy is tall	Was there a chicken?	N
The boy has finished his dinner over there	Was the boy cooking dinner?	N
The son that calls the mother is blond	Was there a boy?	Y
The woman is driving the car over there	Was the car moving?	Y
The chicken that the cow has been seeing is brown	Was there a elephant?	N
The boy made the cake over there	Was there a pizza?	N
The oven is cooking the chicken for a while	Was the chicken cooked?	Y
The fish that is chasing the dolphin is green.	Was there a dolphin?	Y

The van transported the bananas over there	Was there a van?	Y
The film has entertained the audience	Was the audience laughing?	Y
The fish have swum a long way	Was there a bird?	N
The man that is eating the whale is hungry.	Was there a whale?	Y
The man puts on a hat over there	Was the man over there?	Y
The girl that the boy has been greeting is happy.	Was there a girl?	Y
The company makes shoes	Was the company selling shoes?	Y
The woman is planting some flowers for a while	Were the flowers in the street?	N
The radio reported the news for a while	Was the news in the paper?	N
The plate that the girl places on the napkin is clean.	Was there a plate?	Y
The child drank the milk	Was the child running?	N
The cat that has been chasing the butterfly is pink	Was the butterfly pink?	N
The people have climbed the mountain	Was there a sea?	N
The shop sold books	Were there some shoes?	N
The bank has made a profit	Was there a cafe?	Y
The radio is playing music for a while	Was the music on the radio?	Y
The book tells a story	Was there a book?	Y

Appendix 13 – Eye-tracking instructions

Slide 1:

We will now set up the eye-tracker.

This will take a couple of minutes. Thank you for your patience.

Slide 2:

Listen to the sentence and look at the pictures

Audio: **The banana is yellow**

Slide 3:

IF you see a question....

Press Y for YES
Press N for NO

Was there a banana?

Y

N

Press any key to move to the next screen

Slide 4:

Green means correct

Red means incorrect

Was there a banana?

YES

NO

Press any key to move to the next screen

Slide 5:

Sometimes you will get a question and sometimes you won't.

If you see a strawberry make sure you look at it.

Press any key to continue.

Slide 6.

Every time you see this circle, look at the white part in the middle and press the space bar.

Slide 7.

Here are some practice sentences.

Press any key to start.

Appendix 14 – Eye-tracking studies and task details

Article	Aim	Example stimulus	N. images	Fixation cross? (ms)	Preview of images? (ms)	Sound (ms)	Post-sound (ms)	Data analysis
Andringa & Curcic (2015)	Measure learners' ability to use DOM as a cue to predict either animate or inanimate direct objects.	*Ese edzo forigas al ese kuzo* (The man is feeding the dog).	2	1000ms	1000ms	Sound length	5000 ms time out (picture choice task)	Time bins: Each frame equalled 8.33 ms. For each frame, one or zero was scored, depending on whether the participant was looking at the correct or the incorrect image. Frames in which data was missing from the interest area (i.e. blinks / look aways) were excluded.
Cunnings et al. (2017)	English L2 processing of sentences with subject or object biased clauses and ambiguous and unambiguous pronouns.	*After Peter spoke to Mrs Jones by the till in the shop, he paid for the expensive ice-cream that looked tasty* (unambiguous) *After Peter spoke to Mr Smith by the till in the shop, he paid for the expensive ice-cream that looked tasty* (ambiguous).	5	Yes	1000ms	-	-	Proportion of looks to both the subject and object antecedents in the four conditions in two time windows; pronoun time window and biasing noun time window.
Dijkgraaf et al. (2017)	Does predictive language processing occur to the same extent when bilinguals listen to their native language vs. a non-native language.	*Mary reads a letter.* *Mary steals a letter.*	4	500ms	2200	Sound	-	Time course analysis - probabilities reflect the number of samples of eye-data within a 50 ms time bin in which there was a fixation on the target picture, averaged over subjects and items.

Dussias et al. (2013)	Use of gender to anticipate roles.	*El estudiante estaba dibujando el reloj que vio ayer* (The student was drawing the clock that he saw yesterday)	2	Yes - no duration given	Images plus sound at same time (no timings given)	-[6]	Mouse click to correct image	Analysis on fixations from the beginning of the Spanish article to 1,000 ms from this critical onset. Three-phase regression model, with each phase described by a linear function. The three phases were: (a) the preconvergence phase, (b) the convergence phase, and (c) the postconvergence phase.
Flecken et al. (2011)	Investigation of event construal by early Dutch–German bilinguals, reflected in their oral depiction of everyday events shown in video clips.	NA (video clips of motion events)	Video clip	8000ms	-	6000ms video (no sound)	-	Attention was measured on the basis of the duration for fixations within both AoIs (the time they spent looking at the area of interest), or by means of the time (measured in ms from video clip onset).
Flecken, Weimar, Carroll, Von Stutterheim (2015)	Use of of motion events to investigate difference in how spatial information is conceptualised in French and German L1 and L2.	NA (video clips of motion events)	Video clip	8000ms	-	6000ms video (no sound)	-	Fixations were registered in two pre-defined areas of interest for each stimulus, relating to the entity in motion and the potential endpoint object displayed in the videos. The areas of interest were defined on a frame-by-frame-basis.
Grüter, Lew-Williams, & Fernald (2012)	Gender marking on the determiner in Spanish and anticipation of roles.	*Tenemos que buscar otra* (fem) *Tenemos que buscar otro* (masc) 'We must find another (one).'	2	-	2000ms	3000ms	1000ms	Manually coded for each 33ms frames if the subject was looking left, right, between the pictures, or away. Then analysis was performed on reaction time (RT), the latency to initiate an eye movement toward the target picture.

[6] No information in the article

Hopp (2017)	Object and subject marking in German - case marking information and anticipation of roles. Adaption of Kamide et al. 2003.	*Der Hase frisst gleich den Kohl.* (The hare will soon eat the cabbage)	4	yes	3000ms	Image and sound appear simulta neo-usly	1500ms	Fixations to the agent and the patient in 20ms time windows for 4000 ms. Based on the marking of the onsets of the words in the audio editor, the mean onsets were calculated for the following: NP 1, the verb, the adverb, NP2, offset of NP2.
Hopp (2016)	Gender marking on adjectives and nouns and anticipation of roles	*Wo ist der/die/das gelbe?* (Where is the [masc/fem/neut] yellow?)	4	yes	3000ms	Sound	-	Eye gaze position after the determiner onset was coded every 20 ms for 3,000 ms. RTs of looking towards the target region (Figure 4) were analysed in a mixed linear regression model with the fixed factors Type (same vs. difference) and Test (pretest vs. posttest).
Hopp & Lemmerth (2016)	Gender marking on adjectives and nouns and anticipation of roles	*Wo ist der/die/das gelbe?* (Where is the [masc/fem/neut] yellow?)	4	yes	800ms	Sound	-	Reaction time - the position of the first fixation on the target picture after article onset or adjective offset was coded every 20 ms for 3,000 ms.
Ito, Corley & Pickering (2018)	Cognitive load of semantically predictive or non-predictive verbs in English.	*The lady will fold the scarf.*	4	Yes but no duration given	1000ms	Sound	Mouse click to correct image	The proportion of time spent fixating on target and semantic competitor objects was calculated separately for each 50 ms bin relative to the target noun onset
Kaushanskaya & Marian (2007)	Recognition and interference of a nontarget language (Russian) during production in a target language (English) were tested in Russian-English bilinguals	*YTKA (duck)* (target image – chicken)	1	500ms	NA	NA (no sound)	-	Proportion of word fixations per condition for each participant

Kim et al. (2015)	Investigate how adult L2 English learners make use of grammatical and extra-grammatical information to interpret reflexives and pronouns.	*Look at Goofy. Have Mickey touch a picture of himself.*	3 strips of 3 images (moveable by the learner)		1000ms	-	-	Number of fixations to the subject character in each group of images
Kohlstedt & Mani (2016)	Investigate L1 and L2 speakers' anticipation of upcoming information in a discourse and second, L1 and L2 speakers' ability to infer the meaning of unknown words in a discourse based on the semantic cues provided in spoken language context.	*Tanja besucht sehr gerne ihren Opa.* (Tanja likes visiting her grandpa very much).	4	500ms	2000ms	20000 ms		Proportion of looks to the target or to the competitor during each trial across three windows (three different prime phases).
Lee & Winke (2017)	How young language learners process their responses on and perceive a computer mediated, timed speaking test. children's attentional foci and performance on speaking tasks were investigated.	*What animal do you like? Why?*	Various (from TOEFL primary)		-	-	-	(a) the total fixation duration, which indicates the total amount of time (in seconds) a child looked at an AOI and (b) the number of eye visits (i.e., visit counts), which reveals how many separate glances the child gave to each AOI.

Lew-Williams & Fernald (2007)	Gender marking in Spanish-learning children.	*Encuentra la pelota* “Find the ball”	2		2000ms	3000ms	1000ms	Coded each trial frame by frame over the time course of orienting to the correct picture on same-gender and different-gender trials.
Lew-Williams and Fernald (2010)	Processing of gender marking on articles by L1 and L2 adults.	*Encuentra la pelota* “Find the ball”	2	?	2000ms	3000ms	1000ms	From the acoustic onset of the article on each trial coder rated whether the participants were looking left or right in each 33ms frame. RTs to the target were calculated.
Marian & Spivey (2003a)	Auditory processing of competing lexical items to investigate cross-linguistic effects in English and Russian.	*Speaker* *Spear* *Spichki* (matches)	4	yes	-	-	-	Proportion of eye movements to the competitor item in a competition condition was compared to the proportion of eye movements to a non-overlapping filler item in the same location in the control condition.
Marian & Spivey (2003b)	Auditory processing of competing lexical items to investigate cross-linguistic effects in English and Russian.	*Shovel* *Shark* *Sharik* (balloon)	3	-	-	-		Trials were coded as containing zero or greater-than-zero fixations of the between-language competitor object (if it was present), the within-language competitor object(if it was present), and their associated filler (control).
McDonough et al. (2015)	Investigation whether joint attention through eye gaze was predictive of second language (L2) speakers’ responses to recasts.	World Records trivia task Story telling task	-		-	-	-	Eye movements to the L2er and the RA were recorded rather than to an image or task. Logistic regression model. Predictor variables included prosody (binary), intonation (binary), length of L2 speaker and RA eye gaze (both numeric), and mutual eye gaze (binary).

McDonough et al. (2017)	Investigation whether joint attention through eye gaze was predictive of second language Esperanto (L2) speakers' responses to recasts.	*Filino mordas pomon* Girl eats apple						Eye movements to the L2er and the RA were recorded rather than to an image or task. For each picture identification and picture description item, the total duration of self-initiated looks to the correct image was summed, the length of eye gaze was calculated from the time participants looked at the correct picture while or after hearing the sentence until the moment they gave an answer. For the picture description items, the eye-gaze duration was computed from when participants were told which picture to describe until they produced a sentence.
Mercier, Pivneva, & Titone (2014)	An investigation whether individual differences in inhibitory control relate to bilingual spoken word recognition.	*Click on the …. field*	4	Yes - click on cross to view images	207ms	Mouse click	-	Latency of response (i.e., the mouse clicks to the target picture) and the proportion of fixations to lexical competitor.
Mercier, Pivneva, & Titone (2016)	Investigation whether speaking in one language affects cross- and within-language activation when subsequently switching to a task performed in the same or different language.	*Click on the …. field*	4	Yes - click on cross to view images	207ms	Mouse click	-	Latency of response (i.e., the mouse clicks to the target picture) and the proportion of fixations to lexical competitor.
Mitsugi	L2 learners of Japanese usage of case markers to	*Onnanohito-ga otokonohito-ni hidoku tatak-are-ta-sou-desu*	2	1500ms	-	Sound	2000ms	The statistical analysis was performed on the critical region, which was from 200ms after the offset of the second noun and lasted 1200ms.

(2018)	whether the active or passive voice.	“I heard that the woman was badly hit by the man”						Analysis was on fixations to the active picture scenes in 100-ms bins for the critical-region duration of 1200 ms.
Mitsugi & MacWhinney (2015)	Japanese L1 and L2 processing of sentences containing either the monotransitive or ditransitive constructions and case markers to predict preceding nomimals.	*Gakkou-de majimena gakusei-ga kibishii sensei-ni shizukani tesuto-o watashita.* (At the school, the serious student quietly handed over the exam to the strict teacher).	4	2000ms	-	Sound	2000ms	Fixations to each object in the visual scene in 100-ms bins were analysed, for the overall duration of 800 ms for the prediction period and 600 ms for the critical word period.
Morales, Paolieri, Dussias, Valdés Kroff, Gerfen, & Teresa Bajo (2016)	Investigation of the gender-congruency effect during a spoken-word recognition task using eye movements of Italian-Spanish bilinguals and Spanish monolinguals	*Encuentra la bufanda* (Find the scarf)	2	-		click on image	-	The total proportions of target fixations (relative to distractor and outside areas) across conditions in predefined time windows.
Pozzan & Trueswell (2015)	Do children’s difficulties revising initial sentence processing commitments characterize the L2 parser. Adult L2 speakers of English acted out temporarily	*Put the frog on the napkin onto the box.*	4	yes	-	-	1500ms or mouse click	Proportion of time participants spent looking at the incorrect goal from the onset of the critical prepositional phrase (e.g., “on the napkin”) until disambiguating information became available to the participant (i.e., the onset of “box”) as a function of group, ambiguity, and referential context.

	ambiguous and unambiguous instructions.							
Sekerina & Sauermann (2015)	Investigation of how bilingual heritage Russian–English misinterpret sentences with the universal quantifier every and make quantifier-spreading errors that are attributed to a preference for a match in number between two sets of objects.	*Kazhdyj alligator lezhit v vanne.* (Every alligator lies in a/the bathtub).	1 picture with 5 objects		-	-	Key press	The proportions of looks were analysed in separate time windows, or regions of interest (ROIs).
Sekerina & Trueswell (2011)	Compared how monolingual Russian (Experiment 1) and heritage Russian–English bilingual (Experiment 2) listeners process contrastiveness online in Russian.	*Krasnuju položite zvezdoˇcku . . .* RED put star	5 actual objects (not images)	Yes, but is permane nt	-	-	-	Participants moved objects around a grid. Moment-by-moment record of fixations to the target as the dependent variable throughout the duration of the utterance with 33 ms resolution.
Singh & Mishra (2012)	Examined the inhibitory control of high and low proficient Hindi–English bilinguals	*Laal* “red” *Neela* “blue” *Hara* “green”,	4	1200ms		no sound - ppts had to look to the word	1500ms	Saccadic latency was calculated only for the correct trials (i.e. time lag between the onset of the display and the initiation of a saccade towards the correct colour patch).

	on an oculomotor Stroop task.	*Peela* “yellow”				and colour that matche d		
Suzuki (2017)	ppt listened to a sentence whilst looking at 4 images followed by a yes/no CQ	*Watashi-wa toshokan-de benkyousuru* (I study in the library)	4	-	5500ms	-	-	Eye-movements were analysed as follows: target looks divided by the sum of target looks and competitor looks. CFA and Multi trait-multimethod models.
Suzuki & DeKeyser (2017)	How EK and IK interact in a naturalistic second language Japanese acquisition context.		4					eye movements were analyzed from the onset of the case marker
Tremblay (2011)	Processing of re-syllabified words by native English speakers at three proficiency levels in French and by native French speakers.	*Fameux élan* (Infamous swing)	4	500ms	-	Mouse click	-	Eye movements were analysed as the proportion of fixation in one of the four regions of interest for each 100-millisecond time window, from 0 to 1,000 milliseconds.
Trenkic, Mirkovic & Altmann (2014)	Online comprehension of English articles by speakers of article-lacking Mandarin.	*The pirate will put the cube inside the can* versus *The pirate will put the cube inside a can*	1 image - 6 objects	Yes but no duration given	4000ms	Sound	Mouse click or 2000ms	Cumulative proportions of looks to objects in the scene were calculated, across 25ms windows from the onset of the target noun

Appendix 15 – GJT lists

List 1

Grammatical	Ungrammatical items	Error types
The boy has been hugged by the bear	The man has been see	base form
The baby has been kissing the man	The horse has frightened by the dog	missing be
The monkey has been heard	The woman has been feed the child	base form
The wine has been ruined by the cake	The cigarette has lit	missing be
The train has been stopping the car	The plant has been covering by the water	+ ing
The car has been chased	The helicopter has been followed the plane	+ ing
The boy has been hidden by the rock	The child has been pulling	+ ing
The person has been cleaning the shower	The robot has been fix by the man	base form
The bike has been carried	The cage has holding the bird	missing be
The picture has been shown by the girl	The wheel has been chase	base form
The camera has been seeing the thief	The boy has filmed by the camera	missing be
The girl has been burnt	The cat has been transport the box	base form
The child is carried by the dog	The boy pushed	missing be
The man is asking the woman	The man is eating by the crocodile	+ ing
The woman is helped	The girl is beaten the boy	+ ing
The computer is messaged by the phone	The boat is transporting	+ ing
The computer is showing the TV	The car is hit by the tree	base form
The gun is destroyed	The bag carrying the car	missing be
The woman is pushed by the door	The boy is follow	base form
The man is finding the helicopter	The snow hit by the boy	missing be
The camera is watched	The toy is take the child	base form
The computer is told by the woman	The helicopter followed	missing be
The car is stopping the girl	The dog is taking by the bag	+ ing
The girl is hit	The man is eaten the donut	+ ing

List 2

Grammatical	Ungrammatical	Error
The boy has been hugged	The boy has been hugging	+ ing / + ed
The man has been kissed by the baby	The man has been kissing by the baby	+ ing
The monkey has been hearing the bird	The monkey has been heard the bird	+ ing
The wine has been ruined	The wine has been ruin	base form

The car has been stopped by the train	The car has stopped by the train	missing be
The car has been chasing the bus	The car has been chase the bus	base form
The boy has been hidden	The boy has hidden	missing be
The shower has been cleaned by the person	The shower has been clean by the person	base form
The bike has been carrying the child	The bike has carrying the child	missing be
The picture has been shown	The picture has been showing	+ ing
The thief has been seen by the camera	The thief has been seeing by the camera	+ ing
The girl has been burning the paper	The girl has been burnt the paper	+ ing
The child is carried	The child is carry	base form
The woman is asked by the man	The woman asked by the man	missing be
The woman is helping the man	The woman is help the man	base form
The computer is messaged	The computer messaged	missing be
The TV is shown by the computer	The TV is show by the computer	base form
The gun is destroying the tank	The gun destroying the tank	missing be
The woman is pushed	The woman is pushing	+ ing
The helicopter is found by the man	The helicopter is finding by the man	+ ing
The camera is watching the girl	The camera is watched the girl	+ ing
The computer is told	The computer is tell	base form
The girl is stopped by the car	The girl stopped by the car	missing be
The girl is hitting the ball	The girl is hit the ball	base form

List 3

Grammatical	**Ungrammatical**	**Error**
The panda has been seen by the man	The boy has been hugged the bear	+ ing / + ed

The horse has been frightening the dog	The man has been kissing	+ ing
The child has been fed	The bird has been hearing by the monkey	+ ing
The fire has been lit by the cigarette	The wine has been ruining the cake	base form
The plant has been covering the water	The car has stopping	missing be
The plane has been followed	The bus has been chased by the car	base form
The car has been pulled by the child	The boy has hiding the rock	missing be
The robot has been fixing the man	The shower has been cleaning	base form
The bird has been held	The child has carried by the bike	missing be
The cat has been chased by the wheel	The picture has been shown the girl	+ ing
The boy has been filming the camera	The thief has been seeing	+ ing
The box has been transported	The paper has been burning by the girl	+ ing
The girl is pushed by the boy	The child is carrying the dog	base form
The man is eating the crocodile	The woman asked	missing be
The boy is beaten	The man is helped by the woman	base form
The car is transported by the boat	The computer messaging the phone	missing be
The car is hitting the tree	The TV is showing	base form
The car is carried	The tank destroyed by the gun	missing be
The truck is followed by the boy	The woman is pushed the door	+ ing
The snow is hitting the boy	The helicopter is finding	+ ing
The child is taken	The girl is watching by the camera	+ ing
The man followed by the helicopter	The computer is asking the woman	base form
The dog is taking the bag	The girl stopped	missing be
The donut is eaten	The ball is hit by the girl	base form

List 4

Grammatical	Ungrammatical	Error
The bear has been hugged by the dog	The panda has been see	base form
The man has been kissing the baby	The dog has frightened by the horse	missing be
The bird has been heard	The child has been feed the woman	base form
The cake has been ruined by the wine	The fire has lit	missing be
The car has been stopping the train	The water has been covering by the plant	+ ing
The bus has been chased	The plane has been followed the helicopter	+ ing
The rock has been hidden by the boy	The car has been pulling	+ ing
The shower has been cleaning the person	The man has been fix by the robot	base form
The child has been carried	The bird has holding the cage	missing be
The girl has been shown by the picture	The cat has been chase	base form
The thief has been seeing the camera	The camera has filmed by the boy	missing be
The paper has been burnt	The box has been transport the cat	base form
The dog is carried by the child	The girl pushing the boy	missing be
The woman is asking the man	The crocodile is eating by the man	+ ing
The man is helped	The boy is beaten the girl	+ ing

The phone is messaged by the computer
The TV is showing the computer
The tree is hit by the car
The tank is destroyed
The door is pushed by the woman
The helicopter is finding the man
The girl is watched
The girl is stopping the car
The ball is hit

The car is transporting	+ ing
The bag carrying the car	missing be
The truck is followed	base form
The boy hit by the snow	missing be
The child is taking the toy	base form
The woman is tell by the computer	base form
The man followed	missing be
The bag is taking by the dog	+ ing
The donut is eaten the man	+ ing

List 5

Grammatical

The bear has been hugged
The baby has been kissed by the man
The bird has been hearing the monkey
The cake has been ruined
The train has been stopped by the car
The bus has been chasing the car
The rock has been hidden
The person has been cleaned by the shower
The child has been carrying the bike
The girl has been shown
The camera has been seen by the thief
The paper has been burning the girl
The dog is carried
The man is asked by the woman
The man is helping the woman
The phone is messaged
The tree is hit
The tank is detroying the gun
The door is pushed
The man is found by the helicopter
The girl is watching the camera
The woman is told

Ungrammatical	**Error**
The panda has been see the man	base form
The dog has frightened	missing be
The woman has been feed by the child	base form
The fire has lighting the cigarette	missing be
The water has been covering	+ ing
The helicopter has been following by the plane	+ ing
The car has been pulled the child	+ ing
The man has been fix	base form
The cage has held by the bird	missing be
The cat has been chase the wheel	base form
The camera has filmed	missing be
The cat has been transport by the box	base form
The girl pushing the boy	missing be
The crocodile is eating	+ ing
The girl is beating by the boy	+ ing
The car is tranported the boat	+ ing
The computer is show by the TV	base form
The bag carried by the car	missing be
The truck is follow the boy	base form
The boy hit	missing be
The toy is take by the child	base form
The man following the helicopter	missing be

The car is stopped by the girl

The ball is hitting the girl

The bag is taking	+ ing
The man is eating by the donut	+ ing

List 6

Grammatical

The man has been seen by the panda

The dog has been frightening the horse

The woman has been fed

The cigarette has been lit by the fire

The plant has been covering the water

The helicopter has been followed

The child has been pulled by the car

The man has been fixing the robot

The cage has been held

The wheel has been chased by the cat

The camera has been filming the boy

The cat has been transported

The boy is pushed by the girl

The crocodile is eating the horse

The girl is beaten

The boat is transported by the car

The tree is hitting the car

The bag is carried

The boy is followed by the truck

The boy is hitting the snow

The toy is taken

The helicopter is followed by the man

The bag is taking the dog

The man is eaten

Ungrammatical	**Error**
The bear has been hugged the dog	+ ing / + ed
The baby has been kissing	+ ing
The monkey has been hearing by the bird	+ ing
The cake has been ruin the wine	base form
The train has stopped	missing be
The car has been chase by the bus	base form
The rock has hiding the boy	missing be
The person has been clean	base form
The bike has carried by the child	missing be
The girl has been shown the picture	+ ing
The camera has been seeing	+ ing
The girl has been burning by the paper	+ ing
The dog is carry the child	base form
The man asked	missing be
The woman is help by the man	base form
The phone messaging the computer	missing be
The computer is show	base form
The gun destroyed by the tank	missing be
The door is pushed the woman	+ ing
The man is finding	+ ing
The camera is watching by the girl	+ ing
The woman is tell the computer	base form
The car stopped	missing be
The girl is hitting by the ball	+ ing

Appendix 16 – Oral production test lists

Version 1

Target	Tense cue	Voice (passive or active)
The award is presented by the man	Since this morning	P
The woman is asking the bird	Since this morning	A
The woman has been driving the car	Since this morning	A
The cat is pushing the dog	Now	A
The woman is baking the biscuit	Now	A
The horse is watched by the child	Since this morning	P
The girl has been asked by the teacher	Now	P
The panda has been asked by the fish	Now	P
The zebra has been beating the tiger	Now	A
The man has been helping the child	Now	A
The window has been destroyed by the boy	Now	P
The ball is hit by the boy	Since this morning	P
The girl is feeding the cat	Since this morning	A
The man is seeing the bird	Now	A
The man is holding the flower	Now	A
The barbecue is lit by the man	Since this morning	P
The girl is asked by the boy	Now	P
The dinner is cooked by the man	Now	P
The rabbit has been paid by the dog	Since this morning	P
The photo is shown by the man	Now	P
The man has been drawing the woman	Since this morning	A
The letter has been sent by the man	Since this morning	P
The boy has been messaging the girl	Since this morning	A
The ball has been thrown by the boy	Since this morning	P
The dog is asked by the girl	Now	P
The picture has been drawn by the boy	Now	P
The girl is paying the woman	Now	A
The fox has been chasing the rabbit	Since this morning	A
The cake has been made by the man	Since this morning	P
The doctor has been explaining the problem	Since this morning	A
The man is offering the biscuit	Now	A
The girl has been eating the dinner	Since this morning	A
The man has been pouring the drink	Now	A

The woman has been frightened by the snake	Since this morning	P
The tower is built by the girl	Now	P
The woman is carrying the man	Since this morning	A

Version 2

Target	**Tense cue**	**Voice (passive or active)**
The dog has been paying the rabbit	Since this morning	A
The bird is seen by the man	Now	P
The fish has been telling the panda	Since this morning	A
The girl has been messaged by the boy	Since this morning	P
The drink has been poured by the man	Since this morning	P
The woman has been drawn by the man	Since this morning	P
The child is helped by the man	Since this morning	P
The man has been making the cake	Since this morning	A
The boy has been throwing the ball	Since this morning	A
The dog is pushed by the cat	Now	P
The snake has been frightening the woman	Since this morning	A
The biscuit is offered by the man	Now	P
The dinner has been eaten by the girl	Since this morning	P
The man is carried by the woman	Now	P
The teacher has been asking the girl	Since this morning	A
The man is showing the photo	Now	A
The girl is building the tower	Now	A
The man is lighting the barbecue	Now	A
The man has been sending the letter	Since this morning	A
The boy has been destroying the window	Since this morning	A
The problem has been explained by the doctor	Since this morning	P

The car has been driven by the woman	Since this morning	P
The tiger has been beaten by the zebra	Since this morning	P
The man is cooking the dinner	Now	A
The biscuit is baked by the woman	Now	P
The boy has been drawing a picture	Since this morning	A
The girl is asking the dog	Now	A
The child is watching the horse	Now	A
The man is presenting the award	Now	A
The flower is held by the man	Now	P
The bird is told by the woman	Now	P
The rabbit has been chased by the fox	Since this morning	P
The cat is fed by the girl	Now	P
The boy is asking the girl	Now	A
The woman is paid by the girl	Now	P
The boy is hitting the ball	Now	A

Version 3

Target	**Tense cue**	**Voice (passive or active)**
The girl has been paying the woman	Since this morning	A
The cake is made by the man	Now	P
The man is drawing the woman	Now	A
The man is helping the child	Now	A
The woman has been carrying the man	Since this morning	A
The woman is frightened by the snake	Now	P
The doctor is explaining the problem	Now	A
The letter is sent by the man	Now	P
The photo has been shown by the man	Since this morning	P
The barbecue has been lit by the man	Since this morning	P

The rabbit is paid by the dog	Now	P
The window is destroyed by the boy	Now	P
The panda is asked by the fish	Now	P
The woman is driving the car	Now	A
The horse has been watched by the child	Since this morning	P
The man has been offering the biscuit	Since this morning	A
The girl has been asked by the boy	Since this morning	P
The cat has been pushing the dog	Since this morning	A
The girl is eating the dinner	Now	A
The girl has been feeding the cat	Since this morning	A
The ball has been hit by the boy	Since this morning	P
The ball is thrown by the boy	Now	P
The girl is asked by the teacher	Now	P
The dog has been asked by the girl	Since this morning	P
The tower has been built by the girl	Since this morning	P
The zebra is beating the tiger	Now	A
The man is pouring the drink	Now	A
The man has been seeing the bird	Since this morning	A
The fox is chasing the rabbit	Now	A
The woman has been baking the biscuit	Since this morning	A
The dinner has been cooked by the man	Since this morning	P
The woman has been asking the bird	Since this morning	A
The picture is drawn by the boy	Now	P
The boy is messaging the girl	Now	A
The man has been holding the flower	Since this morning	A
The award has been presented by the man	Since this morning	P

Appendix 17 – Writing test lists

Version 1

Sentence stem	Voice cue	Voice
Since this morning the ball...	Who?	passive
Since this morning the dog...	What?	active
Since this morning the rabbit…	Who?	passive
Since this morning the problem...	Who?	passive
Now the boy...	What?	active
Now the flower...	Who?	passive
Now the child...	What?	active
Now the bird...	Who?	passive
Since this morning the snake...	What?	active
Since this morning the dinner…	Who?	passive
Since this morning the girl…	Who?	passive
Since this morning the boy...	What?	active
Now the man...	What?	active
Since this morning the drink...	Who?	passive
Since this morning the woman…	Who?	passive
Now the girl...	What?	active
Since this morning the child…	Who?	passive
Since this morning the car…	Who?	passive
Since this morning the teacher...	What?	active
Since this morning the fish...	What?	active
Now the man...	What?	active
Since this morning the tiger…	Who?	passive
Now the cat...	Who?	passive
Since this morning the man...	What?	active

Now the girl...	What?	active
Since this morning the man...	What?	active
Now the man...	What?	active
Now the woman...	Who?	passive
Now the biscuit...	Who?	passive
Now the dog…	Who?	passive
Now the bird...	Who?	passive
Now the man...	What?	active
Now the man...	Who?	passive
Since this morning the boy...	What?	active
Now the biscuit...	Who?	passive
Now the boy...	What?	active

Version 2

Sentence stem	**Voice cue**	**Voice**
Since this morning the picture...	Who?	Passive
Since this morning the ball...	Who?	Passive
Since this morning the rabbit...	Who?	Passive
Since this morning the window...	Who?	Passive
Since this morning the cake...	Who?	Passive
Since this morning the letter...	Who?	Passive
Since this morning the girl...	Who?	Passive
Since this morning the woman...	Who?	Passive
Since this morning the panda...	Who?	Passive
Now the tower...	Who?	Passive
Now the barbecue...	Who?	Passive
Now the award...	Who?	Passive
Now the dog...	Who?	Passive

Now the photo...	Who?	Passive
Now the horse...	Who?	Passive
Now the girl...	Who?	Passive
Now the dinner...	Who?	Passive
Now the ball...	Who?	Passive
Since this morning the man...	What?	Active
Since this morning the fox...	What?	Active
Since this morning the man...	What?	Active
Since this morning the doctor...	What?	Active
Since this morning the girl...	What?	Active
Since this morning the woman...	What?	Active
Since this morning the man...	What?	Active
Since this morning the boy...	What?	Active
Since this morning the zebra...	What?	Active
Now the cat...	What?	Active
Now the man...	What?	Active
Now the man...	What?	Active
Now the girl...	What?	Active
Now the woman...	What?	Active
Now the woman...	What?	Active
Now the woman...	What?	Active
Now the girl...	What?	Active
Now the man...	What?	Active

Version 3

Sentence stem	**Voice cue**	**Voice**
Since this morning the dog...	Who?	Passive

Since this morning the biscuit...	Who?	Passive
Since this morning the flower...	Who?	Passive
Since this morning the woman...	Who?	Passive
Since this morning the bird...	Who?	Passive
Since this morning the man...	Who?	Passive
Since this morning the biscuit...	Who?	Passive
Since this morning the cat...	Who?	Passive
Since this morning the bird...	Who?	Passive
Now the drink...	Who?	Passive
Now the rabbit...	Who?	Passive
Now the woman...	Who?	Passive
Now the problem...	Who?	Passive
Now the dinner...	Who?	Passive
Now the car...	Who?	Passive
Now the child...	Who?	Passive
Now the girl...	Who?	Passive
Now the tiger...	Who?	Passive
Since this morning the girl...	What?	Active
Since this morning the man...	What?	Active
Since this morning the man...	What?	Active
Since this morning the girl...	What?	Active
Since this morning the man...	What?	Active
Since this morning the child...	What?	Active
Since this morning the boy...	What?	Active
Since this morning the man...	What?	Active
Since this morning the boy...	What?	Active
Now the boy...	What?	Active
Now the boy...	What?	Active

Now the dog...	What?	Active
Now the boy...	What?	Active
Now the man...	What?	Active
Now the man...	What?	Active
Now the teacher...	What?	Active
Now the snake...	What?	Active
Now the fish...	What?	Active

Appendix 18 – Stepwise multi-level model development for eye-tracking data

Learners – trained construction

Model 1:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time * animacy * voice * reduced | sub) + (voice * time | item)

Model 2:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time * animacy * voice * reduced | sub) + (voice + time | item)

Model 3:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time * animacy * voice + reduced | sub) + (voice + time | item)

Model 4:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time * animacy + voice + reduced | sub) + (voice + time | item)

Model 5:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time * animacy + voice + reduced | sub) + (time | item)

Model 6:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time * animacy + voice | sub) + (time | item)

Model 7:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time * animacy | sub) + (time | item)

Model 8:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time | sub) + (time | item)

Model 9:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (time | sub) + (1 | item)

Model 10:

new = lmer(TW800_720 ~ animacy * time * group * voice * reduced + (1 | sub) + (1 | item)

Appendix 19 – Eye-tracking model results

Trained constructions (present simple) – main effect of group and test time after the verb ending

Signif. codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Time window	Main effect															
	Immediate post-test (compared to pre-test)				Delayed post-test (compared to pre-test)				Noun focused group (compared to test-only group)				Cue focused group (compared to test-only group)			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	-1.800e-01	1.881e-01	-.957	.339	-2.216e-01	1.906e-01	-1.163	.245	1.864e-01	1.921e-01	.970	.332	-2.318e-01	1.915e-01	-1.210	.226
100 – 180	-3.840e-01	1.940e-01	-1.980	.048*	-4.627e-01	1.966e-01	-2.353	.019*	-1.270e-01	2.017e-01	-.630	.529	-2.247e-01	2.007e-01	-1.120	.263
200 – 280	-7.009e-02	1.562e-01	-.449	.654	-2.507e-01	1.583e-01	-1.584	.113	-5.435e-02	1.614e-01	-.337	.736	-1.655e-01	1.607e-01	-1.030	.303
300 – 380	-6.707e-02	1.630e-01	-.411	.680	-2.858e-01	1.652e-01	-1.730	.083.	-5.911e-02	1.684e-01	-.351	.726	-2.087e-01	1.677e-01	-1.244	.214
400 – 480	-3.161e-02	1.696e-01	-.186	.852	-2.697e-01	1.719e-01	-1.569	.117	-3.488e-02	1.747e-01	-.200	.842	-1.809e-01	1.740e-01	-1.039	.299
500 – 580	1.877e-02	1.249e-01	.150	.881	-1.476e-01	1.266e-01	-1.166	.244	6.137e-02	1.298e-01	.473	.636	-1.357e-01	1.292e-01	-1.050	.294
600 – 680	-2.960e-01	1.774e-01	-1.669	.095.	-3.392e-01	1.797e-01	-1.887	.059.	-1.781e-01	1.815e-01	-.981	.327	-3.151e-01	1.810e-01	-1.741	.082.

700 – 780	-3.256e-01	1.775e-01	-1.834	.067.	-3.905e-01	1.799e-01	-2.170	.030*	-1.812e-01	1.813e-01	-.999	.318	-2.952e-01	1.809e-01	-1.632	.103
800 – 880	-1.545e-01	1.783e-01	-.867	.386	-2.940e-01	1.807e-01	-1.627	.104	-2.121e-01	1.815e-01	-1.169	.243	-1.948e-01	1.810e-01	-1.076	.282
900 – 980	-1.560e-01	1.810e-01	-.862	.389	-1.205e-01	1.834e-01	-.657	.511	-1.605e-01	1.849e-01	-.868	.385	-2.268e-01	1.844e-01	-1.230	.219

Interactions – group * time

Time window	Main effect															
	Noun focused group (compared to test-only group) * Immediate post-test (compared to pre-test)				Noun focused group (compared to test-only group) * Delayed post-test (compared to pre-test)				Cue focused group (compared to test-only group) * Immediate post-test (compared to pre-test)				Cue focused group (compared to test-only group) * Delayed post-test (compared to pre-test)			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	-7.011e-02	2.712e-01	-.259	.796	-2.299e-01	2.719e-01	-.845	.398	2.583e-01	2.751e-01	.939	.348	7.079e-02	2.768e-01	.256	.798
100 – 180	2.880e-01	2.797e-01	1.029	.303	1.001e-01	2.804e-01	.357	.721	3.998e-01	2.837e-01	1.409	.159	1.339e-01	2.857e-01	.469	.639
200 – 280	7.427e-02	2.252e-01	.330	.742	7.956e-02	2.257e-01	.352	.724	4.853e-02	2.284e-01	.212	.832	1.712e-01	2.300e-01	.744	.457
300 – 380	5.975e-02	2.351e-01	.254	.799	1.316e-01	2.356e-01	.559	.576	1.236e-01	2.384e-01	.519	.604	1.615e-01	2.400e-01	.673	.501

400 – 480	2.486e-02	2.446e-01	.102	.919	1.488e-01	2.451e-01	.607	.544	9.738e-02	2.480e-01	.393	.695	1.517e-01	2.497e-01	.607	.544
500 – 580	-7.309e-02	1.802e-01	-.406	.685	1.951e-03	1.805e-01	.011	.991	9.346e-02	1.827e-01	.512	.609	2.729e-01	1.841e-01	1.483	.138
600 – 680	1.130e-01	2.558e-01	.442	.659	8.303e-02	2.564e-01	.324	.746	4.221e-01	2.594e-01	1.627	.104	2.709e-01	2.611e-01	1.037	.300
700 – 780	5.917e-02	2.561e-01	.231	.817	2.370e-01	2.566e-01	.924	.356	4.120e-01	2.597e-01	1.586	.113	3.128e-01	2.614e-01	1.197	.232
800 – 880	-1.618e-02	2.572e-01	-.063	.950	3.287e-01	2.577e-01	1.275	.202	2.221e-01	2.608e-01	.852	.394	2.143e-01	2.624e-01	.816	.414
900 – 980	-6.700e-02	2.611e-01	-.257	.798	1.846e-01	2.616e-01	.706	.480	3.386e-01	2.648e-01	1.279	.201	1.700e-01	2.665e-01	.638	.524

Group * test time * voice

Time window	Main effects															
	Noun focused group * Immediate post-test * voice (passive)				Noun focused group * Delayed post-test * voice (passive)				Cue focused group * Immediate post-test * voice (passive)				Cue focused group * Delayed post-test * voice (passive)			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	6.875e-03	4.087e-01	.017	.987	-3.559e-01	4.092e-01	-.870	.385	-2.288e-01	4.209e-01	-.544	.587	1.653e-01	4.216e-01	.392	.695

100 – 180	-3.670e-01	4.214e-01	-.871	.384	-3.030e-01	4.218e-01	-.718	.473	1.051e-02	4.338e-01	.024	.981	-2.200e-01	4.347e-01	-.506	.613
200 – 280	-9.347e-03	3.393e-01	-.028	.978	-2.871e-01	3.396e-01	-.846	.398	1.515e-01	3.493e-01	.434	.664	-3.535e-01	3.500e-01	-1.010	.313
300 – 380	4.383e-02	3.541e-01	.124	.902	-2.138e-01	3.544e-01	-.603	.546	1.126e-01	3.646e-01	.309	.757	-2.063e-01	3.653e-01	-.565	.572
400 – 480	7.194e-02	3.685e-01	.195	.845	-1.921e-01	3.688e-01	-.521	.602	1.626e-01	3.794e-01	.429	.668	-1.219e-01	3.801e-01	-.321	.749
500 – 580	6.561e-02	2.715e-01	.242	.809	-1.680e-01	2.715e-01	-.619	.536	1.315e-03	2.794e-01	.005	.996	-1.954e-01	2.800e-01	-.698	.485
600 – 680	-3.963e-01	3.855e-01	-1.028	.304	-1.909e-01	3.857e-01	-.495	.620	-4.179e-01	3.969e-01	-1.053	.292	-3.335e-01	3.976e-01	-.839	.402
700 – 780	-4.356e-01	3.859e-01	-1.129	.259	-5.838e-01	3.861e-01	-1.512	.131	-5.498e-01	3.973e-01	-1.384	.166	-4.451e-01	3.980e-01	-1.118	.264
800 – 880	-2.224e-01	3.876e-01	-.574	.566	-5.917e-01	3.878e-01	-1.526	.127	-2.422e-01	3.990e-01	-.607	.544	-3.559e-01	3.997e-01	-.890	.373
900 – 980	-9.673e-02	3.934e-01	-.246	.806	-2.849e-01	3.937e-01	-.724	.469	-4.176e-01	4.050e-01	-1.031	.303	-2.802e-01	4.058e-01	-.691	.490

Voice * group

Time window	Main effects							
	Noun focused group * voice (passive)				Cue focused group * voice (passive)			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	-7.230e-02	2.869e-01	-.252	.801	6.796e-02	2.922e-01	.233	.816
100 – 180	1.530e-01	2.957e-01	.517	.605	1.944e-01	3.012e-01	.645	.519
200 – 280	9.982e-02	2.381e-01	.419	.675	2.103e-01	2.425e-01	.867	.386
300 – 380	1.434e-01	2.484e-01	.577	.564	2.096e-01	2.531e-01	.828	.408
400 – 480	1.840e-01	2.586e-01	.712	.477	2.219e-01	2.634e-01	.843	.399
500 – 580	-3.646e-02	1.903e-01	-.192	.848	1.620e-01	1.940e-01	.835	.404
600 – 680	4.579e-01	2.704e-01	1.693	.091 .	5.388e-01	2.755e-01	1.956	.051 .
700 – 780	5.744e-01	2.707e-01	2.122	.034 *	6.033e-01	2.758e-01	2.188	.029 *
800 – 880	4.776e-01	2.719e-01	1.756	.079 .	3.923e-01	2.770e-01	1.416	.157

900 – 980	2.975e-01	2.760e-01	1.078	.281	4.220e-01	2.812e-01	1.501	.133

Reduced passives compared to full passives – main effect of group and test time after the verb ending

Signif. codes: 0 ‘***’ .001 ‘**’ .01 ‘*’ .05 ‘.’ .1 ‘ ’ 1

Time window	Main effect				Interactions							
	Reduced				Reduced * immediate post-test				Reduced * delayed post-test			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	-9.585e-02	2.124e-01	-.451	.652	1.794e-01	2.908e-01	.617	.537	2.024e-01	2.876e-01	.704	.4817
100 – 180	2.700e-02	2.167e-01	.125	.901	3.468e-02	2.997e-01	.116	.908	-8.905e-03	2.965e-01	-.030	.976
200 – 280	2.123e-01	1.737e-01	1.222	.222	-1.621e-01	2.414e-01	-.672	.502	-2.283e-01	2.387e-01	-.956	.339
300 – 380	2.273e-01	1.811e-01	1.255	.210	-1.286e-01	2.519e-01	-.510	.610	-1.441e-01	2.491e-01	-.579	.563
400 – 480	2.662e-01	1.894e-01	1.405	.160	-1.414e-01	2.621e-01	-.540	.590	-9.821e-02	2.593e-01	-.379	.705
500 – 580	1.481e-01	1.372e-01	1.079	.281	-2.631e-01	1.931e-01	-1.363	.173	-1.300e-01	1.909e-01	-.681	.496
600 – 680	3.079e-01	1.972e-01	1.561	.119	-2.367e-01	2.742e-01	-.863	.388	9.099e-03	2.712e-01	.034	.973

700 – 780	3.653e-01	1.975e-01	1.850	.065.	-3.126e-01	2.745e-01	-1.139	.255	-1.970e-01	2.714e-01	-.726	.468
800 – 880	2.572e-01	1.989e-01	1.293	.196	-2.149e-01	2.757e-01	-.780	.436	-1.774e-01	2.727e-01	-.651	.515
900 – 980	2.543e-01	2.019e-01	1.260	.208	-1.939e-01	2.798e-01	-.693	.488	-6.825e-02	2.768e-01	-.247	.805

Interactions – group * reduced * test time

Time window	Interaction															
	Noun focused * reduced * immediate post-test				Noun focused * reduced * delayed post-test				Cue focused * reduced * immediate post-test				Cue focused * reduced * delayed post-test			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	1.364e-01	4.102e-01	.333	.740	3.949e-01	4.099e-01	.963	.335	9.205e-02	4.229e-01	.218	.828	-2.721e-01	4.237e-01	-.642	.521
100 – 180	5.012e-01	4.230e-01	1.185	.236	-1.858e-01	4.225e-01	-.440	.660	-7.641e-02	4.360e-01	-.175	.861	1.607e-02	4.367e-01	.037	.971
200 – 280	4.974e-01	3.406e-01	1.460	.144	2.165e-01	3.402e-01	.637	.524	2.354e-01	3.510e-01	.671	.503	3.383e-01	3.517e-01	.962	.336
300 – 380	3.768e-01	3.555e-01	1.060	.289	1.091e-01	3.550e-01	.307	.759	1.381e-01	3.663e-01	.377	.706	1.089e-02	3.670e-01	.030	.976

400 – 480	5.393e-01	3.699e-01	1.458	.145	1.288e-01	3.694e-01	.349	.727	2.356e-01	3.812e-01	.618	.537	-6.080e-02	3.819e-01	-.159	.874
500 – 580	5.575e-01	2.725e-01	2.046	.041*	4.501e-01	2.720e-01	1.655	.098.	3.334e-01	2.808e-01	1.188	.235	1.530e-02	2.813e-01	.054	.957
600 – 680	7.439e-01	3.869e-01	1.922	.055.	3.658e-02	3.864e-01	.095	.925	5.083e-01	3.988e-01	1.275	.203	-1.934e-01	3.995e-01	-.484	.628
700 – 780	7.259e-01	3.873e-01	1.874	.061.	3.889e-01	3.868e-01	1.006	.315	7.324e-01	3.992e-01	1.835	.067.	1.371e-01	3.999e-01	.343	.732
800 – 880	4.603e-01	3.890e-01	1.183	.237	4.126e-01	3.885e-01	1.062	.288	5.231e-01	4.009e-01	1.305	.192	1.830e-01	4.017e-01	.456	.649
900 – 980	4.019e-01	3.949e-01	1.018	.309	3.427e-01	3.944e-01	.869	.385	5.179e-01	4.070e-01	1.273	.203	4.748e-02	4.077e-01	.116	.907

Animacy – main effects before and after the verb

Signif. codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Time window	Main effect											
	Animate - inanimate				Inanimate - animate				Inanimate – inanimate			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-5.106e-02	1.575e-01	-.324	.746	3.841e-01	1.469e-01	2.614	.009**	1.638e-01	1.434e-01	1.142	.254
-700 - -620	-1.844e-01	1.625e-01	-1.134	.257	2.561e-01	1.516e-01	1.689	.091.	8.143e-02	1.480e-01	.550	.582
-600 - -520	-1.356e-01	1.714e-01	-.791	.429	3.170e-01	1.599e-01	1.982	.048*	1.178e-01	1.561e-01	.754	.451
-500 - -420	-.113	.181	-.624	.533	.332	.169	1.960	.050.	.101	.165	.613	.540
-400 - -320	-.126	.192	-.658	.511	.316	.179	1.763	.078.	.134	.175	.766	.444
-300 - -220	-5.090e-02	2.036e-01	-.250	.803	2.836e-01	1.903e-01	1.490	.136	1.765e-01	1.858e-01	.950	.342
-200 - -120	-1.085e-01	1.406e-01	-.772	.440	2.344e-01	1.312e-01	1.787	.074.	2.746e-02	1.280e-01	.214	.830
100 - -20	1.816e-03	2.103e-01	.009	.993	1.680e-01	1.965e-01	.855	.393	4.329e-02	1.918e-01	.226	.822
0 – 80	-6.404e-02	2.159e-01	-.297	.767	2.176e-01	2.016e-01	1.079	.281	-2.951e-02	1.968e-01	-.150	.881

100 – 180	-3.263e-01	2.204e-01	-1.480	.139	-8.644e-02	2.056e-01	-.420	.674	-2.495e-01	2.008e-01	-1.243	.214
200 – 280	-1.985e-01	1.766e-01	-1.124	.261	1.624e-01	1.648e-01	.985	.324	1.747e-02	1.609e-01	.109	.914
300 – 380	-2.191e-01	1.841e-01	-1.190	.234	1.558e-01	1.718e-01	.907	.365	2.747e-02	1.677e-01	.164	.870
400 – 480	-1.522e-01	1.926e-01	-.790	.430	9.832e-02	1.797e-01	.547	.584	4.600e-02	1.755e-01	.262	.793
500 – 580	-1.455e-01	1.395e-01	-1.042	.297	2.710e-01	1.301e-01	2.083	.037*	5.205e-02	1.270e-01	.410	.682
600 – 680	-2.355e-01	2.004e-01	-1.175	.240	3.536e-03	1.870e-01	.019	.985	-1.967e-01	1.826e-01	-1.077	.282
700 – 780	-2.332e-01	2.007e-01	-1.162	.245	-2.692e-02	1.873e-01	-.144	.886	-3.167e-01	1.829e-01	-1.732	.083.
800 – 880	-7.242e-02	2.020e-01	-.358	.720	8.333e-02	1.886e-01	.442	.659	-2.746e-01	1.842e-01	-1.491	.136
900 – 980	-2.593e-02	2.051e-01	-.126	.899	1.414e-01	1.915e-01	.738	.461	-2.147e-01	1.870e-01	-1.148	.251

Voice * animacy

Time window	Interaction											
	Voice * animate-inanimate				Voice * inanimate-animate				Voice * inanimate-inanimate			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*

-800 - -720	1.932e-01	2.239e-01	.863	.388	-4.857e-01	2.184e-01	-2.224	.026*	-2.740e-01	2.161e-01	-1.268	.205
-700 - -620	2.917e-01	2.310e-01	1.263	.207	-3.688e-01	2.253e-01	-1.637	.102	-2.233e-01	2.230e-01	-1.002	.317
-600 - -520	1.555e-01	2.435e-01	.639	.523	-4.643e-01	2.376e-01	-1.954	.051.	-2.290e-01	2.351e-01	-.974	.330
-500 - -420	.061	.258	.238	.812	-.533	.252	-2.120	.034*	-.110	.249	-.442	.659
-400 - -320	-.039	.273	-.144	.886	-.535	.266	-2.008	.045*	-.309	.263	-1.174	.241
-300 - -220	-1.062e-01	2.895e-01	-.367	.714	-4.113e-01	2.825e-01	-1.456	.146	-4.846e-01	2.795e-01	-1.734	.083.
-200 - -120	1.860e-01	1.998e-01	.931	.352	-3.217e-01	1.949e-01	-1.651	.099.	-1.568e-01	1.928e-01	-.813	.416
100 - -20	5.543e-02	2.990e-01	. 185	.853	-1.632e-01	2.918e-01	-.559	.576	-1.908e-01	2.887e-01	-.661	.509
0 – 80	9.379e-02	3.069e-01	.306	.760	-1.488e-01	2.994e-01	-.497	.619	-1.241e-01	2.962e-01	-.419	.675
100 – 180	4.789e-01	3.131e-01	1.530	.126	2.406e-01	3.054e-01	.788	.431	3.633e-02	3.022e-01	.120	.904
200 – 280	3.987e-01	2.510e-01	1.588	.112	-1.464e-01	2.448e-01	-.598	.550	-8.336e-02	2.423e-01	-.344	.731
300 – 380	4.102e-01	2.617e-01	1.567	.117	-1.164e-01	2.552e-01	-.456	.648	-1.194e-01	2.525e-01	-.473	.636
400 – 480	3.800e-01	2.737e-01	1.388	.165	-1.878e-02	2.670e-01	-.070	.944	-1.541e-01	2.642e-01	-.583	.560
500 – 580	2.574e-01	1.982e-01	1.299	.194	-3.761e-01	1.933e-01	-1.945	.052.	-1.640e-01	1.913e-01	-.857	.391
600 – 680	3.700e-01	2.848e-01	1.299	.194	7.627e-02	2.779e-01	.274	.784	1.977e-01	2.749e-01	.719	.472

700 – 780	4.555e-01	2.852e-01	1.597	.110	1.215e-01	2.783e-01	.437	.663	3.732e-01	2.753e-01	1.356	.175
800 – 880	2.546e-01	2.872e-01	.886	.376	-1.501e-01	2.802e-01	-.536	.592	1.378e-01	2.773e-01	.497	.619
900 – 980	1.823e-01	2.916e-01	.625	.533	-1.778e-01	2.845e-01	-.625	.532	-4.559e-03	2.815e-01	-.016	.987

Voice * animacy * group

Time window	Interaction															
	Voice * animate-inanimate * noun focused				Voice * animate-inanimate * cue focused				Voice * inanimate-animate * noun focused				Voice * inanimate-animate * cue focused			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	2.606e-02	3.145e-01	.083	.934	-2.638e-01	3.249e-01	-.812	.417	7.588e-01	3.071e-01	2.471	.014*	3.820e-01	3.126e-01	1.222	.222
-700 - -620	8.691e-03	3.234e-01	.027	.979	-3.797e-01	3.341e-01	-1.136	.256	7.723e-01	3.158e-01	2.445	.015*	2.743e-01	3.215e-01	.853	.394
-600 - -520	8.462e-02	3.386e-01	.250	.803	-2.048e-01	3.497e-01	-.586	.558	7.685e-01	3.306e-01	2.325	.020*	3.736e-01	3.364e-01	1.111	.267
-500 - -420	.264	.354	.746	.456	-.129	.366	-.352	.725	.800	.346	2.313	.021*	.471	.352	1.338	.181

-400 - -320	.258	.367	.703	.482	-.067	.379	-.178	.859	.846	.358	2.361	.018*	.462	.364	1.270	.204
-300 - -220	2.259e-01	3.881e-01	.582	.561	-8.194e-02	4.007e-01	-.205	.838	9.270e-01	3.791e-01	2.445	.015*	4.224e-01	3.854e-01	1.096	.273
-200 - -120	-1.788e-01	2.785e-01	-.642	.521	-2.390e-01	2.877e-01	-.831	.406	4.903e-01	2.719e-01	1.803	.072.	2.291e-01	2.768e-01	.828	.408
100 - -20	-1.153e-01	4.080e-01	-.283	.777	-1.919e-01	4.213e-01	-.456	.649	5.963e-01	3.985e-01	1.496	.135	2.010e-01	4.053e-01	.496	.620
0 – 80	-3.233e-01	4.181e-01	-.773	.439	-2.302e-01	4.317e-01	-.533	.594	4.592e-01	4.083e-01	1.125	.261	3.331e-01	4.153e-01	.802	.423
100 – 180	-6.228e-01	4.310e-01	-1.445	.149	-2.333e-01	4.451e-01	-.524	.600	-1.254e-01	4.208e-01	-.298	.766	-6.176e-03	4.281e-01	-.014	.989
200 – 280	-5.448e-01	3.470e-01	-1.570	.116	-4.975e-01	3.584e-01	-1.388	.165	2.407e-01	3.388e-01	.711	.477	2.710e-01	3.447e-01	.786	.432
300 – 380	-6.469e-01	3.621e-01	-1.787	.074.	-4.378e-01	3.740e-01	-1.171	.242	1.126e-01	3.535e-01	.318	.750	7.224e-02	3.597e-01	.201	.841
400 – 480	-6.147e-01	3.768e-01	-1.631	.103	-3.891e-01	3.892e-01	-1.000	.318	1.556e-02	3.680e-01	.042	.966	-2.136e-01	3.744e-01	-.571	.568
500 – 580	-2.059e-01	2.775e-01	-.742	.458	-3.503e-01	2.866e-01	-1.222	.222	4.176e-01	2.708e-01	1.542	.123	2.457e-01	2.757e-01	.891	.373
600 – 680	-4.426e-01	3.941e-01	-1.123	.261	-4.510e-01	4.070e-01	-1.108	.268	3.523e-02	3.848e-01	.092	.927	-1.643e-01	3.916e-01	-.420	.675
700 – 780	-5.024e-01	3.945e-01	-1.274	.203	-5.130e-01	4.074e-01	-1.259	.208	-5.935e-02	3.852e-01	-.154	.878	-2.095e-01	3.920e-01	-.534	.593
800 – 880	-3.146e-01	3.962e-01	-.794	.427	-3.126e-01	4.092e-01	-.764	.445	1.756e-01	3.870e-01	.454	.650	-1.455e-01	3.937e-01	-.369	.712
900 – 980	-2.477e-01	4.022e-01	-.616	.538	-1.753e-01	4.154e-01	-.422	.673	1.810e-01	3.928e-01	.461	.645	-3.813e-01	3.997e-01	-.954	.340

Time window

	Voice * inanimate-inanimate * noun focused				Voice * inanimate-inanimate * cue focused			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	2.934e-01	3.035e-01	.967	.334	1.039e-01	3.117e-01	.333	.739
-700 - -620	2.673e-01	3.122e-01	.856	.392	2.265e-02	3.205e-01	.071	.944
-600 - -520	1.895e-01	3.268e-01	.580	.562	-3.533e-02	3.354e-01	-.105	.916
-500 - -420	.106	.342	.309	.757	-.087	.351	-.248	.804
-400 - -320	.270	.354	.762	.446	.222	.363	.613	.540
-300 - -220	5.754e-01	3.747e-01	1.536	.125	4.168e-01	3.843e-01	1.084	.278
-200 - -120	8.333e-02	2.688e-01	.310	.757	2.856e-03	2.759e-01	.010	.992
100 - -20	2.652e-01	3.939e-01	.673	.501	7.469e-03	4.042e-01	.018	.985
0 – 80	2.036e-01	4.036e-01	.505	.614	-1.846e-01	4.141e-01	-.446	.656
100 – 180	-1.602e-01	4.159e-01	-.385	.700	-1.217e-01	4.268e-01	-.285	.776
200 – 280	-7.004e-02	3.349e-01	-.209	.834	-1.499e-01	3.437e-01	-.436	.663

300 – 380	-9.677e-02	3.495e-01	-.277	.782	-1.145e-01	3.586e-01	-.319	.750
400 – 480	-1.087e-01	3.637e-01	-.299	.765	-1.285e-01	3.732e-01	-.344	.731
500 – 580	4.388e-02	2.677e-01	.164	.870	2.254e-02	2.749e-01	.082	.935
600 – 680	-4.105e-01	3.804e-01	-1.079	.281	-5.376e-01	3.904e-01	-1.377	.169
700 – 780	-5.854e-01	3.808e-01	-1.537	.124	-7.310e-01	3.908e-01	-1.871	.062.
800 – 880	-4.861e-01	3.825e-01	-1.271	.204	-3.957e-01	3.926e-01	-1.008	.314
900 – 980	-2.231e-01	3.883e-01	-.575	.566	-3.232e-01	3.985e-01	-.811	.417

Animacy * group * time * voice

Time window	Interaction															
	Voice * animate-inanimate * noun focused * immediate post-test				Voice * animate-inanimate * noun focused * delayed post-test				Voice * inanimate-animate * noun focused * immediate post-test				Voice * inanimate-animate * noun focused * delayed post-test			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*

-800 - -720	1.411e-01	4.490e-01	.314	.753	4.775e-01	4.516e-01	1.057	.290	-9.984e-01	4.394e-01	-2.272	.023*	-5.131e-01	4.400e-01	-1.166	.244
-700 - -620	3.440e-01	4.616e-01	.745	.456	5.739e-01	4.644e-01	1.236	.217	-1.033e+00	4.518e-01	-2.286	.022*	-4.566e-01	4.525e-01	-1.009	.313
-600 - -520	2.169e-01	4.830e-01	.449	.653	3.158e-01	4.860e-01	.650	.516	-1.166e+00	4.727e-01	-2.467	.013*	-5.685e-01	4.736e-01	-1.200	.230
-500 - -420	-.215	.505	-.425	.671	.051	.508	.100	.920	-1.122	.494	-2.269	.023*	-.496	.496	-1.001	.317
-400 - -320	-.360	.523	-.688	.491	.221	.526	.420	.674	-.654	.512	-1.279	.201	-.189	.513	-.369	.712
-300 - -220	-1.612e-01	5.530e-01	-.292	.771	5.496e-01	5.567e-01	.987	.324	-4.546e-01	5.413e-01	-.840	.401	5.741e-02	5.428e-01	.106	.916
-200 - -120	3.329e-01	3.974e-01	.838	.402	7.275e-01	3.998e-01	1.819	.069.	-4.278e-01	3.889e-01	-1.100	.272	1.843e-01	3.896e-01	.473	.636
100 - -20	5.180e-01	5.817e-01	.891	.373	1.208e+00	5.854e-01	2.063	.039*	-1.406e-01	5.694e-01	-.247	.805	5.334e-01	5.707e-01	.935	.350
0 – 80	5.793e-01	5.960e-01	.972	.331	1.528e+00	5.999e-01	2.548	.012*	-1.897e-01	5.833e-01	-.325	.745	3.936e-01	5.847e-01	.673	.501
100 – 180	6.418e-01	6.145e-01	1.044	.296	2.115e-01	6.186e-01	.342	.732	-2.413e-02	6.013e-01	-.040	.968	-3.051e-01	6.027e-01	-.506	.613
200 – 280	4.223e-01	4.949e-01	.853	.394	5.049e-01	4.981e-01	1.014	.311	-5.968e-01	4.843e-01	-1.232	.218	2.807e-02	4.853e-01	.058	.954
300 – 380	4.672e-01	5.165e-01	.905	.366	4.595e-01	5.198e-01	.884	.377	-6.238e-01	5.054e-01	-1.234	.217	4.445e-02	5.064e-01	.088	.930
400 – 480	4.766e-01	5.374e-01	.887	.375	2.945e-01	5.409e-01	.544	. 586	-6.267e-01	5.259e-01	-1.192	.234	-5.890e-02	5.271e-01	-.112	.911
500 – 580	3.031e-01	3.959e-01	.765	.444	5.013e-01	3.984e-01	1.258	.208	-5.048e-01	3.874e-01	-1.303	.193	-2.695e-02	3.881e-01	-.069	.945
600 – 680	7.590e-01	5.621e-01	1.350	.177	-1.685e-03	5.657e-01	-.003	.998	-1.627e-01	5.502e-01	-.296	.768	-2.580e-01	5.513e-01	-.468	.640

700 – 780	6.050e-01	5.627e-01	1.075	.282	3.384e-01	5.662e-01	.598	.550	-3.962e-02	5.507e-01	-.072	.943	2.242e-01	5.518e-01	.406	.685
800 – 880	3.827e-01	5.651e-01	.677	.498	1.914e-01	5.686e-01	.337	.736	-3.399e-01	5.532e-01	-.614	.539	1.914e-01	5.686e-01	.337	.736
900 – 980	3.099e-01	5.737e-01	.540	.589	6.480e-02	5.773e-01	.112	.911	-2.405e-01	5.615e-01	-.428	.668	2.480e-01	5.627e-01	.441	.659

Time window	Interaction															
	Voice * inanimate-inanimate * noun focused * immediate post-test				Voice * inanimate-inanimate * noun focused * delayed post-test				Voice * animate-inanimate * cue focused * immediate post-test				Voice * animate-inanimate * cue focused * delayed post-test			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-3.380e-01	4.341e-01	-.779	.436	-7.347e-02	4.334e-01	-.169	.865	3.734e-01	4.569e-01	.817	.413	5.534e-01	4.659e-01	1.188	.235
-700 - -620	-4.080e-01	4.464e-01	-.914	.361	8.736e-02	4.457e-01	.196	.845	4.194e-01	4.698e-01	.893	.372	6.017e-01	4.790e-01	1.256	.209
-600 - -520	-3.827e-01	4.671e-01	-.819	.413	8.233e-02	4.665e-01	.176	.860	7.708e-02	4.916e-01	.157	.875	3.337e-01	5.012e-01	.666	.506
-500 - -420	-.425	.488	-.871	.384	.124	.488	.255	.799	-.243	.514	-.471	.638	.261	.524	.498	.618
-400 - -320	-.548	.505	-1.084	.279	.244	.505	.483	.629	-.062	.532	-.117	.907	.451	.543	.831	.406

-300 - -220	-6.913e-01	5.348e-01	-1.293	.196	-6.884e-02	5.346e-01	-.129	.898	2.418e-01	5.633e-01	.429	.668	7.108e-01	5.740e-01	1.238	.216
-200 - -120	-1.371e-01	3.843e-01	-.357	.721	4.715e-01	3.838e-01	1.229	.219	3.094e-01	4.045e-01	.765	.444	5.621e-01	4.124e-01	1.363	.173
100 - -20	1.640e-01	5.625e-01	.292	.771	5.968e-01	5.622e-01	1.062	.289	3.472e-01	5.923e-01	.586	.558	6.882e-01	6.037e-01	1.140	.254
0 – 80	9.560e-02	5.763e-01	.166	.868	7.641e-01	5.760e-01	1.327	.185	4.090e-01	6.069e-01	.674	.500	6.425e-01	6.186e-01	1.039	.299
100 – 180	6.594e-01	5.941e-01	1.110	.267	9.035e-01	5.937e-01	1.522	.128	1.432e-01	6.257e-01	.229	.819	2.214e-01	6.380e-01	.347	.729
200 – 280	1.188e-01	4.784e-01	.248	.804	4.764e-01	4.780e-01	.997	.319	4.109e-01	5.038e-01	.816	.415	8.601e-01	5.137e-01	1.674	.094.
300 – 380	1.975e-01	4.993e-01	.396	.692	4.037e-01	4.988e-01	.809	.418	4.477e-01	5.257e-01	.851	.395	6.468e-01	5.361e-01	1.206	.228
400 – 480	1.498e-01	5.196e-01	.288	.773	2.703e-01	5.192e-01	.521	.603	3.034e-01	5.471e-01	.555	.579	3.809e-01	5.578e-01	.683	.495
500 – 580	-2.472e-01	3.828e-01	-.646	.518	4.007e-01	3.823e-01	1.048	.295	3.316e-01	4.029e-01	.823	.411	5.579e-01	4.109e-01	1.358	.175
600 – 680	3.747e-01	5.436e-01	.689	.491	3.066e-01	5.430e-01	.565	.572	5.954e-01	5.722e-01	1.040	.298	3.373e-01	5.834e-01	.578	.563
700 – 780	3.757e-01	5.441e-01	.691	.490	6.356e-01	5.435e-01	1.169	.242	6.143e-01	5.728e-01	1.072	.284	5.052e-01	5.839e-01	.865	.387
800 – 880	-7.147e-02	5.465e-01	-.131	.896	4.401e-01	5.460e-01	.806	.420	3.142e-01	5.754e-01	.546	.585	3.849e-01	5.864e-01	.656	.512
900 – 980	-4.107e-01	5.547e-01	-.740	.459	2.392e-02	5.542e-01	.043	.966	3.142e-01	5.840e-01	.538	.591	1.860e-01	5.953e-01	.312	.755

Time window	Interaction															
	Voice * inanimate-animate * cue focused * immediate post-test				Voice * inanimate-animate * cue focused * delayed post-test				Voice * inanimate-inanimate * cue focused * immediate post-test				Voice * inanimate-inanimate * cue focused * delayed post-test			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-5.878e-01	4.493e-01	-1.308	.19092	-8.215e-01	4.528e-01	-1.815	.07.	6.598e-02	4.437e-01	.149	.882	-2.832e-01	4.471e-01	-.634	.526
-700 - -620	-4.662e-01	4.620e-01	-1.009	.3130	-7.708e-01	4.655e-01	-1.656	.098.	1.425e-01	4.563e-01	.312	.755	-2.192e-01	4.597e-01	-.477	.633
-600 - -520	-6.483e-01	4.835e-01	-1.341	.180	-8.959e-01	4.872e-01	-1.839	.066.	1.956e-02	4.776e-01	.041	.967	-1.554e-01	4.810e-01	-.323	.747
-500 - -420	-.757	.506	-1.497	.134	-.865	.510	-1.697	.090.	-.133	.500	-.266	.790	.019	.503	.039	.970
-400 - -320	-.398	.523	-.760	.447	-.359	.527	-.681	.496	-.304	.517	-.587	.557	-.044	.521	-.085	.932
-300 - -220	-1.271e-01	5.540e-01	-.229	.819	1.532e-01	5.580e-01	.275	.784	-3.807e-01	5.472e-01	-.696	.487	-2.636e-01	5.510e-01	-.478	.632
-200 - -120	-3.053e-01	3.978e-01	-.768	.443	-2.184e-01	4.008e-01	-.545	.586	8.890e-03	3.929e-01	.023	.982	1.801e-01	3.957e-01	.455	.649
100 - -20	-9.117e-02	5.826e-01	-.156	.876	-3.738e-02	5.868e-01	-.064	.949	3.784e-01	5.754e-01	.658	.512	1.265e-01	5.795e-01	.218	.827
0 – 80	-7.426e-02	5.969e-01	-.124	.901	-4.836e-01	6.012e-01	-.804	.421	6.224e-01	5.895e-01	1.056	.291	3.363e-01	5.937e-01	.567	.571
100 – 180	-5.996e-01	6.152e-01	-.975	.330	-5.125e-01	6.198e-01	-.827	.408	-5.390e-02	6.076e-01	-.089	.929	2.578e-01	6.120e-01	.421	.674

200 – 280	-7.606e-01	4.954e-01	-1.535	.125	4.452e-01	4.991e-01	-.892	.372	-1.157e-01	4.893e-01	-.237	.813	1.305e-01	4.928e-01	.265	.791
300 – 380	-5.524e-01	5.170e-01	-1.068	.285	-4.733e-01	5.209e-01	-.909	.364	-1.792e-03	5.106e-01	-.004	.997	3.738e-02	5.143e-01	.073	.942
400 – 480	-6.182e-01	5.380e-01	-1.149	.251	-2.451e-01	5.420e-01	-.452	.651	-1.219e-02	5.314e-01	-.023	.982	6.473e-03	5.352e-01	.012	.990
500 – 580	-4.134e-01	3.962e-01	-1.043	.297	-3.685e-01	3.993e-01	-.923	.356	-1.207e-01	3.913e-01	-.308	.758	1.448e-01	3.942e-01	.367	.714
600 – 680	-3.039e-01	5.628e-01	-.540	.589	-1.990e-01	5.670e-01	-.351	.726	4.071e-01	5.558e-01	.732	.464	6.551e-01	5.599e-01	1.170	.242
700 – 780	9.547e-02	5.634e-01	.169	.865	1.512e-02	5.675e-01	.027	.979	5.813e-01	5.564e-01	1.045	.296	1.512e-02	5.675e-01	.027	.979
800 – 880	4.213e-02	5.659e-01	.074	.941	1.172e-01	5.701e-01	.206	.837	6.126e-02	5.589e-01	.110	.913	4.660e-01	5.629e-01	.828	.408
900 – 980	4.669e-01	5.744e-01	.813	.416	2.244e-01	5.787e-01	.388	.698	8.493e-02	5.673e-01	.150	.881	2.454e-01	5.714e-01	.429	.668

Untrained constructions (present simple) – main effect of group and test time after the verb ending

Signif. codes: 0 ‘***’ .001 ‘**’ .01 ‘*’ .05 ‘.’ .1 ‘ ’ 1

Time window	Main effect															
	Immediate post-test (compared to pre-test)				Delayed post-test (compared to pre-test)				Noun focused group (compared to test-only group)				Cue focused group (compared to test-only group)			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-2.074e-02	1.546e-01	-.134	.893	2.948e-01	1.547e-01	1.906	.057.	9.878e-02	1.559e-01	.633	.527	3.336e-02	1.601e-01	.208	.835
-700 - -620	-3.085e-02	1.577e-01	-.196	.845	2.759e-01	1.578e-01	1.748	.08.	1.055e-01	1.595e-01	.661	.508	5.375e-02	1.637e-01	.328	.743
-600 - -520	-2.262e-02	1.671e-01	-.135	.892	2.666e-01	1.672e-01	1.595	.111	1.171e-01	1.690e-01	.692	.489	4.430e-02	1.735e-01	.255	.799
-500 - -420	1.116e-02	1.741e-01	.064	.949	1.953e-01	1.742e-01	1.121	.262	1.290e-01	1.771e-01	.728	.467	7.930e-02	1.817e-01	.436	.663
-400 - -320	-5.343e-02	1.829e-01	-.292	.770	1.373e-01	1.830e-01	.750	.453	-5.552e-03	1.871e-01	-.030	.977	3.304e-02	1.918e-01	.172	.863
-300 - -220	-7.293e-02	1.929e-01	-.378	.705	9.026e-02	1.930e-01	.468	.640	-1.025e-01	1.973e-01	-.519	.604	-6.543e-02	2.022e-01	-.324	.746
-200 - -120	-8.396e-02	1.389e-01	-.605	.545	1.009e-01	1.389e-01	.726	.468	-7.182e-02	1.410e-01	-.509	.611	5.416e-03	1.447e-01	.037	.971
100 - -20	5.713e-04	1.999e-01	.003	.998	4.757e-02	2.000e-01	.238	.812	-1.026e-01	2.033e-01	-.505	.614	-8.354e-02	2.086e-01	-.400	.689

0 – 80	-2.798e-02	2.082e-01	-.134	.893	-1.137e-02	2.083e-01	-.055	.956	-1.867e-01	2.122e-01	-.880	.379	-1.442e-01	2.177e-01	-.662	.508
100 – 180	-.430	.216	-1.990	.047*	-.301	.216	-1.390	.165	-.063	.218	-.289	.772	.014	.224	.065	.949
200 – 280	-1.799e-01	1.677e-01	-1.072	.284	-1.009e-02	1.678e-01	-.060	.952	-6.295e-02	1.699e-01	-.371	.711	1.239e-02	1.744e-01	.071	.943
300 – 380	-8.487e-02	1.738e-01	-.488	.625	5.876e-03	1.739e-01	.034	.973	-6.955e-02	1.759e-01	-.395	.693	4.992e-02	1.805e-01	.276	.782
400 – 480	-6.524e-02	1.815e-01	-.359	.719	2.568e-02	1.816e-01	.141	.888	-8.810e-02	1.834e-01	-.480	.631	8.263e-02	1.883e-01	.439	.661
500 – 580	-4.890e-02	1.378e-01	-.355	.723	1.339e-01	1.379e-01	.971	.332	-9.162e-03	1.395e-01	-.066	.948	4.244e-02	1.433e-01	.296	.767
600 – 680	-4.599e-02	1.876e-01	-.245	.806	7.077e-02	1.877e-01	.377	.706	6.652e-03	1.899e-01	.035	.972	2.505e-01	1.949e-01	1.285	.199
700 – 780	-7.538e-03	1.919e-01	-.039	.969	1.414e-01	1.920e-01	.736	.462	-1.796e-02	1.947e-01	-.092	.927	7.451e-02	1.998e-01	.373	.709
800 – 880	-3.401e-04	1.938e-01	-.002	.999	1.169e-01	1.939e-01	.603	.547	-7.847e-03	1.970e-01	-.040	.968	2.272e-02	2.022e-01	.112	.911
900 – 980	-5.073e-02	1.943e-01	-.261	.794	4.440e-02	1.944e-01	.228	.819	-3.271e-02	1.981e-01	-.165	.869	-6.223e-02	2.032e-01	-.306	.759

Interactions – group * time

Time window	Main effect															
	Noun focused group (compared to test-only group) * Immediate post-test (compared to pre-test)				Noun focused group (compared to test-only group) * Delayed post-test (compared to pre-test)				Cue focused group (compared to test-only group) * Immediate post-test (compared to pre-test)				Cue focused group (compared to test-only group) * Delayed post-test (compared to pre-test)			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	7.685e-02	2.212e-01	.347	.728	-3.425e-01	2.215e-01	-1.546	.122	7.051e-02	2.245e-01	.314	.753	-1.312e-01	2.278e-01	-.576	.565
-700 - -620	8.307e-02	2.257e-01	.368	.713	-3.432e-01	2.260e-01	-1.518	.129	6.729e-02	2.291e-01	.294	.769	-1.564e-01	2.325e-01	-.673	.501
-600 - -520	5.907e-02	2.391e-01	.247	.805	-3.248e-01	2.394e-01	-1.356	.175	7.340e-02	2.427e-01	.302	.762	-1.767e-01	2.463e-01	-.717	.473
-500 - -420	3.615e-03	2.491e-01	.015	.989	-2.856e-01	2.495e-01	-1.145	.252	9.190e-02	2.529e-01	.363	.716	-1.693e-01	2.566e-01	-.660	.509
-400 - -320	1.224e-01	2.617e-01	.468	.640	-1.105e-01	2.621e-01	-.422	.673	1.821e-01	2.656e-01	.686	.493	-1.684e-01	2.697e-01	-.624	.532
-300 - -220	1.229e-01	2.759e-01	.445	.656	6.538e-02	2.764e-01	.236	.813	3.599e-01	2.802e-01	1.285	.199	-2.309e-01	2.844e-01	-.812	.417
-200 - -120	1.810e-01	1.988e-01	.911	.363	-4.400e-02	1.989e-01	-.221	.825	1.739e-01	2.016e-01	.862	.389	-7.636e-02	2.048e-01	-.373	.709
100 - -20	-2.854e-02	2.860e-01	-.100	.921	1.406e-01	2.864e-01	.491	.622	4.450e-01	2.903e-01	1.533	.125	-1.784e-01	2.947e-01	-.605	.545

0 – 80	2.980e-02	2.979e-01	.100	.920	1.964e-01	2.983e-01	.659	.510	4.613e-01	3.023e-01	1.526	.127	-7.505e-02	3.070e-01	-.245	.807
100 – 180	.373	.310	1.204	.229	.110	.310	.356	.722	.165	.314	.526	.599	-.077	.319	-.242	.809
200 – 280	1.741e-01	2.400e-01	.725	.468	-2.178e-01	2.403e-01	-.906	.365	1.440e-01	2.436e-01	.591	.554	-1.548e-02	2.472e-01	-.063	.950
300 – 380	5.316e-02	2.488e-01	.214	.831	-1.890e-01	2.491e-01	-.759	.448	6.318e-03	2.524e-01	.025	.980	-2.325e-02	2.562e-01	-.091	.928
400 – 480	2.946e-02	2.598e-01	.113	.910	-1.902e-01	2.601e-01	-.731	.465	-8.060e-02	2.636e-01	-.306	.760	2.226e-03	2.676e-01	.008	.993
500 – 580	1.721e-01	1.973e-01	.872	.383	-1.363e-01	1.974e-01	-.690	.490	-1.221e-02	2.001e-01	-.061	.951	-4.776e-02	2.032e-01	-.235	.814
600 – 680	1.326e-01	2.685e-01	.494	.621	-2.283e-01	2.688e-01	-.849	.396	-2.332e-01	2.724e-01	-.856	.392	-4.267e-02	2.765e-01	-.154	.877
700 – 780	1.731e-01	2.747e-01	.630	.529	-3.014e-01	2.750e-01	-1.096	.273	-6.041e-02	2.787e-01	-.217	.828	9.628e-02	2.829e-01	.340	.734
800 – 880	1.927e-01	2.774e-01	.695	.487	-2.008e-01	2.777e-01	-.723	.470	6.098e-02	2.815e-01	.217	.829	1.436e-01	2.857e-01	.503	.615
900 – 980	1.799e-01	2.781e-01	.647	.518	2.834e-02	2.784e-01	.102	.919	1.465e-01	2.822e-01	.519	.604	2.670e-01	2.866e-01	.932	.352

Voice * Test time

Time window	Main effects															
	Noun focused group * Immediate post-test * voice (passive)				Noun focused group * Delayed post-test * voice (passive)				Cue focused group * Immediate post-test * voice (passive)				Cue focused group * Delayed post-test * voice (passive)			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-3.657e-02	3.129e-01	-.117	.907	8.307e-01	3.127e-01	2.656	.008**	9.599e-02	3.176e-01	.302	.762	4.587e-01	3.222e-01	1.424	.155
-700 - -620	-3.607e-02	3.192e-01	-.113	.910	8.332e-01	3.191e-01	2.611	.009**	9.229e-02	3.240e-01	.285	.776	5.054e-01	3.287e-01	1.537	.124
-600 - -520	-2.570e-02	3.381e-01	-.076	.939	8.577e-01	3.380e-01	2.538	.011*	1.024e-01	3.432e-01	.298	.766	4.097e-01	3.482e-01	1.177	.239
-500 - -420	-7.605e-02	3.522e-01	-.216	.829	6.445e-01	3.522e-01	1.830	.067.	4.238e-02	3.576e-01	.118	.906	1.064e-01	3.627e-01	.293	.769
-400 - -320	-1.544e-01	3.701e-01	-.417	.676	5.422e-01	3.699e-01	1.466	.143	-3.157e-02	3.756e-01	-.084	.933	1.132e-01	3.811e-01	.297	.767
-300 - -220	1.249e-01	3.902e-01	.320	.749	6.528e-01	3.902e-01	1.673	.094.	1.767e-01	3.962e-01	.446	.656	3.354e-01	4.018e-01	.835	.404
-200 - -120	5.548e-03	2.811e-01	.020	.984	6.400e-01	2.808e-01	2.279	.023*	1.407e-01	2.851e-01	.493	.622	4.368e-01	2.894e-01	1.509	.131
100 - -20	5.815e-01	4.044e-01	1.438	.151	9.237e-01	4.043e-01	2.285	.022*	3.799e-01	4.105e-01	.925	.355	8.552e-01	4.165e-01	2.053	.040*

0 – 80	4.057e-01	4.213e-01	.963	.336	8.082e-01	4.210e-01	1.919	.055.	3.410e-01	4.276e-01	.798	.425	7.367e-01	4.338e-01	1.698	.090.
100 – 180	-.458	.4379	-1.046	.296	.313	.438	.715	.475	-.127	.444	-.286	.775	.693	.451	1.536	.125
200 – 280	-1.901e-01	3.394e-01	-.560	.576	5.173e-01	3.392e-01	1.525	.127	7.787e-03	3.445e-01	.023	.982	2.711e-01	3.495e-01	.776	.438
300 – 380	-1.417e-01	3.518e-01	-.403	.687	4.403e-01	3.516e-01	1.252	.211	2.339e-01	3.570e-01	.655	.512	3.024e-01	3.623e-01	.835	.404
400 – 480	-1.883e-01	3.674e-01	-.513	.608	4.092e-01	3.672e-01	1.114	.265	1.408e-01	3.729e-01	.378	.706	1.441e-01	3.783e-01	.381	.703
500 – 580	-1.192e-01	2.790e-01	-.427	.669	4.921e-01	2.787e-01	1.766	.078.	1.480e-01	2.830e-01	.523	.601	2.817e-01	2.872e-01	.981	.327
600 – 680	-1.823e-01	3.797e-01	-.480	.631	5.239e-01	3.795e-01	1.381	.168	3.214e-01	3.853e-01	.834	.404	1.621e-01	3.910e-01	.415	.679
700 – 780	-2.352e-01	3.884e-01	-.606	.545	4.778e-01	3.882e-01	1.231	.218	1.686e-01	3.942e-01	.428	.669	-9.880e-04	4.000e-01	-.002	.998
800 – 880	-7.349e-02	3.922e-01	-.187	.851	4.636e-01	3.920e-01	1.183	.237	9.958e-02	3.980e-01	.250	.803	6.714e-02	4.039e-01	.166	.868
900 – 980	1.057e-01	3.933e-01	.269	.788	2.583e-01	3.930e-01	.657	.511	4.627e-03	3.991e-01	.012	.991	9.261e-03	4.050e-01	.023	.982

Native speakers – main effect of voice, tense and reduced passive after the verb ending

Signif. codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Time window	Main effect												Interaction			
	Voice (passive)				Tense (present simple)				Reduced passive (full passive)				Voice * tense			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	1.304e-01	1.839e-01	.709	.479	3.074e-01	1.729e-01	1.778	.076.	4.688e-02	1.747e-01	.268	.789	-4.283e-01	2.525e-01	-1.696	.09.
100 – 180	2.655e-02	1.738e-01	.153	.879	2.926e-01	1.634e-01	1.791	.074.	3.472e-02	1.650e-01	.210	.833	-5.114e-01	2.386e-01	-2.144	.032*
200 – 280	.021	.174	.120	.905	.215	.163	1.315	.189	.028	.165	.172	.864	-.455	.238	-1.908	.057.
300 – 380	-.027	.172	-.159	.874	.044	.162	.270	.787	.002	.164	.011	.991	-.308	.237	-1.301	.194
400 – 480	-.091	.172	-.530	.597	-.093	.162	-.575	.566	.021	.164	.126	.900	-.133	.236	-.563	.574
500 – 580	-.113	.168	-.671	.503	-.110	.158	-.699	.486	.044	.160	.276	.783	.112	.230	.485	.628
600 – 680	-.112	.173	-.651	.516	-.143	.1624	-.883	.378	.054	.165	.330	.742	.155	.237	.652	.515
700 – 780	-.160	.168	-.952	.341	-.113	.158	-.713	.476	.032	.160	.202	.840	.188	.231	.812	.417
800 – 880	-.110	.162	-.668	.504	-.035	.154	-.229	.819	-.084	.156	-.542	.588	-.022	.225	-.097	.923

900 – 980	-1.079e-01	1.621e-01	-.665	.506	-8.292e-02	1.524e-01	-.544	.587	-4.181e-02	1.540e-01	-.272	.786	-2.236e-02	2.226e-01	-.100	.920

Native speakers - main effects of animacy

Time window	Main effects											
	Animate - inanimate				Inanimate - animate				Inanimate - inanimate			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-.048	.171	-.279	.780	.003	.173	.022	.983	-.089	.175	-.506	.613
-700 - -620	-.065	.183	-.353	.724	-.037	.185	-.200	.842	-.116	.188	-.619	.537
-600 - -520	.069	.202	-.342	.732	-.067	.205	-.326	.745	-.067	.207	-.323	.747
-500 - -420	-5.403e-02	1.954e-01	-.277	.782	3.060e-02	1.975e-01	.155	.877	9.175e-02	2.004e-01	.458	.647
-400 - -320	-.130	.191	-.676	.499	-.011	.193	-.056	.956	.033	.196	.170	.865
-300 - -220	-.017	.191	-.091	.928	.115	.193	.594	.553	.200	.196	1.018	.309

-200 - -120	-.031	.191	-.161	.873	.057	.193	.298	.766	.205	.196	1.047	.296
100 - -20	-.089	.188	-.474	.636	.111	.189	.584	.560	.189	.193	.979	.328
0 – 80	-1.205e-01	1.846e-01	-.653	.514	1.546e-01	1.854e-01	.834	.404	1.596e-01	1.894e-01	.842	.340

Native and learners (t1) and voice after the verb ending

Signif. codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Time window	Main effect								Interaction			
	Group (native)				Voice (passive)				Group * voice			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
0 – 80	-2.401e-01	1.571e-01	-1.528	.126	-2.133e-01	1.250e-01	-1.706	.089.	3.383e-01	2.147e-01	1.575	.115
100 – 180	-2.359e-01	1.547e-01	-1.525	.127	-2.370e-01	1.248e-01	-1.899	.059.	2.716e-01	2.115e-01	1.284	.199
200 – 280	-2.430e-01	1.404e-01	-1.731	.083.	-2.741e-01	1.087e-01	-2.522	.012*	2.916e-01	1.913e-01	1.525	.128
300 – 380	-1.425e-01	1.414e-01	-1.008	.313	-2.482e-01	1.106e-01	-2.245	.026 *	2.197e-01	1.931e-01	1.138	.255

400 – 480	-6.054e-02	1.438e-01	-.421	.674	-2.191e-01	1.107e-01	-1.980	.049 *	1.300e-01	1.959e-01	.663	.507
500 – 580	-7.516e-02	1.262e-01	-.595	.552	-1.417e-01	9.376e-02	-1.511	.131	2.640e-02	1.711e-01	.154	.877
600 – 680	-1.207e-01	1.434e-01	-.842	.340	-2.661e-01	1.080e-01	-2.463	.015 *	1.560e-01	1.948e-01	.801	.423
700 – 780	6.455e-03	1.453e-01	.044	.965	-2.049e-01	1.101e-01	-1.861	.064.	4.639e-02	1.976e-01	.235	.815
800 – 880	8.749e-02	1.429e-01	.612	.541	-1.556e-01	1.134e-01	-1.372	.171	4.993e-02	1.954e-01	.256	.798
900 – 980	1.087e-01	1.436e-01	.757	.449	-1.439e-01	1.138e-01	-1.264	.208	3.686e-02	1.961e-01	.188	.851

Group and animacy main effects – effects before the verb

Time window	Main effect															
	Group				Animate - inanimate				Inanimate – animate				Inanimate - inanimate			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-2.225e-01	1.252e-01	-1.778	.076.	8.032e-02	1.018e-01	.789	.431	-8.853e-03	1.036e-01	-.085	.932	-6.779e-02	1.018e-01	-.666	.506
-700 - -620	-2.028e-01	1.309e-01	-1.549	.122	5.453e-02	1.080e-01	.505	.614	4.063e-03	1.100e-01	.037	.971	-8.125e-02	1.080e-01	-.752	.453

-600 - -520	-1.778e-01	1.445e-01	-1.231	.219	3.337e-02	1.198e-01	.279	.781	-1.776e-02	1.219e-01	-.146	.884	-8.743e-02	1.198e-01	-.730	.466
-500 - -420	-.253	.146	-1.730	.084.	.019	.120	.151	.880	-.025	.122	-.203	.840	-.079	.120	-.658	.511
-400 - -320	-1.671e-01	1.495e-01	-1.118	.262	6.923e-02	1.178e-01	.588	.558	6.323e-02	1.200e-01	.527	.599	-4.597e-02	1.178e-01	-.390	.670
-300 - -220	-.214	.155	-1.385	.166	.105	.125	.839	.402	.181	.127	1.422	.156	.011	.125	.085	.933
-200 - -120	-2.767e-01	1.370e-01	-2.019	.0436 *	5.094e-02	1.055e-01	.483	.630	3.543e-02	1.074e-01	.330	.742	-4.413e-02	1.055e-01	-.418	.676
100 - -20	-2.092e-01	1.579e-01	-1.324	.186	4.738e-02	1.289e-01	.368	.714	1.560e-01	1.312e-01	1.189	.235	3.591e-02	1.289e-01	.279	.781
0 – 80	-2.401e-01	1.571e-01	-1.528	.127	1.116e-01	1.253e-01	.890	.374	1.977e-01	1.276e-01	1.550	.122	5.525e-02	1.253e-01	.441	.660

Time window	Interaction											
	Group * animate - inanimate				Group * inanimate – animate				Group * inanimate - inanimate			
	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*	Estimate	SE	*t*	*p*
-800 - -720	-1.327e-01	1.709e-01	-.777	.437	1.862e-02	1.740e-01	.107	.915	-2.109e-02	1.741e-01	-.121	.904
-700 - -620	-1.262e-01	1.785e-01	-.707	.480	-3.762e-02	1.821e-01	-.207	.836	-3.525e-02	1.819e-01	-.194	.846
-600 - -520	-1.124e-01	1.971e-01	-.570	.569	-4.177e-02	2.011e-01	-.208	.836	2.158e-02	2.008e-01	.108	.914

-500 - -420	-.084	.199	-.420	.674	.065	.203	.321	.749	.172	.202	.847	.397
-400 - -320	-2.048e-01	2.027e-01	-1.010	.312	-6.972e-02	2.060e-01	-.338	.735	8.106e-02	2.065e-01	.393	.695
-300 - -220	-.130	.210	-.621	.535	-.066	.214	-.308	.758	.188	.214	.877	.381
-200 - -120	5.094e-02	1.055e-01	.483	.630	3.543e-02	1.074e-01	.330	.742	-4.413e-02	1.055e-01	-.418	.676
100 - -20	-1.395e-01	2.149e-01	-.649	.516	-4.119e-02	2.191e-01	-.188	.851	1.575e-01	2.189e-01	.719	.472
0 – 80	-2.365e-01	2.137e-01	-1.106	.269	-4.449e-02	2.173e-01	-.205	.838	1.044e-01	2.177e-01	.479	.632

Appendix 20 – First noun animacy and learners – immediate and delayed post-test

Immediate post-test – active sentences

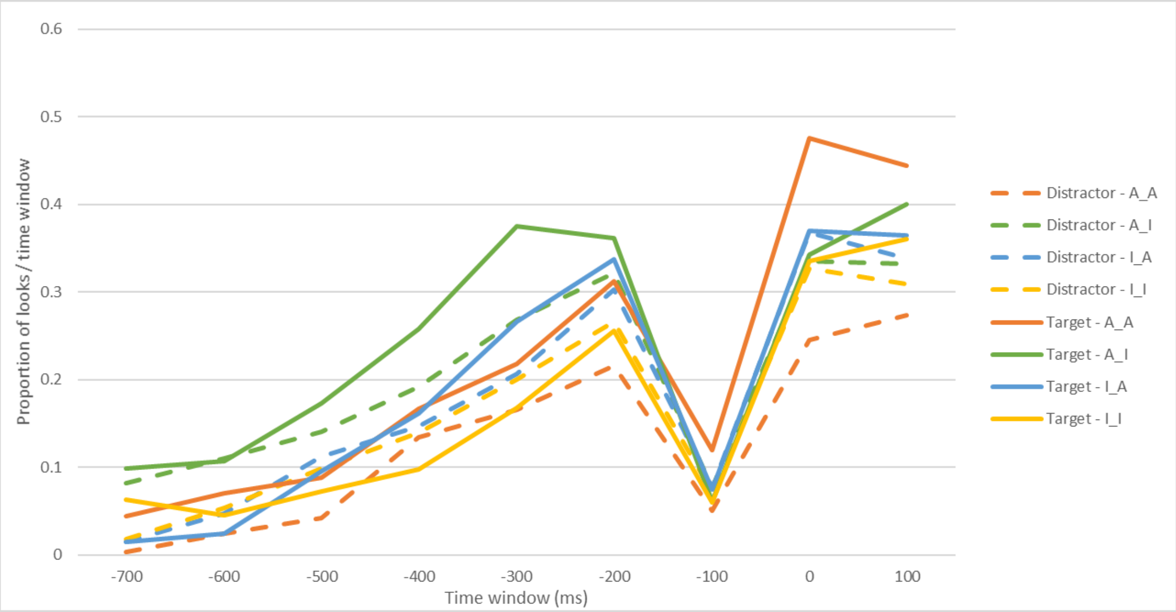


Figure 1. Cue focused group at immediate post-test- pre-verbal animacy and active voice

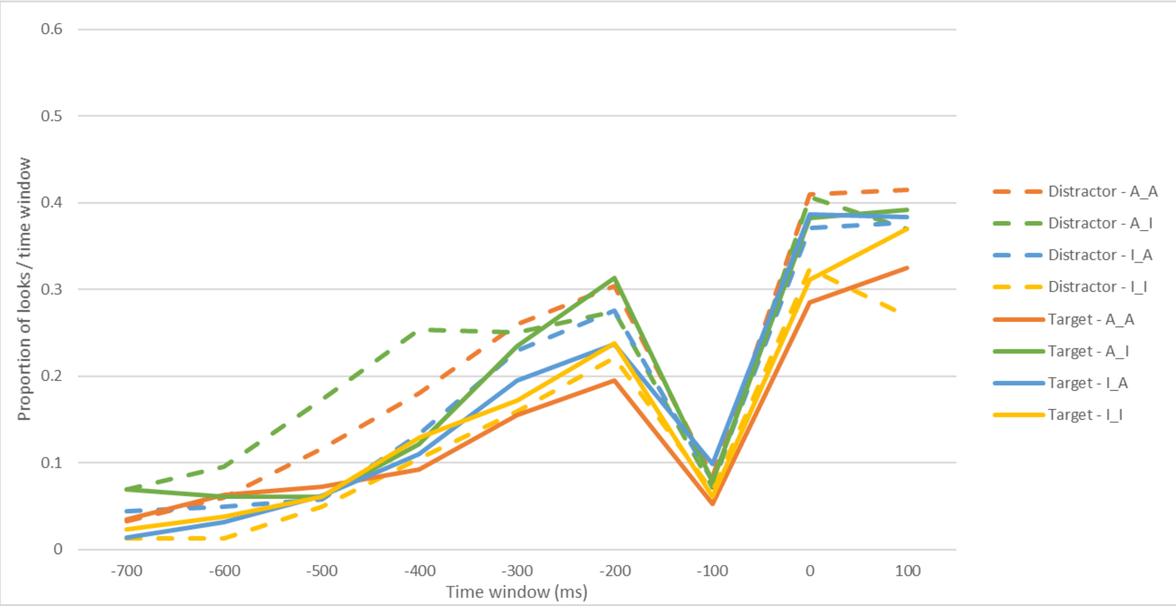


Figure 2. Noun focused group at immediate post-test- pre-verbal animacy and active voice

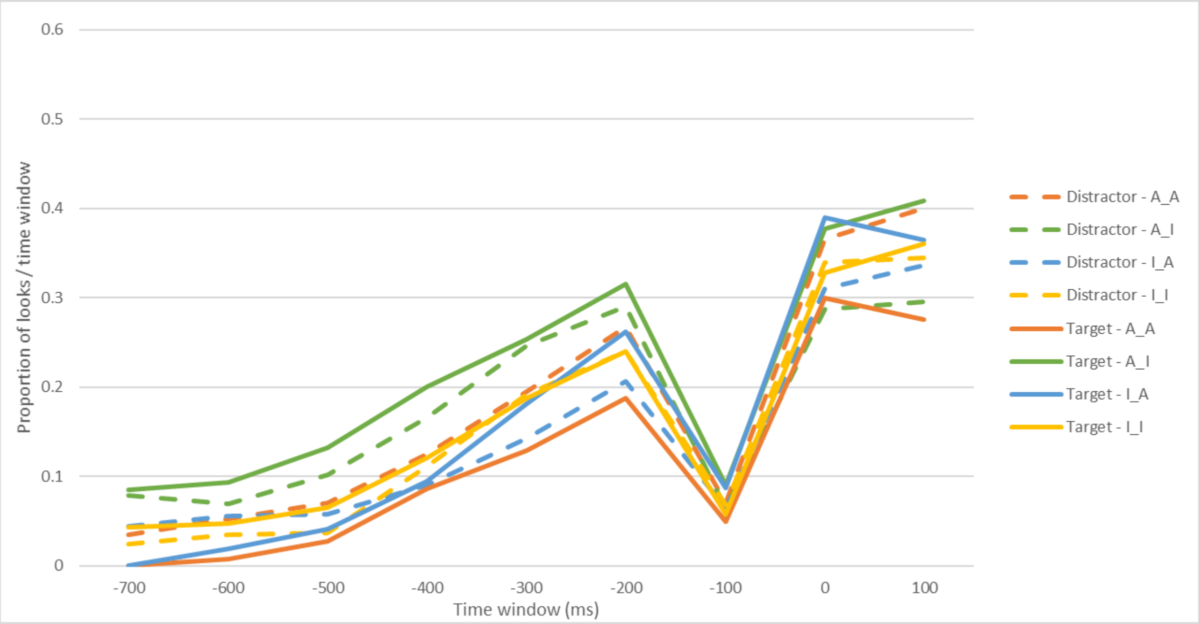


Figure 3. Test only group at immediate post-test- pre-verbal animacy and active voice

Delayed post-test – passive and active sentences

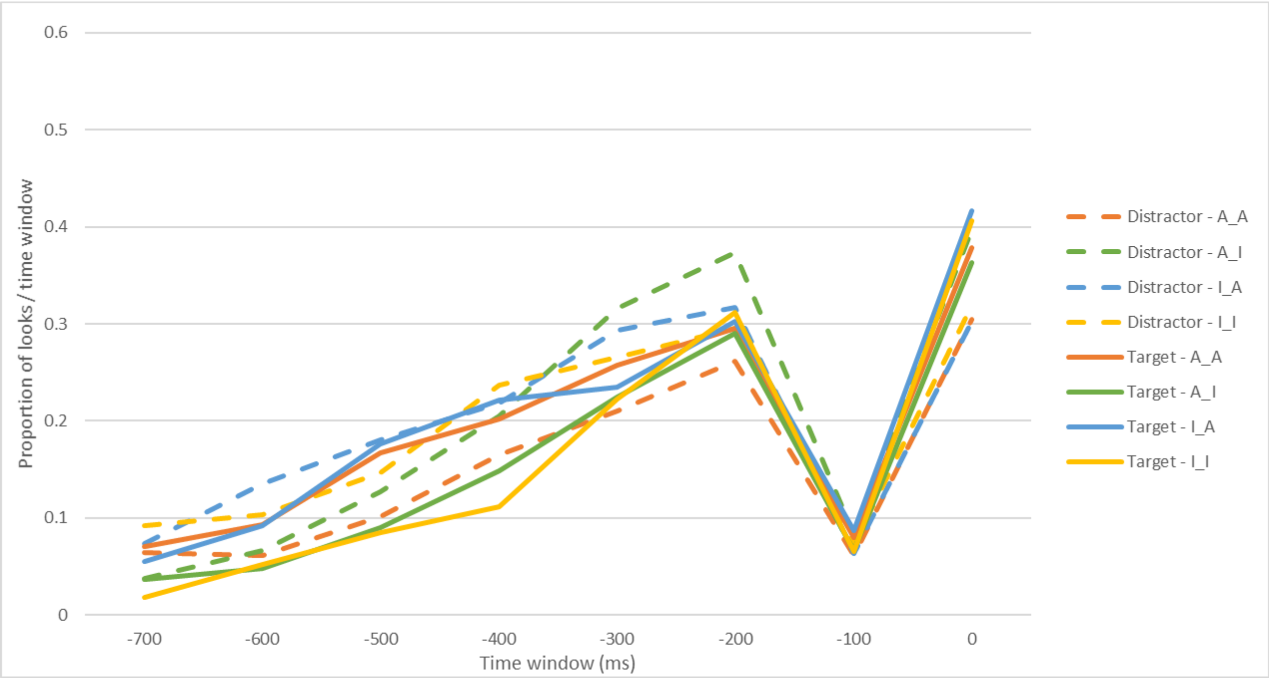


Figure 4. Cue focused group and passives and animacy at delayed post-test

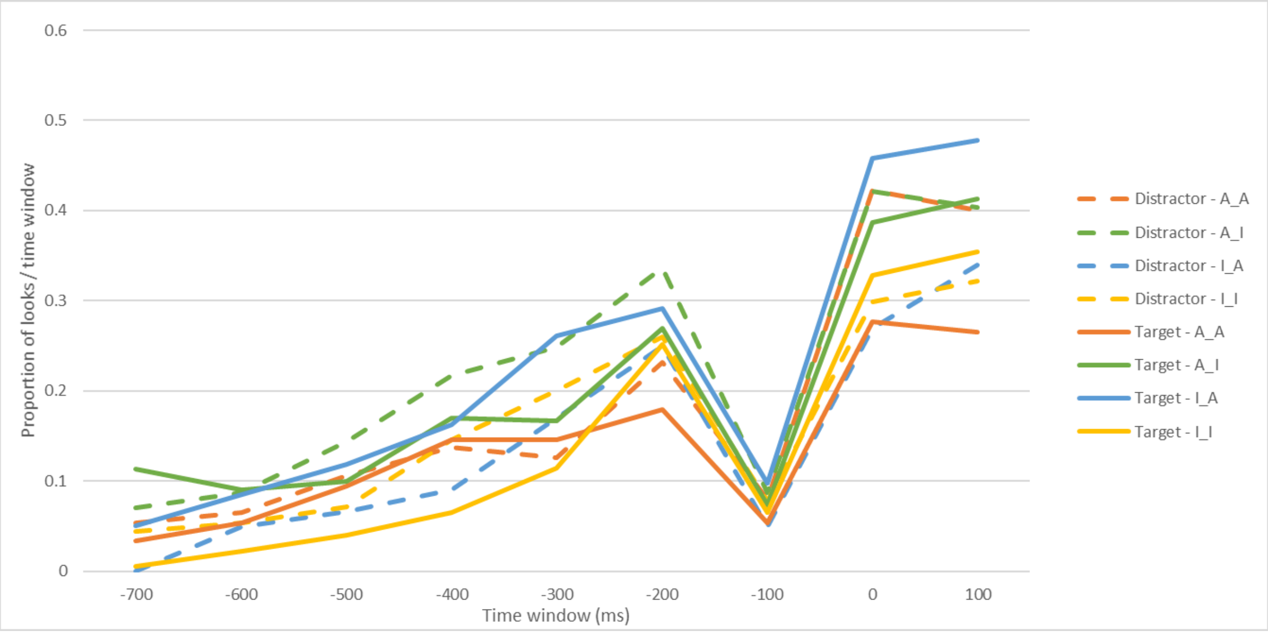


Figure 5. Cue focused group and actives and animacy at delayed post-test

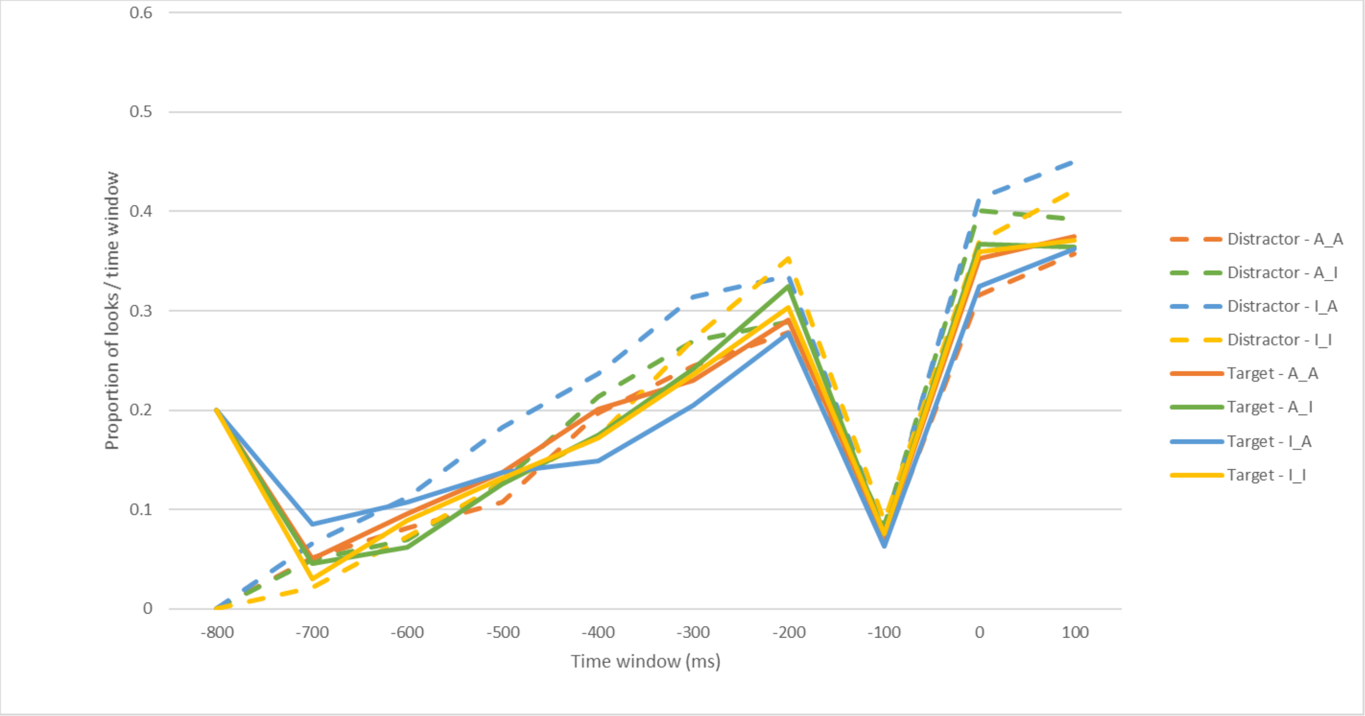


Figure 6. Noun focused group and passives and animacy at delayed post-test

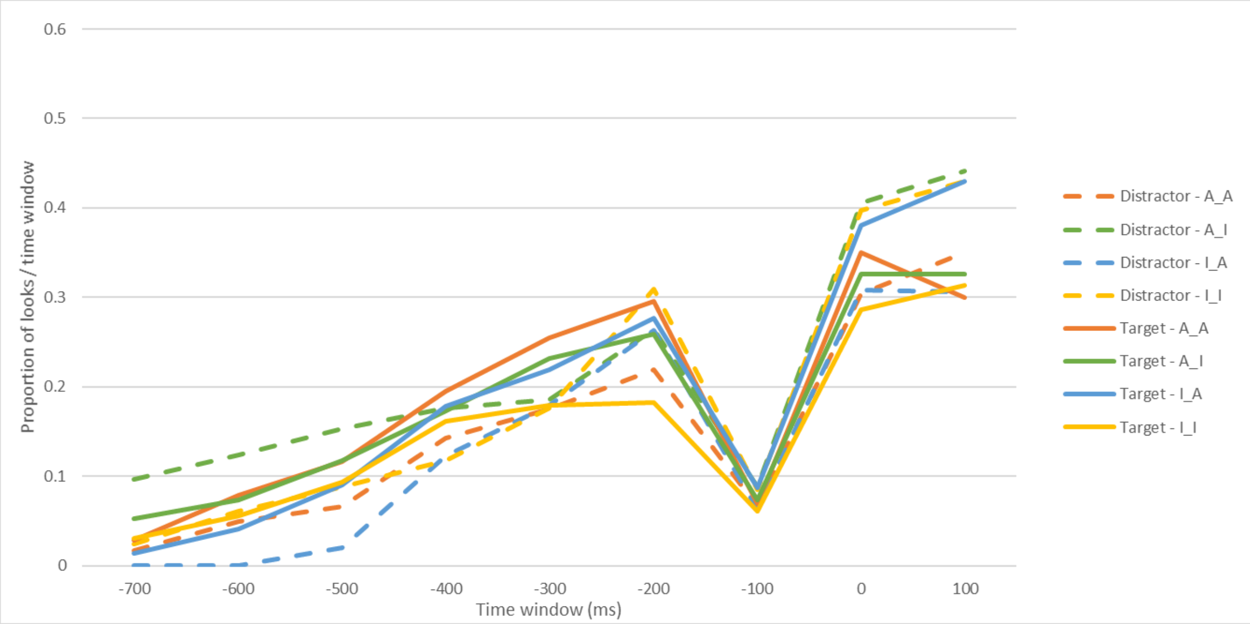


Figure 7. Noun focused group and actives and animacy at delayed post-test

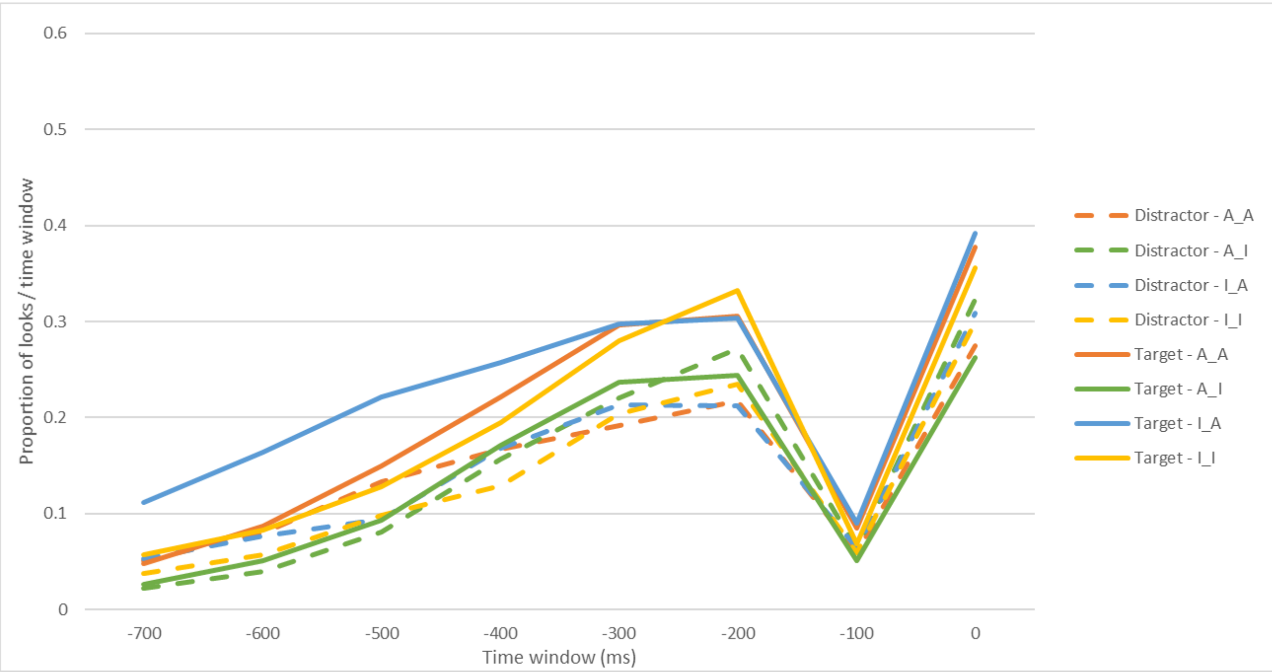


Figure 8. Passives and test-only group and animacy at delayed post-test

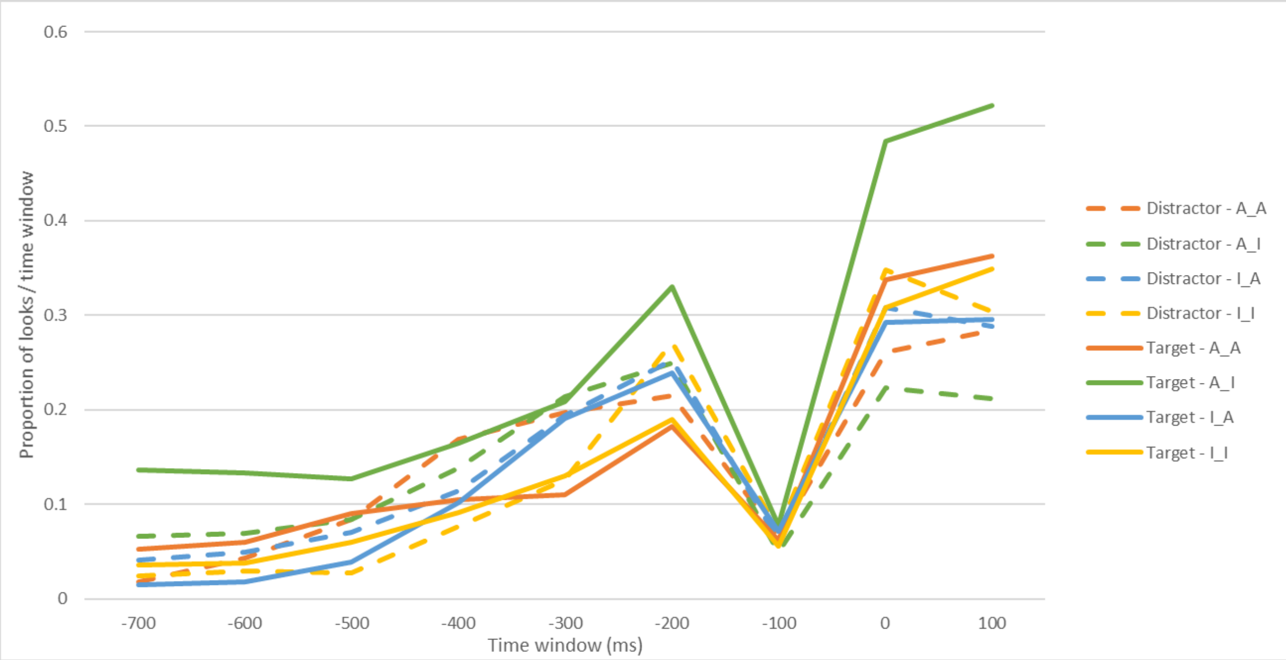


Figure 9. Actives and test-only group and animacy at delayed post-test

Appendix 21 – Untrained passives – reduced compared to full

Reduced / full passives - Pre-test

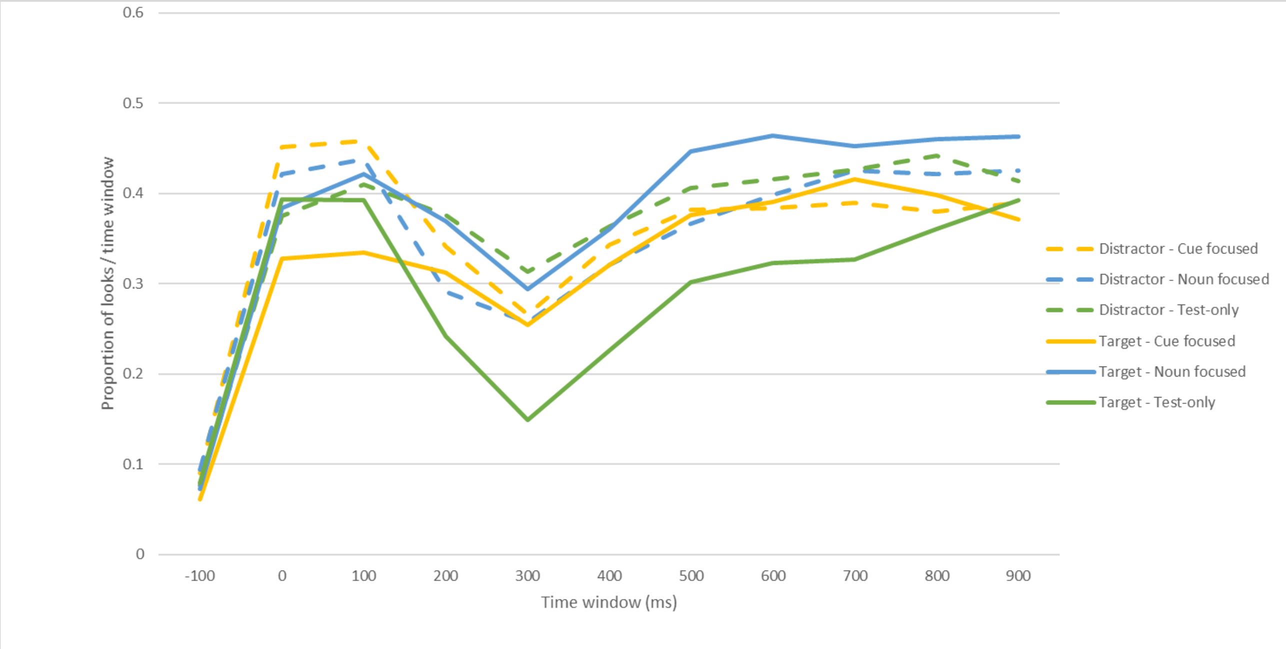


Figure 1. Pre-test - full passives and group

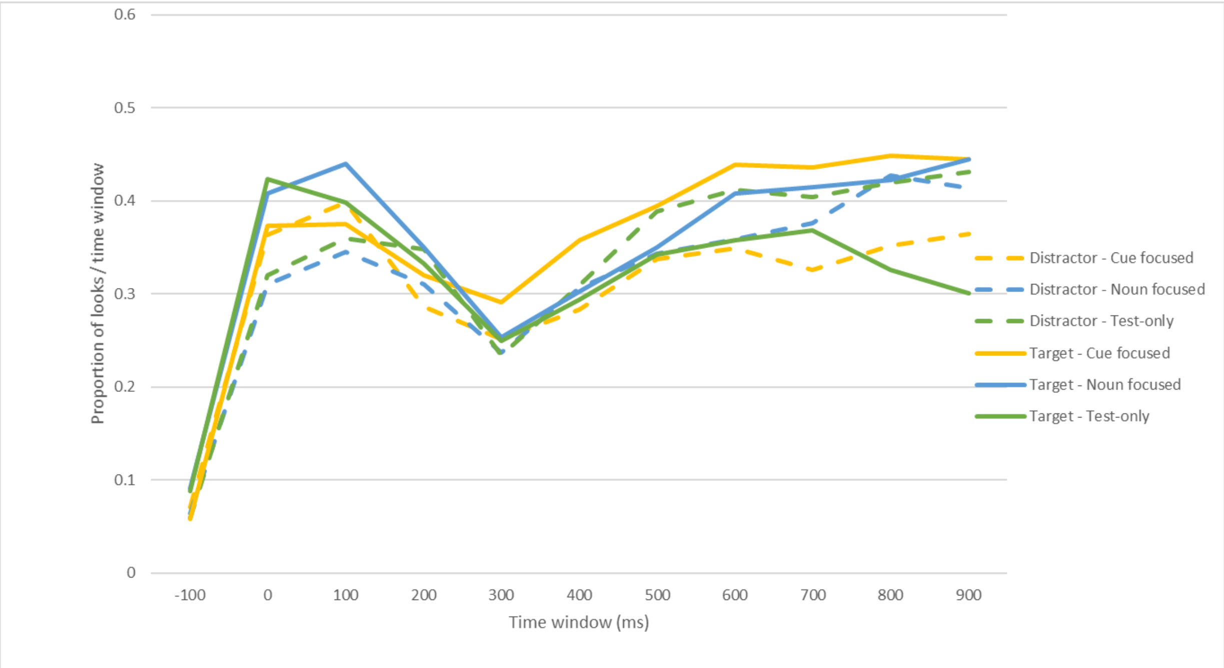


Figure 2. Pre-test - reduced passives and group

Reduced / full passives – Immediate post-test

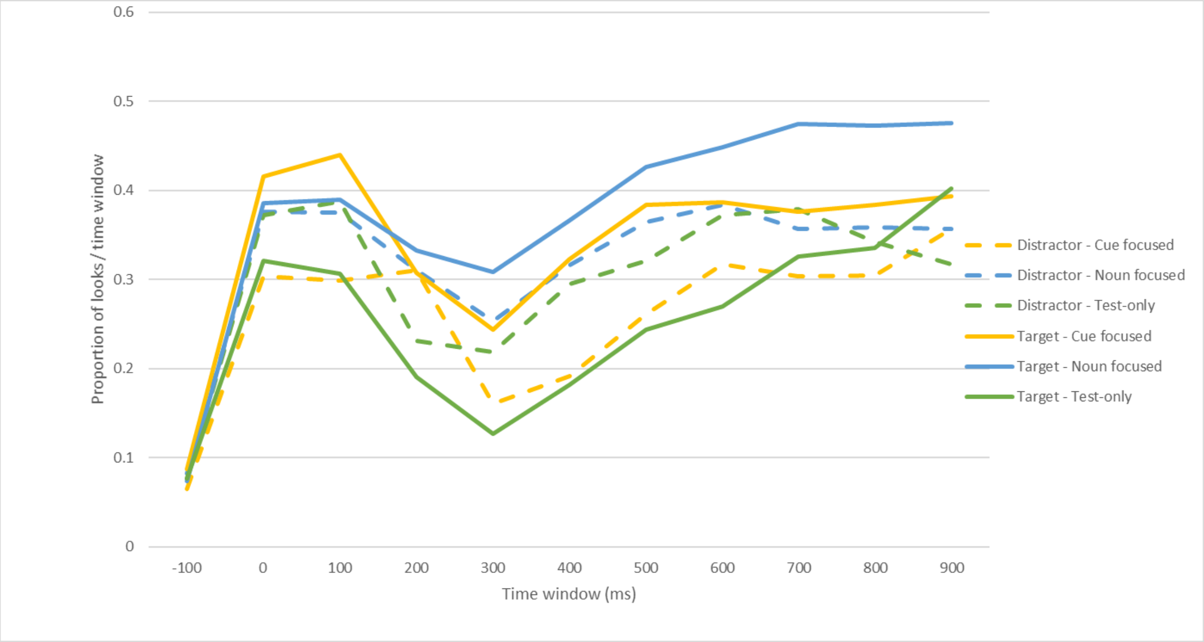


Figure 3. Immediate post-test - full passives and group

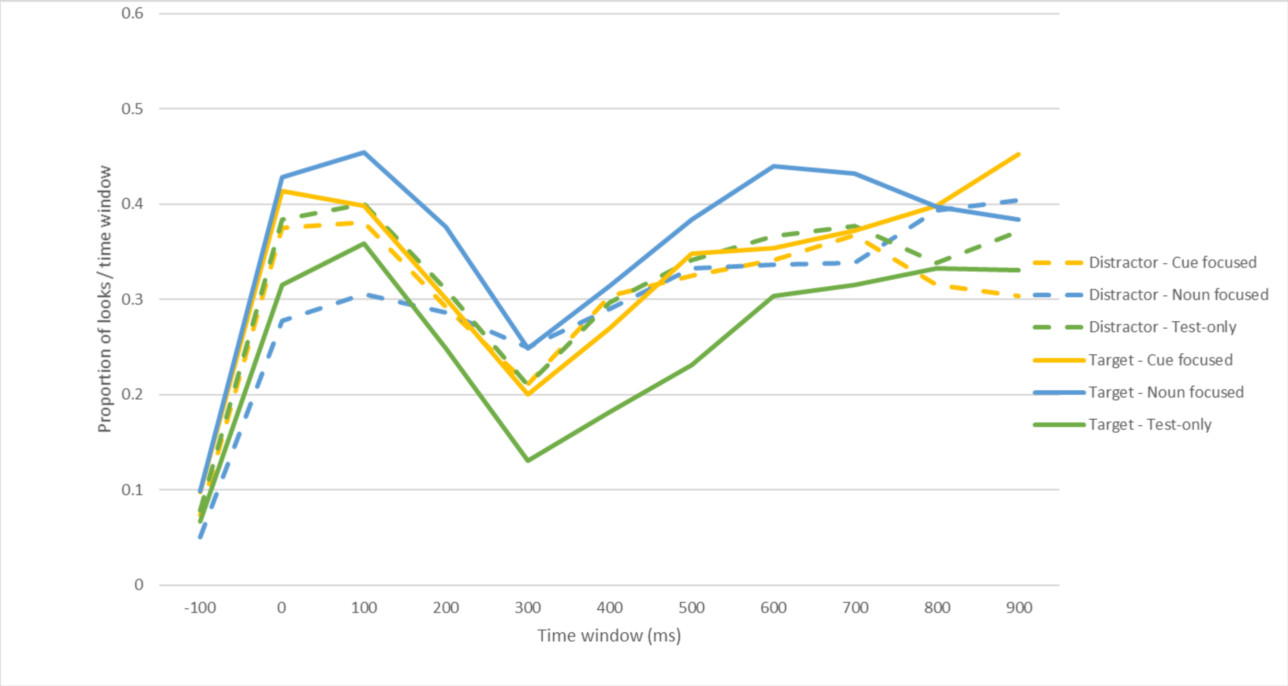


Figure 4. Immediate post-test - reduced passives and group

Reduced / full passives – Delayed post-test

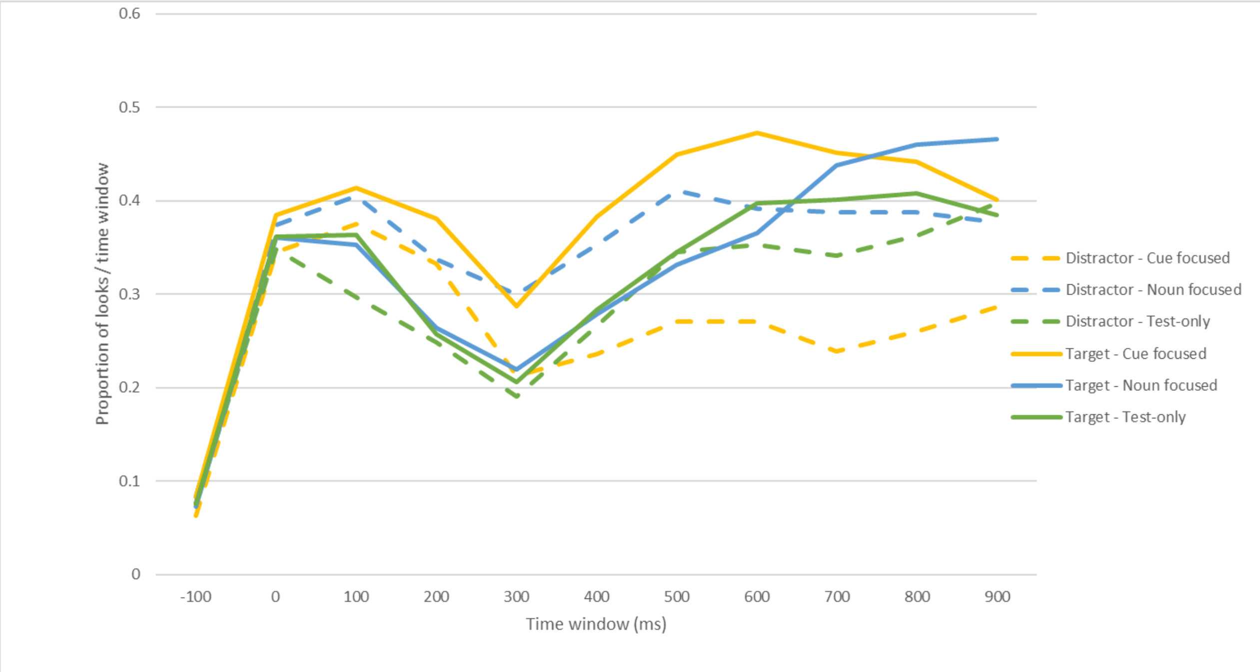


Figure 5. Delayed post-test - full passives and group

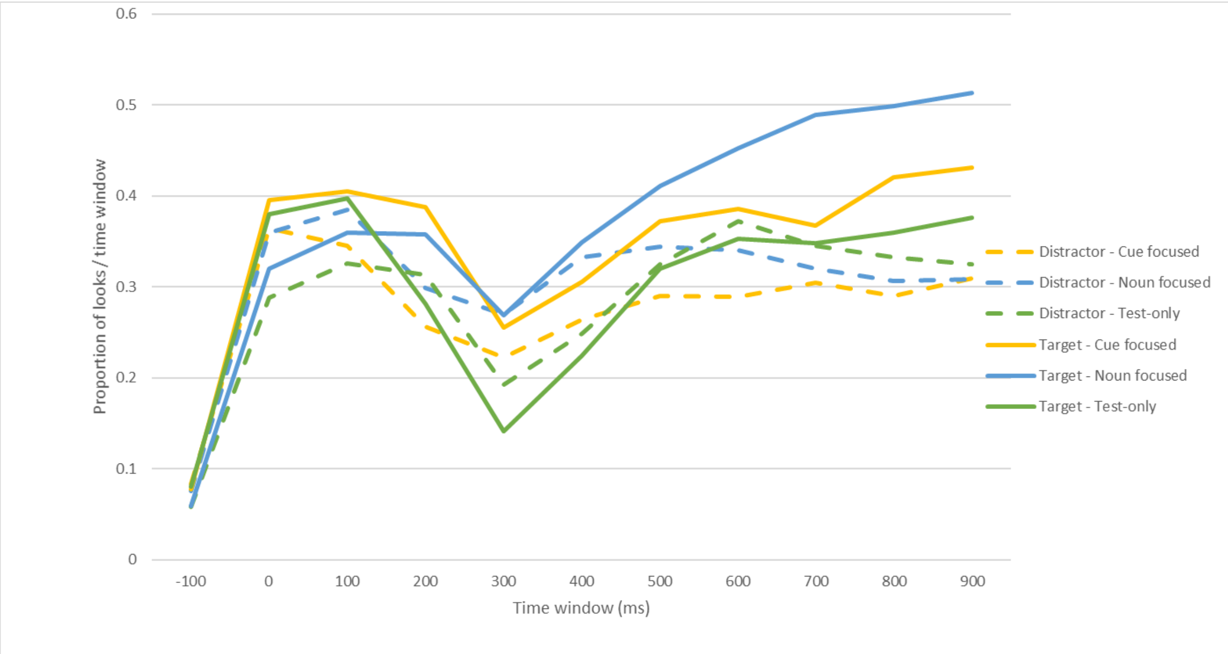


Figure 6. Delayed post-test - reduced passives and group

Appendix 22 – Oral production test - model results

*lmer(score ~ group_sum * time_sum * voice_sum * animacy_sum * tense_sum * reverse_sum + (1 | subject) + (voice |item) + (time | version), data=new)*

Main effect / interaction	Estimate	SE	*t*	*p*
Intercept	1.635	.070	23.467	.00 ***
Group – noun focused	.090	.083	1.086	.277
Group – cue focused	-.107	.073	-1.469	.142
Immediate post-test	.076	.087	.869	.387
Delayed post-test	.092	.064	1.436	.151
Voice (passive)	.136	.059	2.326	.020 *
Reversible	-.003	.032	-.080	.936
Animacy A-I	.000	.056	.007	.995
Animacy I-A	.059	.106	.548	.584
Trained / untrained	-.052	.042	-1.245	.214

Group – noun focused * immediate post-test	-.1408	.0983	-1.432	.152
Group – cue focused * immediate post-test	-.048	.090	-.534	.594
Group – noun focused * delayed post-test	.127	.103	1.229	.219
Group – cue focused * delayed post-test	.129	.092	1.396	.163
Group – noun focused * voice (passive)	-.043	.079	-.544	.586
Group – cue focused * voice (passive)	.011	.069	.164	.870
Immediate post-test * voice (passive)	.003	.081	.038	.969
Delayed post-test * voice (passive)	-.038	.061	-.622	.534
Group – noun focused * reversible	-.022	.024	-.897	.370
Group – cue focused * reversible	.010	.027	.370	.712
Immediate post-test * reversible	.021	.024	.888	.374
Delayed post-test * reversible	-.003	.025	-.131	.896
Voice (passive) * reversible	-.090	.036	-2.507	.013 *
Group – noun focused * A-I	-.054	.082	-.658	.510

Group – cue focused * A-I	.064	.071	.905	.366
Group – noun focused * I-A	.096	.157	.610	.542
Group – cue focused * I-A	-.001	.133	-.006	.995
Immediate post-test * A-I	-.015	.084	-.174	.862
Delayed post-test * A-I	.007	.066	.102	.919
Immediate post-test * I-A	-.009	.161	-.054	.957
Delayed post-test * I-A	-.049	.121	-.402	.688
Voice (passive) * A-I	.0413	.056	.733	.464
Voice (passive) * I-A	-.007	.101	-.065	.948
Untrained * A-I	-.002	.036	-.064	.949
Trained * I-A	-.053	.065	-.812	.417
Group – noun focused * immediate post-test * voice	.078	.093	.845	.398
Group – cue focused * immediate post-test * voice	-.015	.096	-.153	.878
Group – noun focused * delayed post-test * voice	.018	.086	.210	.834

Group – cue focused * delayed post-test * voice	-.045	.085	-.531	.595
Group – noun focused * immediate post-test * trained/untrained	.047	.032	1.467	.143
Group – cue focused * immediate post-test * trained/untrained	.000	.034	.008	.994
Group – noun focused * delayed post-test * trained/untrained	.035	.033	1.062	.288
Group – cue focused * delayed post-test * trained/untrained	-.018	.034	-.533	.594
Immediate post-test * A-I * voice	-.023	.083	-.271	.786
Delayed post-test * A-I * voice	-.053	.064	-.828	.408
Immediate post-test * I-A * voice	-.074	.157	-.470	.638
Delayed post-test * I-A * voice	.026	.116	.221	.825

Signif. codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 ' ' 1

Likelihood ratio test

Main effect	
Time	$\chi 2(127)=183.30$, $p=.00$
Group	$\chi 2(127)=153.91$, $p=.05$
Trained / untrained	$\chi 2(94)=111.74$, $p=.10$
Reversibility	$\chi 2(90)=82.07$, $p=.71$
Voice	$\chi 2(85)=170.39$, $p=.00$
Animacy	$\chi 2(212)=98.04$, $p=.94$

Appendix 23 – Oral production test mean scores and effect sizes (Cohen's *d*)

Effect sizes adjusted for baseline differences

Passive items

Table 1: Oral production test effect sizes corrected for baseline differences comparing mean scores for passives between groups (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.54	.47
Cue vs. test-only	.17	.12
Noun vs. test-only	-.35	-.40

Active items

Table 2: Oral production test effect sizes corrected for baseline differences comparing mean scores for actives between groups (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.36	.36
Cue vs. test-only	-.44	.45
Noun vs. test-only	-.91	.14

Correct verb inflection

Table 3: Oral production test effect sizes corrected for baseline differences comparing mean scores for correct verb usage between groups (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.18	.50
Cue vs. test-only	1.16	.88
Noun vs. test-only	1.06	.53

Use of by

Table 4: Oral production test effect sizes corrected for baseline differences comparing mean scores for correct use of by between groups (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.00	.00
Cue vs. test-only	.55	.09
Noun vs. test-only	.67	.06

Trained / untrained constructions

Table 5: Oral production test - effect sizes corrected for baseline differences comparing mean scores for trained and untrained items between groups (Cohen's d).

Group (n)		Immediate post (*d*, CIs)	Delayed post (*d*, CIs)
Cue focused vs. noun focused	untrained	-.43	-.53
	trained	.07	-.17
Cue focused vs. control	untrained	-.40	-.52
	trained	.42	.22
Noun focused vs. control	untrained	.00	.00
	trained	.89	.47

Table 6. Oral production test mean % score for trained (k =18) and untrained (k =18) items across time (SDs)

Group (*n*)	Construction	Pre	Immediate post	Delayed post
Cue (21)	untrained	83 (10)	82 (9)	84 (12)
	trained	84 (11)	89 (11)	89 (10)
Noun (24)	untrained	78 (9)	81 (11)	84 (10)
	trained	80 (10)	84 (11)	87 (9)

Test-only (19)	untrained	78 (9)	81 (7)	84 (11)
	trained	86 (6)	82 (11)	89 (7)

Table 7. Oral production test within-group effect size (Cohen's d) between test times for trained (simple) and untrained (perfect) items

Group (*n*)	Construction	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (21)	untrained	-.11 [-.69, .48 [a]]	.19 [-.40, .78 [a]]	.09 [-.49, .67 [a]]
	trained	.45 [-.13, 1.04 [a]]	.00 [-.60, .60 [a]]	.45 [-.13, 1.04 [a]]
Noun (24)	untrained	.30 [-.27, .87 [a]]	.29 [-.28, .85 [a]]	.63 [.05, 1.21]
	trained	.38 [-.19, .95 [a]]	.30 [-.27, .87 [a]]	.74 [.15, 1.32]
Test-only (19)	untrained	.37 [-.21, .96 [a]]	.33 [-.27, .92 [a]]	.60 [-.01, 1.20 [a]]
	trained	-.45 [-1.05, .15 [a]]	.76 [.15, 1.37]	.46 [-.14, 1.06 [a]]

[a] 95% confidence intervals pass through zero

Table 8. Oral production test between-groups effect sizes (Cohen's d) for trained and untrained items

Group (*n*)		Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	untrained	.53 [-.07, 1.12 [a]]	.10 [-.49, .68 [a]]	.00 [-.59, .59 [a]]
	trained	.38 [-.21, .97 [a]]	.45 [-.14, 1.05 [a]]	.21 [-.38, .80 [a]]
Cue vs. test-only	untrained	.52 [-.11, 1.16 [a]]	.12 [-.50, .74 [a]]	.00 [-.62, .62 [a]]
	trained	-.22 [-.85, .40 [a]]	.64 [.00, 1.27]	.00 [-.62, .62 [a]]
Noun vs. test-only	untrained	.00 [-.60, .60 [a]]	.00 [-.60, .60 [a]]	.00 [-.60, .60 [a]]
	trained	-.71 [-1.33, -.09]	.18 [-.42, .78 [a]]	-.24 [-.85, .36 [a]]

[a] 95% confidence intervals pass through zero

Effect sizes for untrained actives and passives

Table 9: Oral production test % means (SDs) untrained passive and active sentences (Cohen's d).

Group (*n*)	Voice	Pre	Immediate post	Delayed post
Cue (21)	Passive	71 (17)	83 (11)	82 (10)
	Active	87 (13)	92 (10)	95 (6)
Noun (24)	Passive	68 (16)	74 (14)	73 (15)
	Active	89 (10)	92 (7)	92 (10)
Test-only (19)	Passive	70 (22)	82 (11)	84 (11)
	Active	84 (20)	90 (8)	88 (14)

Table 10: Oral production test within-subjects effect sizes comparing untrained passive and active sentences (Cohen's d).

Group (*n*)	Voice	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (21)	Passive	.82 [.21, 1.47]	-.10 [-.70, .15 [a]]	.79 [.16, 1.42]
	Active	.43 [-.18, 1.04 [a]]	.36 [-.25, .97 [a]]	.79 [.16, 1.42 [a]]
Noun (24)	Passive	.40 [-.17. .97 [a]]	-.07 [-.64, .50 [a]]	.32 [-.25, .89 [a]]
	Active	.35 [-.22, .92 [a]]	.00 [-.57, .57 [a]]	.30 [-.27, .87 [a]]
Test-only (19)	Passive	.69 [.04, 1.35]	.182 [-.46, .82 [a]]	.80 [.14, 1.47]

	Active	.39 [-.25, 1.04 [a]]	-.175 [-.81, .46 [a]]	.23 [-.41, .87 [a]]

[a] 95% confidence intervals pass through zero

Table 11: Oral production test between-groups effect sizes comparing untrained passive and active sentences (Cohen's d).

Group (*n*)		Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	Passive	.18 [-.41, .77 [a]]	.71 [.12, 1.31]	.70 [.09, 1.30]
	Active	-.17 [-.76, .41 [a]]	.00 [-.59. .59 [a]]	.36 [-.232, .95 [a]]
Cue vs. test-only	Passive	.05 [-.57, .67 [a]]	.09 [-.53, .71 [a]]	-.19 [-.81, .43 [a]]
	Active	.18 [-.44, .80 [a]]	.22 [-.40, .84 [a]]	.66 [.03, 1.30]
Noun vs. test-only	Passive	-.11 [-.70, .50 [a]]	-.63 [-1.24, -.01 [a]]	-.82 [-1.45, -.20 [a]]
	Active	.33 [-.28, .93 [a]]	.27 [-.34, .87 [a]]	.34 [-.27, .94 [a]]

[a] 95% confidence intervals pass through zero

Table 12: Oral production test between-groups effect sizes comparing untrained passive and active sentences adjusted for baseline differences (Cohen's d).

Group (*n*)		Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	Passive	.53	.52
	Active	.17	.53
Cue vs. test-only	Passive	.04	-.24
	Active	.04	.48
Noun vs. test-only	Passive	-.52	-.71
	Active	-.06	.01

Appendix 24 – Writing test model results

*lmer(score ~ group_sum * time_sum * voice_sum * animacy_sum * tense_sum * reverse_sum + (1 | subject) + (voice |item) + (time | version), data=new)*

Main effect / interaction	Estimate	SE	*t*	*p*
Intercept	1.692	.050	33.982	.000 ***
Group – noun focused	.054	.041	1.312	.190
Group – cue focused	-.075	.040	-1.868	.062 .
Immediate post-test	-.003	.049	-.056	.955
Delayed post-test	-.007	.048	-.144	.886
Voice (passive)	.125	.039	3.157	.002 **
Reversible	.003	.028	.122	.903
Animacy A-I	-.002	.033	-.069	.945
Animacy I-A	.039	.063	.615	.538
Trained / untrained	-.053	.041	-1.291	.197

Group – noun focused * immediate post-test	.063	.059	1.064	.287
Group – cue focused * immediate post-test	-.078	.046	-1.67	.165
Group – noun focused * delayed post-test	-.085	.061	-1.389	.092 .
Group – cue focused * delayed post-test	.042	.049	.846	.398
Group – noun focused * voice (passive)	-.037	.034	-1.091	.276
Group – cue focused * voice (passive)	.045	.033	1.362	.173
Immediate post-test * voice (passive)	-.042	.050	-.839	.402
Delayed post-test * voice (passive)	.041	.044	.933	.351
Group – noun focused * reversible	-.003	.013	-.264	.792
Group – cue focused * reversible	-.001	.013	-.072	.942
Immediate post-test * reversible	-.006	.015	-.414	.679
Delayed post-test * reversible	.015	.013	1.189	.234
Voice (passive) * reversible	-.044	.027	-1.639	.104
Group – noun focused * A-I	-.045	.036	-1.225	.221

Group – cue focused * A-I	.053	.036	1.472	.141
Group – noun focused * I-A	.094	.069	1.362	.173
Group – cue focused * I-A	-.055	.066	-.834	.405
Immediate post-test * A-I	-.051	.049	-1.034	.301
Delayed post-test * A-I	.015	.046	.322	.747
Immediate post-test * I-A	.126	.097	1.294	.196
Delayed post-test * I-A	-.043	.089	-.483	.629
Voice (passive) * A-I	.016	.034	.470	.638
Voice (passive) * I-A	-.023	.061	-.375	.708
Untrained * A-I	-.028	.033	-.842	.400
Trained * I-A	.032	.062	.517	.605
Group – noun focused * immediate post-test * voice	-.032	.057	-.571	.568
Group – cue focused * immediate post-test * voice	.020	.057	.351	.726
Group – noun focused * delayed post-test * voice	.047	.043	1.083	.279

Group – cue focused * delayed post-test * voice	.005	.045	.117	.907
Group – noun focused * immediate post-test * trained/untrained	-.027	.019	-1.429	.153
Group – cue focused * immediate post-test * trained/untrained	.006	.020	.293	.770
Group – noun focused * delayed post-test * trained/untrained	.019	.018	1.102	.271
Group – cue focused * delayed post-test * trained/untrained	.006	.018	.347	.729
Immediate post-test * A-I * voice	.079	.051	1.549	.121
Delayed post-test * A-I * voice	-.024	.046	-.521	.602
Immediate post-test * I-A * voice	-.090	.093	-.977	.328
Delayed post-test * I-A * voice	-.005	.086	-.054	.957

Signif. codes: 0 ‘***’ .001 ‘**’ .01 ‘*’ .05 ‘.’ .1 ‘ ’ 1

Likelihood ratio test

Main effect	
Time	$\chi2(127)=183.3$, $p=.00$
Group	$\chi2(127)=153.91$, $p=.05$
Trained / untrained	$\chi2(94)=111.74$, $p=.10$
Reversibility	$\chi2(90)=82.067$, $p=.71$
Voice	$\chi2(85)=170.39$, $p=.00$
Animacy	$\chi2(121)=98.04$, $p=.94$

Appendix 25 – Writing test mean scores and effect sizes (Cohen's d)

Effect sizes adjusted for baseline differences

Passive items

Table 1: Writing test effect sizes corrected for baseline differences comparing mean scores for passives between groups (Cohen's d).

Group (*n*)	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	-.14	-.05
Cue vs. test-only	.88	.32
Noun vs. test-only	1.04	.41

Active items

Table 2: Writing test effect sizes corrected for baseline differences comparing mean scores for active between groups (Cohen's d).

Group (*n*)	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	.00	-.15
Cue vs. test-only	.00	-.16
Noun vs. test-only	.00	.00

Correct verb inflection

Table 3: Writing test effect sizes corrected for baseline differences comparing mean scores for correct verb form between groups (Cohen's d).

Group (*n*)	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	.35	.82
Cue vs. test-only	.44	.65
Noun vs. test-only	.10	-.12

Production of by

Table 4: Writing test effect sizes corrected for baseline differences comparing mean scores for correct use of by between groups (Cohen's d).

Group (*n*)	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	.12	.09
Cue vs. test-only	.04	.76
Noun vs. test-only	-.21	.54

Untrained/trained

Table 5. Writing test mean % score for trained and untrained items across time (SDs) (k=18)

Group (*n*)		Pre	Immediate post	Delayed post
Cue (22)	untrained	79 (12)	87 (7)	89 (5)
	trained	79 (11)	87 (7)	91 (6)
Noun (24)	untrained	78 (11)	83 (8)	82 (9)
	trained	78 (8)	83 (9)	85 (8)
Test-only (22)	untrained	77 (20)	86 (7)	86 (9)
	trained	78 (14)	86 (9)	86 (8)

Table 6. Writing test within-group effect size (Cohen's d) for trained / untrained items

Group (*n*)		Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (22)	untrained	.81 [.20, 1.43]	.33 [-.27, .92 [a]]	1.09 [.45, 1.72]
	trained	.87 [.25, 1.49]	.61 [.01, 1.22]	1.35 [.70, 2.01]
Noun (24)	untrained	.52 [-.06, 1.10 [a]]	-.12 [-.68, .45 [a]]	.40 [-.17, .97 [a]]
	trained	.59 [.01, 1.17]	.23 [-.33, .80 [a]]	.88 [.28, 1.47]
Test-only (22)	untrained	.60 [-.12, 1.22 [a]]	.00 [-.59, .59 [a]]	.58 [-.04, 1.20 [a]]
	trained	.68 [.06, 1.30]	.00 [-.59, .59 [a]]	.70 [.08, 1.32]

[a] 95% confidence intervals pass through zero

Table 7. Writing test between-groups effect sizes (Cohen's d) for trained / untrained items

Group (*n*)		Pre *d* [CIs]	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	untrained	.09 [-.49, .67 [a]]	.53 [-.06, 1.12 [a]]	.95 [.34, 1.56]
	trained	.10 [-.47, .68 [a]]	.49 [-.09, 1.08 [a]]	.84 [-.24, 1.45 [a]]
Cue vs. test-only	untrained	.12 [-.47, .71 [a]]	.12 [-.47, .72 [a]]	.41 [-.19, 1.01 [a]]
	trained	.08 [-.51, .67 [a]]	.12 [-.47, .72 [a]]	.65 [.05, 1.26]
Noun vs. test-only	untrained	.06 [-.52, .64 [a]]	-.35 [-.94, .23 [a]]	-.44 [-1.03, .14 [a]]
	trained	.00 [-.58, .58 [a]]	-.33 [-.92, .25 [a]]	-.12 [-.70, .46 [a]]

[a] 95% confidence intervals pass through zero

Table 8. Writing test % means (SDs) untrained passive and active sentences (Cohen's d).

Group (*n*)	Construction	Pre	Immediate post	Delayed post
Cue (21)	Passive	75 (17)	76 (15)	81 (16)
	Active	87 (10)	88 (10)	88 (15)
Noun (24)	Passive	70 (17)	74 (18)	78 (15)
	Active	86 (10)	89 (9)	89 (10)
Test-only (19)	Passive	72 (14)	71 (16)	80 (14)
	Active	86 (8)	90 (7)	89 (8)

Table 9. Writing test within-subjects effect sizes comparing untrained passive and active sentences (Cohen's d).

Group (*n*)	Construction	Pre vs. immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (21)	Passive	.06 [-.54, .67[a]]	.32 [-.29, .93[a]]	.36 [-.25, .97[a]]
	Active	.10 [-.51, .71[a]]	.00 [-.61, .61[a]]	.08 [-.53, .68[a]]
Noun (24)	Passive	.23 [-.334, .79[a]]	.24 [-.33, .81[a]]	.50 [-.08, 1.07[a]]
	Active	.32 [-.25, .89[a]]	.00 [-.57, .57[a]]	.30 [-.27, .87[a]]
Test-only (19)	Passive	-.07 [-.70, .57[a]]	.60 [-.05, 1.25[a]]	.57 [-.08, 1.22[a]]
	Active	.53 [-.12, 1.18[a]]	-.13 [-.77, .50[a]]	.38 [-.27, 1.02[a]]

[a] 95% confidence intervals pass through zero

Table 10. Writing test between-groups effect sizes comparing untrained passive and active sentences (Cohen's d).

Group (*n*)		Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	Passive	.29 [-.30, .88[a]]	.12 [-.47, .71[a]]	.19 [-.39, .78[a]]
	Active	.10 [-.49, .69[a]]	-.11 [-.69, .48[a]]	-.08 [-.67, .51[a]]
Cue vs. test-only	Passive	.19 [-.43, .81[a]]	.32 [-.30, .95[a]]	.07 [-.56, .69[a]]
	Active	.11 [-.51, .73[a]]	-.23 [-.85, .39[a]]	-.08 [-.70, .54[a]]
Noun vs. test-only	Passive	-.13 [-.73, .48 [a]]	.17 [-.43, .78[a]]	-.14 [-.74, .47[a]]
	Active	.00 [-.60, .60[a]]	-.12 [-.73, .48[a]]	.00 [-.60, .60 [a]]

[a] 95% confidence intervals pass through zero

Table 11. Writing test between-groups effect sizes comparing untrained passive and active sentences adjusted for baseline differences (Cohen's d).

Group (*n*)		Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	Passive	-.17	-.10
	Active	-.21	-.18
Cue vs. test-only	Passive	.13	-.12
	Active	-.34	-.19
Noun vs. test-only	Passive	.30	-.01
	Active	-.12	.00

Appendix 26 – JT model results

*glmer(score ~ group * voice * tense * animacy * grammatical + (1 | subject) + (1 + voice | item), data=new, family=binomial, control = glmer, Control(optimizer = "bobyqa"))*

Time 1

Main effect / interaction	Estimate	SE	*z*	*p*
Intercept	2.127	2.814	.756	.450
Group – noun focused	.529	5.623	.094	.925
Group – cue focused	-.103	2.814	-.037	.971
Voice (passive)	.337	2.813	.120	.904
Animate-inanimate	-.384	2.816	-.136	.891
Inanimate-animate	-.326	2.814	-.116	.908
Inanimate-inanimate	.489	3.936	.124	.901
Trained / untrained	-.532	2.812	-.189	.850
Grammaticality	.043	2.659	.016	.987
Grammaticality * voice	-.353	2.659	-.133	.894
Group – noun focused * voice (passive)	.577	5.622	.103	.918

Group – cue focused * voice (passive)	-.215	2.812	-.076	.939
Group – noun focused * grammaticality	.011	5.316	.002	.998
Group – cue focused * grammaticality	.112	2.659	.042	.966
Animate-inanimate * voice	-.417	2.816	-.148	.882
Inanimate-animate * voice	-.398	2.814	-.141	.888
Inanimate-inanimate * voice	.445	3.936	.113	.910
Animate-inanimate * Group – noun focused	-.659	5.623	-.117	.907
Inanimate-animate * Group – noun focused	-.537	5.623	-.096	.924
Inanimate-inanimate * Group – noun focused	.6391	7.867	.081	.935
Animate-inanimate * Group – cue focused	.351	2.815	.125	.901
Inanimate-animate * Group – cue focused	.389	2.814	.138	.890
Inanimate-inanimate * Group – cue focused	-.513	3.936	-.130	.896
Group – noun focused * trained / untrained	-.628	5.622	-.112	.911
Group – cue focused * trained / untrained	.158	2.812	.056	.955
Voice * trained / untrained	-.488	2.812	-.173	.862

Signif. codes: 0 ‘***’ .001 ‘**’ .01 ‘*’ .05 ‘.’ .1 ‘ ’ 1

Likelihood ratio test

Main effect	
Group	$\chi2(64)=80.25$, $p=.08$
Trained / untrained	$\chi2(48)=87.79$, $p=.00$
Animacy	$\chi2(72)=56.91$, $p=.90$
Voice	$\chi2(48)=79.41$, $p=.00$
Grammaticality	$\chi2(48)=92.00$, $p=.00$

Time 2

Main effect / interaction	Estimate	SE	*z*	*p*
Intercept	2.983	31.709	.094	.925
Group – noun focused	.124	5.172	.002	.998
Group – cue focused	-.947	31.710	-.030	.976
Voice (passive)	.615	31.709	.019	.985
Animate-inanimate	-.293	53.471	-.005	.996
Inanimate-animate	-.554	31.709	-.017	.986

Inanimate-inanimate	.540	59.268	.009	.993
Trained / untrained	-.629	31.709	-.020	.984
Grammaticality	.448	31.709	.014	.989
Grammaticality * voice	-.165	31.709	-.005	.996
Group – noun focused * voice (passive)	.356	5.173	.007	.994
Group – cue focused * voice (passive)	-.911	31.709	-.029	.977
Group – noun focused * grammaticality	-.216	5.173	-.004	.997
Group – cue focused * grammaticality	.107	31.709	.003	.997
Animate-inanimate * voice	-.487	53.471	-.009	.993
Inanimate-animate * voice	-.650	31.709	-.020	.984
Inanimate-inanimate * voice	.568	59.268	.010	.992
Animate-inanimate * Group – noun focused	1.123	99.658	.011	.991
Inanimate-animate * Group – noun focused	-.158	5.173	-.003	.997
Inanimate-inanimate * Group – noun focused	-1.358	7.884	-.019	.985
Animate-inanimate * Group – cue focused	.341	53.471	.006	.995
Inanimate-animate * Group – cue focused	.453	31.710	.014	.989
Inanimate-inanimate * Group – cue focused	-.462	59.268	-.008	.994
Group – noun focused * trained / untrained	.332	5.173	.007	.995

Group – cue focused * trained / untrained	.537	31.709	.017	.986
Voice * trained / untrained	-.548	31.709	-.017	.986

Likelihood ratio test

Main effect	
Group	χ2(64)=96.212, *p*=.00 **
Trained / untrained	χ2(48)=70.167, *p*=.02 *
Animacy	χ2(72)=87.87, *p*=.10 .
Voice	χ2(48)=71.053, *p*=.02 *
Grammaticality	χ2(48)=85.107, *p*=.00 ***

Time 3

Main effect / interaction	Estimate	SE	*z*	*p*
Intercept	2.841	5.504	.516	.606
Group – noun focused	-.042	7.091	-.006	.995
Group – cue focused	-.663	5.505	-.120	.904

Voice (passive)	.364	4.778	.076	.939
Animate-inanimate	.971	1.378	.094	.925
Inanimate-animate	.050	6.887	.007	.994
Inanimate-inanimate	-.720	5.504	-.131	.896
Trained / untrained	-.023	5.300	-.004	.997
Grammaticality	.424	4.952	.086	.932
Grammaticality * voice	-.373	5.300	-.070	.944
Group – noun focused * voice (passive)	.498	6.647	.075	.940
Group – cue focused * voice (passive)	-.405	4.778	-.085	.932
Group – noun focused * grammaticality	-.179	6.769	-.026	.979
Group – cue focused * grammaticality	.098	4.952	.020	.984
Animate-inanimate * voice	1.486	11.114	.134	.894
Inanimate-animate * voice	-.962	8.355	-.115	.908
Inanimate-inanimate * voice	-.558	4.779	-.117	.907
Animate-inanimate * Group – noun focused	1.975	16.671	.118	.906
Inanimate-animate * Group – noun focused	-.907	8.376	-.108	.914
Inanimate-inanimate * Group – noun focused	-.249	7.090	-.035	.972
Animate-inanimate * Group – cue focused	-1.268	1.378	-.122	.903

Inanimate-animate * Group – cue focused	.212	6.887	.031	.975
Inanimate-inanimate * Group – cue focused	.841	5.505	.153	.879
Group – noun focused * trained / untrained	-.125	7.062	-.018	.986
Group – cue focused * trained / untrained	-.111	5.300	-.021	.983
Voice * trained / untrained	-.393	4.952	-.079	.937

Likelihood ratio test

Main effect	
Group	$\chi 2(64)=64.582$, $p=.45$
Trained / untrained	$\chi 2(48)=136.96$, $p=.00$
Animacy	$\chi 2(72)=77.815$, $p=.30$
Voice	$\chi 2(48)=156.79$, $p=.00$ ***
Grammaticality	$\chi 2(48)=159.75$, $p=.00$ ***

Error identification for ungrammatical items

*Glmer (score ~ group * time * passive * error + (1 | subject) + (1 | item), data=new, family=binomial, control = glmerControl(optimizer = "bobyqa"))*

Main effect / interaction	Estimate	SE	*z*	*p*
Intercept	2.152e+00	6.901e-01	3.118	.002 **
Immediate post-test	2.287e+00	1.189e+00	1.924	.054 .
Delayed post-test	2.147e+01	6.902e+03	.003	.998
Group – noun focused	6.149e-01	9.013e-01	.682	.495
Group – cue focused	-3.933e-02	8.438e-01	-.047	.963
Voice (passive)	-2.918e-01	7.303e-01	-.400	.690
Error – ing	6.656e-01	8.569e-01	.777	.437
Error – missing be	4.391e-01	8.272e-01	.531	.595
Group – noun focused * voice (passive)	1.064e+00	1.037e+00	1.025	.305
Group – cue focused * voice (passive)	1.753e+00	9.897e-01	1.772	.076 .
Group – noun focused * immediate post-test	-1.818e+00	1.459e+00	-1.246	.213
Group – noun focused * delayed post-test	-1.849e+00	1.001e+04	.000	1.000
Group – cue focused * immediate post-test	1.754e+01	7.257e+03	.002	.998
Group – cue focused * delayed post-test	-1.507e+00	1.005e+04	.000	1.000

Error – ing * voice	-2.305e-01	1.025e+00	-.225	.822
Error – missing be * voice	1.003e+00	1.064e+00	.943	.346
Error – ing * cue focused	5.062e-01	1.064e+00	.476	.634
Error – missing be * noun focused	2.397e-01	1.155e+00	.208	.836
Error – ing * immediate post-test	1.789e+01	6.262e+03	.003	.998
Error – missing be * delayed post-test	-1.909e+01	6.902e+03	-.003	.998
Noun focused * voice * immediate post-test	-1.957e+01	5.407e+03	-.004	.997
Noun focused * voice * delayed post-test	6.848e-02	1.001e+04	.000	1.000
Cue focused * voice * immediate post-test	-1.976e+01	1.058e+04	-.002	.999
Cue focused * voice * delayed post-test	2.226e-01	1.005e+04	.000	1.000
Noun focused * missing be * immediate post-test	-1.910e+01	6.414e+03	-.003	.998
Noun focused * missing be * delayed post-test	1.939e+01	1.156e+04	.002	.999
Noun focused * ing * immediate post-test	-1.874e+01	6.262e+03	-.003	.999
Noun focused * ing * delayed post-test	1.925e+01	1.179e+04	.002	.999
Cue focused * missing be * immediate post-test	-1.857e+01	1.210e+04	-.002	.999
Cue focused * missing be * delayed post-test	1.878e+00	1.005e+04	.000	1.000
Cue focused * ing * immediate post-test	-1.826e+01	1.151e+04	-.002	.999
Cue focused * ing * delayed post-test	1.459e+00	1.005e+04	.000	1.000

Appendix 27 – JT test mean scores and effect sizes (Cohen's *d*)

Acceptability judgements of passives - effect sizes adjusted for baseline differences

Table 1. JT effect sizes corrected for baseline differences comparing mean scores for passives between-groups (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.76	.43
Cue vs. test-only	.33	.46
Noun vs. test-only	-.46	.10

Acceptability judgements of actives - effect sizes adjusted for baseline differences

Table 2. JT effect sizes corrected for baseline differences comparing mean scores for actives between-groups (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.66	-.09
Cue vs. test-only	.26	.27
Noun vs. test-only	.39	.34

Acceptability judgements by grammatical or ungrammatical items - effect sizes adjusted for baseline differences

Table 3. JT between subjects effect sizes corrected for baseline differences for acceptability judgements: passive and active items and grammaticality (Cohen's d).

Group (*n*)	Type	Grammaticality	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	Passives	Grammatical	.12	.22
		Ungrammatical	.82	.54
	Actives	Grammatical	.57	-.02
		Ungrammatical	.51	.19
Cue vs. test-only	Passives	Grammatical	-.14	.14
		Ungrammatical	.61	.55
	Actives	Grammatical	.00	.08
		Ungrammatical	1.08	.70
Noun vs. test-only	Passives	Grammatical	-.21	-.06
		Ungrammatical	-.33	-.04
	Actives	Grammatical	-.29	.29
		Ungrammatical	-.45	.03

Acceptability judgements by error type - effect sizes adjusted for baseline differences

Table 4. JT between subject effect sizes corrected for baseline differences for error type (Cohen's d).

Group (*n*)	Error type	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	Base form	.55	.26
	Ing	.81	.62
	Missing be	.44	.36
Cue vs. test-only	Base form	.31	.42
	Ing	.14	-.01
	Missing be	.09	.03
Noun vs. test-only	Base form	-.23	.03
	Ing	-.74	-.68
	Missing be	-.37	-.34

Acceptability judgements and animacy effects

Animate first nouns

Table 5. JT mean % score across time (SDs) - animate first nouns and voice by group

		Pre		Immediate post		Delayed post	
Group (*n*)	Voice	A-A (*k*=12)	A-I (*k*=12)	A-A (*k*=12)	A-I (*k*=12)	A-A (*k*=12)	A-I (*k*=12)
Cue (21)	Passive	69 (33)	70 (24)	76(28)	86 (16)	87 (12)	91 (12)
	Active	80 (16)	80 (18)	87 (11)	92 (12)	85 (23)	87 (12)
Noun (22)	Passive	74 (30)	78 (21)	71 (32)	74 (19)	84 (14)	87 (16)
	Active	84 (18)	81 (19)	89 (10)	74 (21)	85 (23)	88 (14)
Test-only (22)	Passive	66 (38)	78 (20)	78 (29)	82 (17)	87 (14)	83 (17)
	Active	80 (13)	78 (16)	86 (13)	80 (19)	89 (15)	84 (16)

Table 6. JT within-subject effect size (Cohen's d) between test times - animate first nouns and voice

		Pre vs. immediate post *d* [CIs]		Immediate vs. delayed post *d* [CIs]		Pre vs delayed post *d* [CIs]	
Group (*n*)		A-A (*k*=12)	A-I (*k*=12)	A-A (*k*=12)	A-I (*k*=12)	A-A (*k*=12)	A-I (*k*=12)
Cue focused (21)	Passive	.23 [-.378, .836 [a]]	.78 [.157, 1.412]	.51 [-.104, 1.125 [a]]	.35 [-.256, .963 [a]]	.72 [.101, 1.349]	1.12 [.457, 1.756]
	Active	.51 [-.105, 1.125 [a]]	.78 [.157, 1.412]	-.11 [-.716, .494 [a]]	-.41 [-1.028, .195 [a]]	.25 [-.355, .860 [a]]	.46 [-.155, 1.070 [a]]
Noun focused (22)	Passive	-.10 [-.695, .502 [a]]	-.20 [-.792, .393 [a]]	.52 [-.086, 1.130 [a]]	.74 [.129, 1.351]	.42 [-.181, 1.029 [a]]	.48 [-.117, 1.082 [a]]
	Active	.34 [-.252, .939 [a]]	-.35 [-.945, .246 [a]]	-.08 [-.673, .509 [a]]	.50 [-.101, 1.010 [a]]	.19 [-.403, .782 [a]]	.24 [-.345, .841 [a]]
Test-only (22)	Passive	.36 [-.241, .951 [a]]	.22 [-.377, .808 [a]]	.39 [-.201, .910 [a]]	.06 [-.532, .650 [a]]	.73 [.123, 1.343]	.26 [-.324, .863 [a]]
	Active	.48 [-.120, 1.079 [a]]	.11 [-.478, .705 [a]]	.21 [-.379, -.806 [a]]	.23 [-.365, .821 [a]]	.64 [.035, 1.247 [a]]	.38 [-.221, .971 [a]]

[a] 95% confidence intervals pass through zero

Table 7. JT between-groups effect sizes (Cohen's d) - animate first nouns and voice by group

		Pre		Immediate post		Delayed post	
		d [CIs]		*d* [CIs]		*d* [CIs]	
Group (*n*)		A-A (*k*=12)	A-I (*k*=12)	A-A (*k*=12)	A-I (*k*=12)	A-A (*k*=12)	A-I (*k*=12)
Cue vs. noun	Passive	-.16 [-.76, .44 [a]]	-.36 [-.96, .25 [a]]	.17 [-.43, .77 [a]]	.68 [.07, 1.30]	.23 [-.37, .83 [a]]	.28 [-.32, .88 [a]]
	Active	-.23 [-.83, .37 [a]]	-.05 [-.65, .54 [a]]	-.19 [-.79, .41 [a]]	1.05 [.41, 1.68]	.00 [-.60, .60 [a]]	-.08 [-.67, .52 [a]]
Cue vs. test-only	Passive	.08 [-.51, .68 [a]]	-.36 [-.97, .24 [a]]	-.07 [-.67, .53 [a]]	.24 [-.36, .84 [a]]	.00 [-.60, .60 [a]]	.54 [-.07, 1.15 [a]]
	Active	.00 [-.60, .60 [a]]	.12 [-.48, .72 [a]]	.08 [-.52, .68 [a]]	.74 [.13, 1.37]	-.20 [-.81, .39 [a]]	.21 [-.39, .81 [a]]
Noun vs. test-only	Passive	.23 [-.36, .83 [a]]	.00 [-.59, .59 [a]]	-.23 [-.82, .36 [a]]	-.44 [-1.04, .15 [a]]	-.2 [-.81, .38 [a]]	.24 [-.35, .84 [a]]
	Active	.25 [-.34, .85 [a]]	.17 [-.42, .76 [a]]	.26 [-.33, .85 [a]]	-.30 [-.89, .29 [a]]	-.2 [-.80, .39 [a]]	.27 [-.33, .86 [a]]

[a] 95% confidence intervals pass through zero

Table 8. JT between-group effect sizes corrected for baseline differences: Animate first nouns and voice by group and across test time (Cohen's d).

		Immediate post *d* [CIs]		Delayed post *d* [CIs]	
Group (*n*)		A-A	A-A	A-A	A-A
Cue vs. noun	Passive	.33	1.04	.39	.44
	Active	.04	1.28	.23	.15
Cue vs. test-only	Passive	-.15	.60	-.08	.90
	Active	.08	.62	-.21	-.09
Noun vs. test-only	Passive	.00	-.44	-.44	.24
	Active	.01	-.47	-.46	.10

Inanimate first nouns

Table 9. JT mean % score across time (SDs) - inanimate first nouns and voice by group

		Pre		Immediate post		Delayed post	
Group (*n*)	Voice	I-I (*k*=12)	I-A (*k*=12)	I-I (*k*=12)	I-A (*k*=12)	I-I (*k*=12)	I-A (*k*=12)
Cue focused (21)	Passive	73 (23)	80 (22)	79 (27)	87 (18)	89 (9)	92 (9)
	Active	75 (17)	80 (16)	83 (16)	87 (14)	77 (25)	83 (23)

Noun focused (22)	Passive	79 (17)	78 (21)	81 (29)	77 (21)	87 (15)	89 (15)
	Active	80 (23)	80 (16)	81 (20)	83 (17)	85 (19)	87 (22)
Test-only (22)	Passive	73 (23)	85 (19)	82 (19)	81 (26)	86 (11)	84 (17)
	Active	76 (14)	80 (16)	81 (16)	81 (14)	74 (22)	85 (17)

Table 10. JT within-subject effect size (Cohen's d) between test times - inanimate first nouns and voice

Groups (*n*)		Pre vs Immediate post *d* [CIs]		Immediate vs delayed post *d* [CIs]		Pre vs delayed post *d* [CIs]	
		I-I (*k*=12)	I-A (*k*=12)	I-I (*k*=12)	I-A (*k*=12)	I-I (*k*=12)	I-A (*k*=12)
Cue focused (21)	Passive	.24 [-.35, .83 [a]]	.35 [-.24, .94 [a]]	.50 [-.10, 1.10 [a]]	.35 [-.26, .96]	.92 [.30, 1.54]	.71 [.11, 1.32]
	Active	.48 [-.13, 1.10 [a]]	.47 [-.15, 1.08 [a]]	.29 [-.32, .89 [a]]	-.21 [-.81, .40 [a]]	-.15 [-.76, .46 [a]]	.15 [-.45, .76 [a]]
Noun focused (22)	Passive	.08 [-.51, .68 [a]]	-.04 [-.64, .54 [a]]	.26 [-.33, .85 [a]]	.66 [.05, 1.26]	.50 [-.10, 1.10 [a]]	.60 [-.00, 1.21[a]]
	Active	.05 [-.55, .64 [a]]	.18 [-.410, .77 [a]]	.21 [-.39, .80 [a]]	.20 [-.39, .80 [a]]	.24 [-.36, .83 [a]]	.36 [-.23, .96 [a]]
Test-only (22)	Passive	.43 [-.17, 1.02 [a]]	-.18 [-.77, .42 [a]]	.26 [-.37 .85 [a]]	.14 [-.46, .73 [a]]	.72 [.11, 1.33]	-.06 [-.65, .54 [a]]
	Active	.33 [-.26, .93 [a]]	.07 [-.53, .66 [a]]	-.36 [-.96, .23 [a]]	.26 [-.34, .85 [a]]	-.11 [-.70, .48 [a]]	.30 [-.29, .90 [a]]

[a] 95% confidence intervals pass through zero

Table 11. JT between-groups effect sizes (Cohen's d) - animate first nouns and voice

Group (*n*)		Pre-test *d* [CIs]		Immediate post-test *d* [CIs]		Delayed post-test *d* [CIs]	
		I-I (*k*=12)	I-A (*k*=12)	I-I (*k*=12)	I-A (*k*=12)	I-I (*k*=12)	I-A (*k*=12)
Cue vs. noun	Passive	-.30 [-.90, .30 [a]]	.09 [-.51, .69 [a]]	-.07 [-.67, .53 [a]]	.51 [-.10, 1.12 [a]]	.16 [-.44, .76 [a]]	.24 [-.36, .84 [a]]
	Active	-.25 [-.85, .35 [a]]	.00 [-.60, .60 [a]]	.11 [-.49, .71 [a]]	.26 [-.34, .86 [a]]	-.36 [-.96, .24 [a]]	-.18 [-.78, .42 [a]]
Cue vs. test-only	Passive	.00 [-.60, .60 [a]]	-.24 [-.84, .36 [a]]	-.13 [-.73, .47 [a]]	.27 [-.33, .87 [a]]	.30 [-.30, .90 [a]]	.58 [-.03, 1.19 [a]]
	Active	-.06 [-.66, .53 [a]]	.00 [-.60, .60 [a]]	.13 [-.47, .72 [a]]	.43 [-.18, 1.03 [a]]	.13 [-.47, .73 [a]]	-.10 [-.70, .50 [a]]
Noun vs. test-only	Passive	.30 [-.30, .89 [a]]	-.35 [-.95, .25 [a]]	-.04 [-.63, .55 [a]]	-.17 [-.76. .42 [a]]	.08 [-.52, .67] [a]	.31 [-.28, .91 [a]]
	Active	.21 [-.38, .80 [a]]	.00 [-.60, .60 [a]]	.00 [-.59, .59 [a]]	.13 [-.46, .72 [a]]	.53 [-.07, 1.14 [a]]	.12 [-.47, .71 [a]]

[a] 95% confidence intervals pass through zero

Table 12. JT between-group effect sizes corrected for baseline differences: Animate first nouns and voice by group and across test time (Cohen's d).

Group (*n*)		Immediate post *d*(CIs)		Delayed post *d*(CIs)	
		A-A	A-I	A-A	A-I
Cue vs. noun	Passive	.23	.42	.46	.15
	Active	.36	.26	-.11	-.18
Cue vs. test-only	Passive	-.13	.51	.30	.82
	Active	.19	.43	.19	-.10
Noun vs. test-only	Passive	-.34	.18	-.22	.66
	Active	-.21	.13	.32	.12

Acceptability judgements - trained / untrained passives and actives

Untrained and trained passives sentences

Table 13. JT mean accuracy % score for untrained and trained passive items across time (SDs)

Groups (*n*)	Pre		Immediate post		Delayed post	
	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)
Cue (21)	81 (15)	79 (15)	87 (11)	87 (13)	90 (8)	89 (9)
Noun (22)	80 (17)	80 (16)	80 (12)	82 (13)	86 (14)	87 (11)
Test-only (22)	76 (15)	80 (13)	80 (14)	84 (12)	85 (12)	85 (11)

Table 14. JT within-subject effect size (Cohen's d) between test times for untrained and trained passive items

Groups (*n*)	Pre vs immediate post *d* [CIs]		Immediate vs delayed post *d* [CIs]		Pre vs delayed post *d* [CIs]	
	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)
Cue (21)	.46 [-.16, 1.07 [a]]	.57 [-.05, 1.19 [a]]	.28 [-.33, .89 [a]]	.18 [-.43, .79 [a]]	.75 [.12, 1.37]	.81 [.18, 1.44]
Noun (22)	.00 [-.59, .59]	.14 [-.45 .73 [a]]	.46 [-.14, 1.06 [a]]	.42 [-.18, 1.01 [a]]	.39 [-.21, .98 [a]]	.51 [-.09, 1.11[a]]
Test-only (22)	.28 [-.32, .87 [a]]	.32 [-.28, .91 [a]]	.38 [-.21, .98 [a]]	.09 [-.50, .68 [a]]	.66 [.06, 1.27]	.42 [-.18, 1.01 [a]]

Table 15. JT between-subjects effect sizes (Cohen's d) for untrained and trained passive items

Groups (*n*)	Pre *d* [CIs]		Immediate *d* [CIs]		Delayed *d* [CIs]	
	Perfect (*k*=12)	Simple (*k*=12)	Perfect (*k*=12)	Simple (*k*=12)	Perfect (*k*=12)	Simple (*k*=12)
Cue vs. noun	.06 [-.54, .66 [a]]	-.06 [-.66, .53 [a]]	.61 [-.00, 1.22 [a]]	.38 [-.22, .99 [a]]	.35 [-.25, .95 [a]]	.20 [-.40, .80 [a]]
Cue vs. test-only	.34 [-.26, .95 [a]]	-.07 [-.67, .53 [a]]	.55 [-.05, 1.16 [a]]	.24 [-.36, .84 [a]]	.49 [-.12, 1.09 [a]]	.40 [-.21, 1.00 [a]]
Noun vs. test-only	.26 [-.34, .85 [a]]	.00 [-.59, .59 [a]]	.00 [-.59, .59 [a]]	-.16 [-.75, .43 [a]]	.08 [-.51, .67 [a]]	.18 [-.41, .77 [a]]

[a] 95% confidence intervals pass through zero

Table 16. JT between subjects effect sizes corrected for baseline differences for passive items (Cohen's d)

Groups (*n*)	Immediate		Delayed	
	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)
Cue vs. noun	.55	.44	.41	.26
Cue vs. test-only	.21	.31	.15	.16
Noun vs. test-only	-.26	-.16	-.18	.18

Untrained and trained active sentences

Table 17. JT mean accuracy % score for untrained and trained active items across time (SDs)

Groups (*n*)	Pre		Immediate		Delayed	
	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)
Cue (21)	71 (21)	76 (21)	78 (16)	88 (18)	76 (18)	91 (17)
Noun (22)	70 (19)	85 (19)	72 (19)	79 (18)	83 (13)	89 (17)
Test-only (22)	68 (22)	84 (19)	76 (17)	86 (20)	77 (17)	88 (17)

Table 18. JT within-subject effect size (Cohen's d) between test times for untrained and trained active items

Group (*n*)	Pre vs. immediate *d* [CIs]		Immediate vs. delayed *d* [CIs]		Pre vs. delayed *d* [CIs]	
	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)
Cue (21)	.31 [-.29, .93 [a]]	.61 [-.01, 1.23 [a]]	-.12 [-.72, .49 [a]]	.17 [-.42, .76 [a]]	.26 [-.35, .86 [a]]	.79 [.16, 1.41]
Noun (22)	.11 [-.48, .70 [a]]	-.32 [-.92 .27 [a]]	.68 [.07, 1.28]	.57 [-.03, 1.17 a]	.80 [.18, 1.41]	.22 [-.37, .81 [a]]
Test-only (22)	.41 [-.19, 1.00 [a]]	.10 [-.49, .69 [a]]	.05 [-.54, .64 [a]]	.11 [-.48, .70 [a]]	.41 [-.19, 1.01 [a]]	.22 [-.37, .81 [a]]

[a] 95% confidence intervals pass through zero

Table 19. JT between-groups effect sizes (Cohen's d) for untrained and trained active items

Groups (*n*)	Pre *d* [CIs]		Immediate post *d* [CIs]		Delayed post *d* [CIs]	
	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)
Cue vs. noun	.05	-.45	.34	.50	-.45	.12
	[-.55, .65 [a]]	[-1.06, .16 [a]]	[-.26, .94 [a]]	[-.11, 1.11 [a]]	[-1.05, .16 [a]]	[-.48, .72 [a]]
Cue vs. test-only	.14	-.4	.12	.11	-.06	.18
	[-.46, .74 [a]]	[-1.00, .20 [a]]	[-.48, .72 [a]]	[-.49, .70 [a]]	[-.66, .54 [a]]	[-.42, .78 [a]]
Noun vs. test-only	.10	.05	-.22	-.37	.40	.06
	[-.49, .69 [a]]	[-.54, .64 [a]]	[-.81, .37 [a]]	[-.96, .23 [a]]	[-.20, .99 [a]]	[-.53, .65 [a]]

[a] 95% confidence intervals pass through zero

Table 20. JT between subjects effect sizes corrected for baseline differences for active items (Cohen's d)

Groups (*n*)	Immediate post		Delayed post	
	Untrained (*k*=12)	Trained (*k*=12)	Untrained (*k*=12)	Trained (*k*=12)
Cue vs. noun	.29	.05	-.50	.57
Cue vs. test-only	-.02	.51	-.02	.58
Noun vs. test-only	-.32	-.42	.30	.01

Error identification

Error identification for passives

Table 21. JT error identification - mean % score for passive items across time (SDs)

Group (*n*)	Pre	Immediate post	Delayed post
Cue (23)	92 (11)	93 (15)	99 (3)
Noun (23)	94 (7)	95 (9)	95 (8)
Test-only (24)	89 (12)	96 (9)	100 (2)

Table 22. JT error identification - within-group effect size (Cohen's d) between test times for passive items

Group (*n*)	Pre vs. Immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (23)	.08 [-.50, .65 [a]]	.56 [-.03, 1.14 [a]]	.87 [.26, 1.47]
Noun (23)	.12 [-.46, .70 [a]]	.00 [-.58, .58 [a]]	.13 [-.45, .71 [a]]
Test-only (24)	.66 [.08, 1.24]	.61 [.04, 1.19]	1.28 [.66, 1.90]

[a] 95% confidence intervals pass through zero

Table 23. JT error identification – between-groups effect sizes (Cohen's d) for passive items.

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	-.22 [-.80, .36 [a]]	-.16 [-.74, .42 [a]]	.66 [.07, 1.26]
Cue vs. test-only	.26 [-.31, .83 [a]]	-.24 [.82, .33 [a]]	-.39 [-.97, .18 [a]]
Noun vs. test-only	.51 [-.07, 1.09 [a]]	-.11 [-.68, .46 [a]]	-.87 [-1.46, -.27]

[a] 95% confidence intervals pass through zero

Table 24. JT between subjects effect sizes corrected for baseline differences for error identification for passive items (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.06	.88
Cue vs. control	-.50	-.65
Noun vs. control	-.62	-1.38

Error identification for actives

Table 25. JT error identification - mean % score for active items across time (SDs)

Group (*n*)	Pre	Immediate post	Delayed post
Cue (23)	88 (16)	98 (8)	97 (7)
Noun (23)	91 (17)	95 (12)	94 (13)
Test-only (24)	90 (15)	94 (14)	97 (8)

Table 26. JT error identification - within-group effect size (Cohen's d) between test times for active items

Group (*n*)	Pre vs. Immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (23)	.79 [.19, 1.39]	-.13 [-.71, .45 [a]]	.73 [.13, 1.33]
Noun (23)	.27 [-.31, .85 [a]]	-.08 [-.66, .50 [a]]	.20 [-.38, .78 [a]]
Test-only (24)	.28 [-.29, .84[a]]	.26 [-.31, .83[a]]	.58 [.01, 1.16]

[a] 95% confidence intervals pass through zero

Table 27. JT error identification – between-groups effect sizes (Cohen's d) for active items.

Group (*n*)	Pre *d* [CIs]	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	-.18 [-.76, .40 [a]]	.29 [-.29, .88 [a]]	.29 [-.29, .86 [a]]
Cue vs. test-only	-.13 [-.70, .44 [a]]	.35 [-.23, .93 [a]]	.00 [-.57, .57 [a]]
Noun vs. test-only	.06 [-.51, .63 [a]]	.08 [-.50, .65 [a]]	-.28 [-.85, .29 [a]]

[a] 95% confidence intervals pass through zero

Table 28. JT between subjects effect sizes corrected for baseline differences for error identification for active items (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.47	.47
Cue vs. control	.48	.13
Noun vs. control	.02	-.34

Correct corrections of ungrammatical items

Correct corrections for passive ungrammatical sentences

Table 29. Correct corrections - mean % score for passive items across time (SDs) (k =12)

Group (*n*)	Pre-test	Post-test	Delayed Post test
Cue focused (24)	86 (18)	95 (6)	95 (6)
Noun focused (23)	83 (16)	91 (10)	92 (10)
Test only (23)	80 (13)	93 (6)	96 (7)

Table 30. Correct corrections - within-subject effect size (Cohen's d) between test times for passive items (k=12)

Group (*n*)	Pre vs. Immediate post *d* [CIs]	Immediate vs. delayed post *d* [CIs]	Pre vs. delayed post *d* [CIs]
Cue focused (24)	.67 [.09, 1.25]	.00 [-.57, .57 [a]]	.67 [.09, 1.25]
Noun focused (23)	.60 [.01, 1.19]	.10 [-.48, .68 [a]]	.67 [.08, 1.27]
Test-only (23)	1.28 [.65, 1.92]	.46 [-.13, 1.05 [a]]	1.53 [.88, 2.19]

[a] 95% confidence intervals pass through zero

Table 31. Correct corrections – between-groups effect sizes (Cohen's d) for passives

Group (*n*)	Pre *d* [CIs]	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	.18 [-.40, .75 [a]]	.49 [-.09, 1.07 [a]]	.37 [-.21, .94 [a]]
Cue vs. test-only	.38 [-.20, .96 [a]]	.33 [-.24, .91 [a]]	-.15 [-.73, .41 [a]]
Noun vs. test-only	.21 [-.38, .79 [a]]	-.24 [-.82, .34 [a]]	-.46 [-1.05. .12 [a]]

[a] 95% confidence intervals pass through zero

Table 32. JT between subjects effect sizes corrected for baseline differences for error identification for passive items (Cohen's d).

Group (*n*)	Immediate post *d* [CIs]	Delayed post *d* [CIs]
Cue vs. noun	.31	.19
Cue vs. test-only	-.05	-.53
Noun vs. test-only	-.45	-.67

Correct corrections for active ungrammatical sentences

Table 33. JT correct corrections - mean % score for active items across time (SDs) (k =12)

	Pre	Immediate	Delayed
Cue (24)	79 (18)	92 (6)	88 (14)
Noun (23)	89 (19)	89 (16)	89 (16)
Test-only (23)	88 (20)	88 (14)	94 (7)

Table 34. JT correct corrections - within-group effect size (Cohen's d) between test times for actives

Group (*n*)	Pre vs. Immediate *d* [CIs]	Immediate vs. delayed *d* [CIs]	Pre vs. delayed *d* [CIs]
Cue (24)	1.12 [.51, 1.73]	-.56 [-1.13, .02 [a]]	.56 [-.02, 1.13 [a]]
Noun (23)	.00 [-.58, .58 [a]]	.00 [-.58, .58 [a]]	.00 [-.58, .58 [a]]
Test-only (23)	.40 [-.18, .98 [a]]	.54 [-.05, 1.13 [a]]	.40 [-.18, .98 [a]]

[a] 95% confidence intervals pass through zero

Table 35. JT correct corrections – between-groups effect sizes (Cohen's d) for actives

Group (*n*)	Pre *d* [CIs]	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	-.54 [-1.12, .04 [a]]	.42 [-.16, 1.00 [a]]	-.07 [-.64, .51 [a]]
Cue vs. test-only	-.47 [-1.05, .11 [a]]	.56 [-.02, 1.14 [a]]	-.54 [-1.12, .04 [a]]
Noun vs. test-only	.05 [-.53, .63 [a]]	.07 [-.51, .64 [a]]	-.40 [-.99, .18 [a]]

[a] 95% confidence intervals pass through zero

Table 36. JT between subjects effect sizes corrected for baseline differences for error identification for active items (Cohen's d).

Group (*n*)	Immediate *d* [CIs]	Delayed *d* [CIs]
Cue vs. noun	.96	.47
Cue vs. test-only	1.03	-.07
Noun vs. test-only	.02	-.45

Judgement test incorrect corrections

Table 37. JT passive sentences all learners at pre-test (mistake count)

Mistake	base form	-ing	missing be	Total
addition of be	1	5	2	8
being not been			1	1
change tense	5	4	5	13
changed noun	2	1		3
infinitive			1	1
ing not past	23			23
past form not ing		1		1
removed been	2			2
removed by		2		2
removed has			1	1
spelling		3		3
tense change	1			1
verb change			12	12
verb form incorrect irregular	1			1
verb form incorrect regular	18	8		26

verb form past simple	2	8		10
Grand Total	45	41	22	108

Table 38. JT active sentences all learners at pre-test (mistake count)

Mistake	base form	-ing	missing be	Total
being not been			1	1
change tense	2	1	2	5
changed noun	1	1		2
incorrect addition of by	6	6		12
past form not ing	17	5		22
removed been	3	1		4
spelling	1	2		3
verb change			11	11
verb form incorrect regular		1		1
verb form past simple	4			4
Total	34	17	14	65

Table 39. JT immediate post-test passive voice correction errors: Cue focused group (mistake count)

Mistake	base form	-ing	missing be	Total
added been	1			1
change noun		1		1
change tense		1		1
change verb			1	1
ing not past	1	3		4
removed by		1		1
verb form incorrect regular	3		1	4
Total	5	6	2	13

Table 4. JT immediate post-test passive voice correction errors: noun focused group (mistake count)

Mistake	base form	-ing	missing be	Total
change noun		1		1
change verb			4	4

infinitive verb		1		1
ing not past	2	1	1	4
no been			1	1
removed by		2		2
verb form incorrect irregular	2	1		3
verb form incorrect regular	2	1		3
verb form past simple		1		1
Total	6	8	6	20

Table 41. JT immediate post-test passive voice correction errors: Test-only group (mistake count)

Mistake	base form	-ing	missing be	Total
added being		1		1
change noun		1		1
change verb			2	2
ing not past	1	2		3
no been			1	1

removed been			1	1
spelling		1		1
verb form incorrect regular	3	1		4
verb form past simple		1		1
(blank)				
Total	4	7	4	15

Table 42. JT immediate post-test active voice correction errors: Cue focused group (mistake count)

Mistake	base form	-ing	missing be	Total
change tense	1			1
past form not ing	3	1		4
removed been	2	1		3
(blank)				
Total	6	2		8

Table 43. JT immediate post-test active voice correction errors: Test-only group (mistake count)

Mistake	base form	-ing	missing be	Total
change verb			3	3

infinitive verb		1		1
past form not ing	5	2		7
removed been	1			1
(blank)				
Total	6	3	3	12

Table 44. JT immediate post-test active voice correction errors: Test-only group (mistake count)

Mistake	base form	-ing	missing be	Total
added by	1			1
change tense		1		1
change verb			3	3
past not ing	3	1		4
removed been	3			3
verb form incorrect regular		1		1
Total	7	3	3	13

Table 45. JT delayed post-test passive voice correction errors: Cue focused group (mistake count)

Mistakes	base form	-ing	missing be	Total
change tense			1	1
change verb			2	2
infinitive		1		1
removed by		1		1
verb form incorrect irregular	1			1
verb form incorrect regular	3	2	1	6
verb form past simple		1		1
Total	4	5	4	13

Table 46. JT delayed post-test passive voice correction errors: noun focused group (mistake count)

Mistakes	base form	-ing	missing be	Total
change noun		3		3
changed tense		1		1
ing not past	6	1		7

removed by		3	3
verb form incorrect regular	1	2	3
verb ing incorrect regular	1		1
Total	8	10	18

Table 47. JT delayed post-test passive voice correction errors: Test-only group (mistake count)

Mistake	base form	-ing	missing be	Total
addition of been		1		1
being not been			2	2
change verb			1	1
spelling		2		2
verb form incorrect regular	1	2		3
verb form past simple	1	2		3
Total	2	7	3	12

Table 48. JT delayed post-test active voice correction errors: Cue focused group (mistake count)

Mistake	base form	-ing	missing be	Total
change tense	1			1

change verb			2	2
incorrect addition of by		1		1
past not ing	2	3	1	6
removed auxiliary			1	1
removed been	1	1		2
Total	4	5	4	13

Table 49. JT delayed post-test active voice correction errors: noun focused group (mistake count)

Mistake	base form	-ing	missing be	Total
change noun		1		1
change verb		1	3	4
incorrect addition of by		4		4
past not ing	2	1		3
removed been	1	1		2
verb form past simple	1			1
Total	4	8	3	15

Table 50. JT delayed post-test active voice correction errors: Test-only group (mistake count)

Mistake	base form	-ing	missing be	Total
change verb			4	4
changed verb		2	1	3
incorrect addition of by	1			1
past not ing	4	3	2	9
verb form past simple	1			1
Total	6	5	7	18

Glossary

A-A	Animate first noun - animate second noun
A-I	Animate first noun - inanimate second noun
AOI	Area of interest (in eye-tracking trial)
CEFR	Common European Framework of Reference
CET	College English Test
CI	95% confidence interval
EEG	Electroencephalography
EFL	English as a foreign language
EI	Explicit information
ERP	Event related potential
ESL	English as a second language
f	Feminine
GJT	Grammaticality judgement test
HSK	International standardized test of Chinese language proficiency
I-A	Inanimate first noun - animate second noun
IELTs	International English Language Test
I-I	Inanimate first noun - inanimate second noun
JT	Judgement test
k	Number of test items
L1	First language
L2	Second language

M	Mean
m	Masculine
ms	Milliseconds
n	Number of participants
n	Neuter
N	Noun
N400	EEG signal indicating sensitivity to a lexical-semantic anomaly
NS	Native speaker
O	Object
P600	EEG signal indicating sensitivity to syntactic violations
PI	Processing instruction
RAGE	Reduced ability to generate expectations
RT	Reaction time
S	Subject
SD	Standard deviation
SPR	Self-paced reading
TW	Time window
V	Verb

References

Adams, R. (2020). Third of non-EU university students in UK come from China. *The Guardian*. Retrieved from https://www.theguardian.com/education/2020/jan/16/third-of-non-eu-university-students-in-uk-come-from-china

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research*, *16*(1), 9–37. https://doi.org/10.1177/1362168811423339

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. W. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). Second Language Teaching and Curriculum Center.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Altmann, L. J. P & Kemper, S. (2006) Effects of age, animacy and activation order on sentence production, *Language and Cognitive Processes, 21*, 1-3, 322-354, https://doi.org/10.1080/0169096054400006

Alzuhairy, A. (2012). *The frequency of the twelve verb tenses in academic papers written by native speakers*. [Unpublished master's thesis]. University of Central Florida, Orlando, Florida.

Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing. *Studies in Second Language Acquisition*, *37*(2), 237–268. https://doi.org/10.1017/S0272263115000017

Andringa, S., de Glopper, K., & Hacquebord, H. (2011). Effect of explicit and implicit instruction on free written response task performance. *Language Learning*, *61*(3), 868–903. https://doi.org/10.1111/j.1467-9922.2010.00623.x

Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning. *Studies in Second Language Acquisition*, *37*(2), 185–196. https://doi.org/10.1017/S027226311500008X

Armon-Lotem, S., Haman, E., Jensen de López, K., Smoczynska, M., Yatsushiro, K., Szczerbinski, M., … van der Lely, H. (2016). A large-scale cross-linguistic investigation of the acquisition of passive. *Language Acquisition*, *23*(1), 27–56. https://doi.org/10.1080/10489223.2015.1047095

Avery, N., & Marsden, E. (2019). A meta-analysis of sensitivity to grammatical information during self-paced reading: Towards a framework of reference for reading time effects sizes. *Studies in Second Language Acquisition*, *41*(5), 1055–1087. https://doi.org/10.1017/S0272263119000196

Bailey, S. (2015). *Academic writing : a handbook for international students* (Fourth ed.). Routledge.

Balcom, P., & Bouffard, P. (2015). The Effect of Input Flooding and Explicit Instruction on Learning Adverb Placement in L3 French. *Canadian Journal of Applied Linguistics*, *18*(2), 1–27. Retrieved from https://journals.lib.unb.ca/index.php/CJAL/article/view/20186

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Bar, M. (2011). *Predictions in the Brain: Using Our Past to Generate a Future*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195395518.001.0001

Bardovi-Harlig, K. (1997). Another piece of the puzzle: The emergence of the present perfect. *Language Learning ,47*(3). 375–422.

Barker, C., & Dowty, D. (1993). Non-verbal thematic proto-roles. *Proceedings of NELS*.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. *The crosslinguistic study of sentence processing*, *3*, 73-112.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bates, E., Chen, S., Tzeng, O., Li, P., & Opie, M. (1991). The noun-verb problem in Chinese aphasia. *Brain and Language*, *41*(2), 203–233. https://doi.org/10.1016/0093-934X(91)90153-R

Bartoń, K. (2018). MuMIn: Multi-model inference. R package version 1.42.1. https://cran.r-project.org/web/packages/MuMIn/index.html

Benati, A. (2001). A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research*, *5*(2), 95–127. https://doi.org/10.1177/136216880100500202

Benati, A. (2004). The effects of processing instruction and its components on the acquisition of gender agreement in Italian. *Language Awareness*, *13*(2), 67–80. https://doi.org/10.1080/09658410408667087

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning—output instruction on the acquisition of the English past simple tense. *Language Teaching Research*, *9*(1), 67–93. https://doi.org/10.1191/1362168805lr154oa

Benati, A., & Lee, J. F. (2012). *Processing Instruction and Discourse.* (1st editio). New York: Bloomsbury.

Bever, T. G. (2013). The cognitive basis for linguistic structures. In *Language Down the Garden Path* (pp. 1–80). https://doi.org/10.1093/acprof:oso/9780199677139.003.0001

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning*, *29*(1), 81–103. https://doi.org/10.1111/j.1467-1770.1979.tb01053.x

Bialystok, E. (1982). On the relationship between knowing and using linguistic forms. *Applied Linguistics*, *III*(3), 181–206. https://doi.org/10.1093/applin/III.3.181

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*(2), 157–168. https://doi.org/10.1017/S0272263100012857

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Bock, J. K. (1989). Closed class immanence in sentence production. *Cognition*, *31*, 163–186.

Bock, J. K. & Loebell, H. (1990). Framing sentences. *Cognition, 35,* 1–39.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language*, *34*(6), 774–806. https://doi.org/10.1006/jmla.1995.1034

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113,* 787–821.

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *33*(2), 247–271. https://doi.org/10.1017/S0272263110000756

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & Cognition*, *41*(2), 281–296. https://doi.org/10.3758/s13421-012-0262-9

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, *27*(2), 185–209. https://doi.org/10.1017/S0142716406060243

Brown, G. D. A. (1992). Cognitive psychology and second language processing: The role of short-term memory. In *Advances in Psychology* (pp. 105–121). https://doi.org/10.1016/S0166-4115(08)61490-9

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving Korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics*, *66*, 45–60. https://doi.org/10.1016/j.pragma.2014.02.011

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal*, *79*(2), 179-193. https://doi.org/10.1111/j.1540-4781.1995.tb05430.x

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Rand McNally.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback. *Studies in Second Language Acquisition*, *15*(3), 357–386.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Heinle & Heinle

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102.

Chang, X., & Wang, P. (2016). Influence of second language proficiency and syntactic structure similarities on the sensitivity and processing of English passive sentences in late Chinese-English bilinguists: An ERP study. *Journal of Psycholinguistic Research*, *45*(1), 85–101. https://doi.org/10.1007/s10936-014-9319-1

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research*, *10*(2), 157–187. https://doi.org/10.1177/026765839401000203

Cheng, A. C. (1995). *Grammar instruction and input processing: The acquisition of Spanish ser and estar*. (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Cheng, A. C. (2002). The effects of processing instruction on the acquisition of ser and estar. *Hispania. 85*(2). https://doi.org/10.2307/4141092.

Chomsky, N. (1965). Persistent topics in linguistic theory. *Diogenes*, *13*(51), 13–20. https://doi.org/10.1177/039219216501305102

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752

Chu, C. C. (1973). The passive construction: Chinese and English. *Journal of Chinese Linguistics*, *1*(3), 437–470. Retrieved from http://www.jstor.org/stable/23752862

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570. https://doi.org/10.1016/j.tics.2006.10.002

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, L., Manion, L., & Morrison, K. (2017). Research methods in education. In *Research Methods in Education*. https://doi.org/10.4324/9781315456539

Collins, P. C., & Hollo, C. (2017). *English grammar: An introduction* (3rd Ed.).

Comer, W. J., & DeBenedette, L. (2011). Processing instruction and Russian: Further evidence is IN. *Foreign Language Annals*, *44*(4), 646–673.

https://doi.org/10.1111/j.1944-9720.2011.01155.x

Cook, V., Iarossi, E., Stellakis, N., & Tokumaru, Y. (2003). Effects of the L2 on the syntactic processing of the L1. In *Effects of the Second Language on the First*.

Cornillie, F., Noortgate, W., Branden, K.V., & Desmet, P. (2017). Examining focused L2 practice: from in vitro to in vivo. *Language Learning & Technology, 21*(1), 121-145.

Council of Europe. (2001). The common European framework of reference for languages: Learning, teaching, assessment. In *Council of Europe*. https://doi.org/10.1017/S0267190514000221

Cowen, J. R., & Reed, B. (1988). Applying an experimental approach to the investigation of syntactic change. In T. Walsh (Ed.), *Synchronic and Diachronic Approaches to Linguistic Variation and change.* (pp. 261–267). Georgetown University Press.

Cowles, H.W., & Wijnen, F. (2015). Object referent prediction in native and highly proficient L2 speakers. *CUNY Conference on Human Sentence Processing*. Los Angeles, CA.

Culman, H., Henry, N. and VanPatten, B. (2009). The role of explicit information in instructed SLA: An on-line study with processing instruction and German accusative case inflections. *Die Unterrichtspraxis / Teaching German, 42*, 19-31. doi:10.1111/j.1756-1221.2009.00032.x

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora resolution and reanalysis during l2 sentence processing: Evidence from the visual world paradigm. *Studies in Second Language Acquisition*, *39*(4), 621-652. https://doi.org/10.1017/S0272263116000292

Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, *69*(S1), 42–71. https://doi.org/10.1111/lang.12321

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*(4), 465–480. https://doi.org/10.1006/jmla.1999.2688

Day, E. M., & Shapson, S. M. (2001). Integrating formal and functional approaches to language teaching in French immersion: An experimental study. *Language Learning*, *51*(S1), 47–80. https://doi.org/10.1111/j.1467-1770.2001.tb00014.x

De Jong, N. (2005). Can second language grammar be learned through listening?: An experimental study. *Studies in Second Language Acquisition*, *27*(2), 205–234. https://doi.org/10.1017/S0272263105050114

DeKeyser, R. M. (1995). Learning second language grammar rules. *Studies in Second Language Acquisition*, *17*(3), 379–410. https://doi.org/10.1017/S027226310001425X

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*(2), 195-221. https://doi.org/10.1017/S0272263197002040

DeKeyser, R. M. (1998). Beyond focus on from: Cognitive perspectives on learning and practicing second language grammar. *Focus on Form in Classroom Second Language Acquisition*.

DeKeyser, R. (2008). Implicit and explicit learning. In *The Handbook of Second Language Acquisition* (pp. 312–348). https://doi.org/10.1002/9780470756492.ch11

DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15-32). Routledge.

Dekeyser, R., & Criado, R. (2012). Automatization, skill acquisition, and practice in second language acquisition. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0067

DeKeyser, R., Salaberry, R., Robinson, P. and Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's "Processing Instruction: An update". *Language Learning*, *52*(4), 805-823. https://doi.org/10.1111/1467-9922.00204

DeKeyser, R. M., & Sokalski, K. J. (2001). The differential role of comprehension and production practice. *Language Learning, 51*, 81-112.

Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-assessment of linguistic skills by bilingual hispanics. *Hispanic Journal of Behavioral Sciences*, *21*(1), 31–46. https://doi.org/10.1177/0739986399211003

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/n*n*504

den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00548

de Wilde, V., Brysbaert, M. and Eyckmans, J. (2020). Learning English through out-of-school exposure: How do word-related variables and proficiency influence receptive vocabulary Learning?. *Language Learning, 70*(2), 349-381. https://doi.org/10.1111/lang.12380

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, *20*(5), 917–930. https://doi.org/10.1017/S1366728916000547

Dörnyei, Z., & Katona, L. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, *9*(2), 187–206. https://doi.org/10.1177/026553229200900206

Doughty, C. J. (2008). Instructed SLA: Constraints, compensation, and enhancement. In *The handbook of second language acquisition* (Vol. 27, pp. 256–310). Blackwell Publishing.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547. https://doi.org/10.2307/415037

Dracos, M. J. (2012). *The effects of form-focused training and working memory on the L2 processing and learning of morphological cues*. The Pennsylvania State University.

Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. Oxford University Press.

Dussias, P. E. (2004). Parsing a first language like a second: The erosion of L1 parsing strategies in Spanish-English Bilinguals. *International Journal of Bilingualism*, *8*(3), 355–371. https://doi.org/10.1177/13670069040080031001

Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand. *Studies in Second Language Acquisition*, *35*(2), 353–387. https://doi.org/10.1017/S0272263112000915

Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, *23*(3), 290–325. https://doi.org/10.1191/0265532206lt330oa

Ellis, N. C. (2016). Salience, cognition, language complexity, and complex adaptive systems. *Studies in Second Language Acquisition*, *38*(2), 341–351. https://doi.org/10.1017/S027226311600005X

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305-352. https://doi.org/10.1017/S027226310505014X

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164–194. https://doi.org/10.1093/applin/aml015

Ellis, N. C. (2007). Blocking and learned attention in language acquisition. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, *29*(29).

Ellis, N. C., & Sagarra, N. (2011). Learned attention in adult language acquisition: A

Replication and generalization study and meta-analysis. *Studies in Second Language Acquisition*, *33*(4), 589-624. https://doi.org/10.1017/S0272263111000325

Ellis, R. (1991). The interaction hypothesis: A critical evaluation. *Regional Language Centre Seminar*.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227–275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-26). Multilingual Matters.

Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition*, *29*(1). 119-126. https://doi.org/10.1017/S0272263107070052

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*(2), 339-368. https://doi.org/10.1017/S0272263106060141

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147–171. https://doi.org/10.1191/1362168805lr161oa

Ervin-Tripp, S. M. (1974). Is second language learning like the first. *TESOL Quarterly*, *8*(2), 111. https://doi.org/10.2307/3585535

Farley, A., & Aslan, E. (2011). The relative effects of processing instruction and meaning-based output instruction on L2 acquisition of the English subjunctive. *ELT Research Journal 1*(2), 120-141.

Feng, S., Legault, J., Yang, L., Zhu, J., Shao, K., & Yang, Y. (2015). Differences in grammatical processing strategies for active and passive sentences: An fMRI study. *Journal of Neurolinguistics*, *33*, 104–117. https://doi.org/10.1016/j.jneuroling.2014.09.002

Fernández, C. (2008). Reexamining the role of explicit information in processing instruction. *Studies in Second Language Acquisition*, *30*(3), 277–305. https://doi.org/10.1017/S0272263108080467

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33*(6), 715-736.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47*(2), 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69*(3), 165–182. https://doi.org/10.1016/j.jml.2013.06.001

Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 324–335.

https://doi.org/10.1037/0278-7393.22.2.324

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-Enough language processing. In *Psychology of Learning and Motivation - Advances in Research and Theory* (pp. 217–247). https://doi.org/10.1016/bs.plm.2016.04.002

Ferreira, F., & Stacey, J. (2000). The misinterpretation of passive sentences.

Ferreira, V. S. (1996). Is it better to give them than to denote? Syntactic flexibility in language production. *Journal of Memory and Language*, *35*, 724–755.

Fiedler, S., Schulte-Mecklenbeck, M., Renkewitz, F., & Orquin, J. L. (2019). Increasing reproducibility of eye-tracking studies. In M. Schulte-Mecklenbeck, A. Kuehberger, & J. G. Johnson (Eds.), *A Handbook of Process Tracing Methods*, *65*. Routledge.

Fine, A. B., & Florian Jaeger, T. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, *37*(3), 578–591. https://doi.org/10.1111/cogs.12022

Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. A. (2010). Is there syntactic adaptation in language comprehension? *Proceedings of ACL: Workshop on Cognitive Modeling and Computational Linguistics*. Uppsala, Sweden.

Flecken, M., Carroll, M., Weimar, K., & Von Stutterheim, C. (2015). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French-German L2 users. *Modern Language Journal*, *99*(S1), 100–122. https://doi.org/10.1111/j.1540-4781.2015.12181.x

Flege, J. E., Mackay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*(4), 567–598. https://doi.org/10.1017/S0142716402004046

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469. https://doi.org/10.1037/a0036756

Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say "book"? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience*, *30*(6), 768–780. https://doi.org/10.1080/23273798.2015.1016047

Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition*, *19*(1), 213–219. https://doi.org/10.1017/S1366728915000486

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research*, *21*(2), 175–198. https://doi.org/10.1191/0267658305sr257oa

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Fuchs, R., Göetz, S., Werner, V., Seoane, E., & Suárez-Gómez, C. (2016). The present perfect in learner Englishes: A corpus-based case study on L1 German intermediate and advanced speech and writing. In V. Werner, E. Seoane & C. Suárez-Gómez (Eds). *Re-assessing the Present Perfect* (pp. 297–338). De Gruyter Mouton. https://doi.org/10.1515/9783110443530-013

Garcia, P., & Gass, S. (2000). Input, interaction, and the second language learner. *TESOL Quarterly*, *34*(4), 789. https://doi.org/10.2307/3587793

Gass, S., & Mackey, A. (2012). *Data Elicitation for Second and Foreign Language Research*. Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203826102

Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning*, *55*(4), 575–611. https://doi.org/10.1111/j.0023-8333.2005.00318.x

Gathercole, V. C. (1986). The acquisition of the present perfect: Explaining differences in the speech of Scottish and American children. *Journal of Child Language*, *13*(3). 537–560.

Godfroid, A. (2019). *Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide (Second Language Acquisition Research Series)*. https://doi.org/10.4324/9781315775616

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., … Yoon, H. (2018). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

Godfroid, A., & Hui, B. (2020). Five common pitfalls in eye-tracking research. *Second Language Research*, *36*(3), 277–305. https://doi.org/10.1177/0267658320921218

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & R. Ellis (2015). Timed and untimed grammaticality judgment measure distinct types of knowledge. *Studies in Second Language Acquisition*, *37*(2), 269–297. https://doi.org/10.1017/S0272263114000850

Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 325–348). https://doi.org/10.1075/sibil.48.14god

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. *Phraseology: Theory, analysis and applications*, 145-160.

Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, *36*(4), 611–645. https://doi.org/10.1017/S0272263113000727

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on `alternations'. *International Journal of Corpus Linguistics*, *9*(1), 97–129. https://doi.org/10.1075/ijcl.9.1.06gri

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: the role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, *35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*(3), 423–449. https://doi.org/10.1017/S0272263113000041

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited. *Studies in Second Language Acquisition*, *32*(3), 465–491. https://doi.org/10.1017/S0272263110000045

Han, Y., & R. Ellis (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*(1), 1–23. https://doi.org/10.1177/136216889800200102

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study With German L2 learners of Dutch. *Language Learning*, *59*(1), 73–112. https://doi.org/10.1111/j.1467-9922.2009.00501.x

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 512–521. https://doi.org/10.1016/S0022-5371(74)80003-4

Hawkins, R., & Liszka, S. (2003). Locating the Source of Defective Past Tense Marking in Advanced L2 English Speakers. In R. Van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The lexicon-syntax interface in second language acquisition* (pp. 21–44). John Benjamins.

Hedgcock, J. (1993). Well-formed vs. ill-formed strings in L2 metalingual tasks: specifying features of grammaticality judgements. *Second Language Research*, *9*(1), 1–21. https://doi.org/10.1177/026765839300900101

Henderson, J. M. (2017) Gaze control as prediction. *Trends in Cognitive Sciences 21*(1), 15–23. https://doi.org/10.1016/j.tics.2016.11.003

Henderson, J.M., & Ferreira, F (2004) Scene perception for psycholinguists. In J. M., Henderson, & F., Ferreira (Eds.) *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). Psychology Press.

Henry, N. (2015). *Morphosyntactic processing, cue interaction, and the effects of instruction: An investigation of Processing Instruction and the acquisition of case markings in L2 German.* [Unpublished doctoral thesis]. Pennsylvania State University.

Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA. *Studies in Second Language Acquisition*, *31*(4), 559. https://doi.org/10.1017/S0272263109990027

Hinkel, E. (2002). Why English passive is difficult to teach (and learn). In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching* (pp. 233–260). Mahway, NJ: Lawrence Erlbaum Associates.

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research, 8*(1), 5–29. https://doi.org/10.1191/1362168804lr132oa

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369–397. https://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. https://doi.org/https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching. 53*(3), 277–306. https://doi.org/10.1515/iral-2015-0014.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307. https://doi.org/10.1177/0267658315624960

Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199. https://doi.org/10.1017/S0272263116000437

Howell, D. (2002). *Statistical Methods For Psychology*. Wadsworth Cengage Learning. https://doi.org/10.2307/2348956.

Hsu, H.-C. (2017). The effect of task planning on L2 performance and L2 development in text-based synchronous computer-mediated communication. *Applied Linguistics*, *38*(3), 359–385. https://doi.org/10.1093/applin/amv032

Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, *69*(4), 589–606. https://doi.org/10.1016/j.jml.2013.08.002

Hulstijn, J. H. (1997). Second language acquisition research in the laboratory. *Studies in Second Language Acquisition*, *19*(2), 131–143. https://doi.org/10.1017/S0272263197002015

Hulstijn, J. H. (2005). Theoretical and empirical issyes in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*(2). https://doi.org/10.1017/S0272263105050084

Hulstijn, J. H., & de Graff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge?: A research proposal. *AILA Review*.

Indrarathne, B., & Kormos, J. (2017). Attentional processing of input in explicit and implicit conditions. *Studies in Second Language Acquisition*, *39*(3), 401–430. https://doi.org/10.1017/S027226311600019X

Indrarathne, B., Ratajczak, M., & Kormos, J. (2018). Modelling changes in the cognitive processing of grammar in implicit and explicit learning conditions: Insights from an eye-tracking study. *Language Learning*, *68*(3), 669–708. https://doi.org/10.1111/lang.12290

Isabelli, C. A. (2008). First noun principle or L1 transfer principle in SLA? *Hispania*, *91*(2), 465. https://doi.org/10.2307/20063732

Ishihara, K., Hiser, E., & Okada, T. (2003). Modifying c-test for practical purposes. *Doshisha Studies in Language and Culture*, *5*(4), 539–568.

Issa, B. I., & Morgan-Short, K. (2019). Effects of external and internal attentional manipulations on second language grammar development. *Studies in Second Language Acquisition*, *41*(2), 389–417. https://doi.org/10.1017/S027226311800013X

Issa, B., Morgan-Short, K., Villegas, B., & Raney, G. (2015). An eye-tracking study on the role of attention and its relationship with motivation. *EUROSLA Yearbook*, *15*, 114–142. https://doi.org/10.1075/eurosla.15.05iss

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Ito, K., & Wong, W. (2019). Processing instruction and the effects of input modality and voice familiarity on the acquisition of the French causitive construction. *Studies in Second Language Acquisition*, *41*(2), 443–468. https://doi.org/10.1017/S0272263118000281

Izumi, S., & Lakshmanan, U. (1998). Learnability, negative evidence and the L2 acquisition of the English passive. *Second Language Research*, *14*(1), 62–101. https://doi.org/10.1191/026765898675700455

Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, *58*(4), 875–909. https://doi.org/10.1111/j.1467-9922.2008.00481.x

Jackson, C., & Roberts, L. (2010). Animacy affects the processing of subject–object ambiguities in the second language: Evidence from self-paced reading with German second language learners of Dutch. *Applied Psycholinguistics*, *31*(4), 671–691. https://doi.org/10.1017/S0142716410000196

Jackson, C., & van Hell, J. G. (2011). The effects of L2 proficiency level on the processing of wh-questions among Dutch second language speakers of English. *IRAL - International Review of Applied Linguistics in Language Teaching*, *49*(3). https://doi.org/10.1515/iral.2011.012

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83. https://doi.org/10.1016/j.cognition.2012.10.013

Jafarpur, A. (1995). Is C-testing superior to cloze? *Language Testing*, *12*(2), 194–216. https://doi.org/10.1177/026553229501200204

Jia, G., Aaronson, D., & Wu, Y. (2002). Long-term language attainment of bilingual immigrants: Predictive variables and language group differences. *Applied Psycholinguistics*, *23*(4), 599–621. https://doi.org/10.1017/S0142716402004058

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, *42*(2), 217–248. https://doi.org/10.1111/j.1467-1770.1992.tb00708.x

Juffs, A. (1998). Some effects of first language argument structure and morphosyntax on second language sentence processing. *Second Language Research*, *14*(4), 406–424. https://doi.org/10.1191/026765898668800317

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition*, *17*(4), 483–516. https://doi.org/10.1017/S027226310001442X

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language Learning*, *46*(2), 283–323. https://doi.org/10.1111/j.1467-1770.1996.tb01237.x

Kaan, E., Ballantyne, J. C., & Wijnen, F. (2015). Effects of reading speed on second-language sentence processing. *Applied Psycholinguistics*, *36*(4), 799–830. https://doi.org/10.1017/S0142716413000519

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003b). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003a). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37–55. https://doi.org/10.1023/A:1021933015362

Kasprowicz, R., & Marsden, E. (2018). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics*, *39*(6), 886–911. https://doi.org/10.1093/applin/amw051

Kasprowicz, R., Marsden, E., & Sephton, N. (2019). Investigating distribution of practice effects for the learning of foreign language verb morphology in the young learner classroom. *The Modern Language Journal*, *103*(3), 580–606. https://doi.org/10.1111/modl.12586

Keating, G. D., & Farley, A. P. (2008). Processing instruction, meaning-based output instruction, and meaning-based drills: impacts on classroom L2 acquisition of Spanish object pronouns. *Hispania*, *91*(3), 639-650.

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*(1), 1–32. https://doi.org/10.1017/S0272263114000187

Keenan, E. L., & Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, *8*(1), 63–69.

Kellerman, E. (1985). If at first you do succeed... In S. M. Gass & C. Madden (Eds.), *Input and second language acquisition* (pp. 345-353). Newbury House.

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning*, *58*(4), 703–746. https://doi.org/10.1111/j.1467-9922.2008.00477.x

Kempe, V., Brooks, P. J., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, *60*(1), 127–153. https://doi.org/10.1111/j.1467-9922.2009.00553.x

Kim, H., & Lee, W. (2019). Validation of the grammaticality judgment test of Korean passive expressions through exploratory factor analysis. *The Korean Language in America*, *22*(2), 122–155. https://doi.org/10.5325/korelangamer.22.2.0122

Kim, E., Montrul, S., & Yoon, J. (2015). The on-line processing of binding principles in second language acquisition: Evidence from eye tracking. *Applied Psycholinguistics*, *36*(6), 1317-1374. https://doi.org/10.1017/S0142716414000307

Kohlstedt, T., & Mani, N. (2018). The influence of increasing discourse context on L1 and L2 spoken language processing. Bilingualism: *Language and Cognition*, *21*(1), 121-136. https://doi.org/10.1017/S1366728916001139

Krashen, S. D. (1982). Acquiring a second language. *World Englishes*, *1*(3), 97–101. https://doi.org/10.1111/j.1467-971X.1982.tb00476.x

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Larsen-Freeman, D. (1997). Grammar and Its Teaching: Challenging the Myths. *ERIC Clearinghouse on Languages and Linguistics*.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. Oxford University.

Larson-Hall, J. (2015). A guide to doing statistics in second language research using SPSS and R. In *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. https://doi.org/10.4324/9781315775661

Lee, J. F., & Malovrh, P. A. (2009). Linguistic and non-linguistic factors affecting OVS processing ofaccusative and dative case pronounsby advanced L2 learners of Spanish. *11th Hispanic Linguistics Symposium*, 105–116. Cascadilla Proceedings Project.

Lee, J. F. (2014). The relationship between learning rate and learning outcome for processing instruction on the Spanish passive voice. In A. G. Benati, C. Laval, & M. Arche (Eds.), *The grammar dimension in instructed second language learning : Advances in instructed second language acquisition research* (pp. 147–163). https://doi.org/10.5040/9781472542113.ch-007

Lee, J. F. (2015). Processing instruction on the Spanish passive with transfer-of-training effects to anaphoric and cataphoric reference contexts. *International Review of Applied Linguistics in Language Teaching*, *53*(2). https://doi.org/10.1515/iral-2015-0010

Lee, J. F., & Doherty, S. (2019). Native and nonnative processing of active and passive sentences. *Studies in Second Language Acquisition*, *41*(4), 853–879. https://doi.org/10.1017/S027226311800027X

Lee, J. F., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345–366.

https://doi.org/10.1093/applin/amu040

Lee, J. F., & VanPatten, B. (2003). *Making communicative language teaching happen*. McGraw-Hill.

Lee, S.-K., & Huang, H. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, *30*(3), 307-331. https://doi.org/10.1017/S0272263108080479

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, *42*(3), 625-651. https://doi.org/10.1017/S0272263120000108

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, *47*(3), 467–505. https://doi.org/10.1111/0023-8333.00017

Leow, R. P. (2015). *Explicit Learning in the L2 Classroom*. Routledge. https://doi.org/10.4324/9781315887074

Leow, R. P., & Donatelli, L. (2017). The role of (un)awareness in SLA. *Language Teaching*, *50*(2), 189–211. https://doi.org/10.1017/S0261444817000039

Leow, R. P., & Martin, A. (2017). Enhancing the input to promote salience of L2: A critical overview. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in Second Language Acquisition*. Taylor & Francis. https://doi.org/10.4324/9781315399027

Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, *33*(1), 33–55. https://doi.org/10.1017/S0272263110000525

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Li, C. N., & Thompson, S. A. (1976). Development of the causative in Mandarin Chinese: Interaction of diachronic processes in syntax. In *The Grammar of Causative Constructions* (pp. 477–492). Brill. https://doi.org/10.1163/9789004368842_020

Li, P., Bates, E., Liu, H., & MacWhinney, B. (1992). Cues as functional constraints on sentence processing in Chinese. *Advances in Psychology, 90*, 207–234. https://doi.org/10.1016/S0166-4115(08)61893-2

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language, 32*(2), 169–192. https://doi.org/10.1006/jmla.1993.1010

Li, P., & MacWhinney, B. (2012). Competition model. In C. Chapelle (Ed.) *The Encyclopedia of Applied Linguistics*. John Wiley & Sons, Ltd.

Li, S., Ellis, R. & Zhu, Y. (2016b). Task-based versus task-supported language instruction: An experimental study. *Annual Review of Applied Linguistics, 36*, 205–229. https://doi.org/10.1017/S0267190515000069

Li, S., Zhu, Y. and Ellis, R. (2016a), The effects of the timing of corrective feedback on the acquisition of a new linguistic structure. *The Modern Language Journal, 100*(1), 276-295. https://doi.org/10.1111/modl.12315

Liao, M.-T., & Chen, C.-H. (2009). Rhetorical strategies in Chinese and English: A comparison of L1 composition textbooks. *Foreign Language Annals*, *42*(4), 695–720. https://doi.org/10.1111/j.1944-9720.2009.01050.x

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*(4), 451–484. https://doi.org/10.1017/S0142716400005762

Liu, L. (2005). Rhetorical education through writing instruction across cultures: A comparative analysis of select online instructional materials on argumentative writing. *Journal of Second Language Writing*, *14*(1), 1–18.

https://doi.org/10.1016/j.jslw.2004.11.001

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (Ed.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Multilingual Matters.

Loewen, S., & Inceoglu, S. (2016). The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense. *Studies in Second Language Learning and Teaching*, *6*(1), 89–110. https://doi.org/10.14746/ssllt.2016.6.1.5

Long, M. H. (2010). Methodological principles for language teaching. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 373-394). Wiley Blackwell.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge University Press.

Lupyan, G., & Clark, A. (2015). Words and the world. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition*, *21*(4), 557–587. https://doi.org/10.1017/S0272263199004027

Mackey, A., & Gass, S. (2016). *Second Language Research: Methodology and Design*. Routledge. https://doi.org/10.4324/9781315750606

MacWhinney, B. (2005). A unified model of language acquisition. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). Oxford University Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Mak, P., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language. 47*(1). 50-68. https://doi.org/10.1006/jmla.2001.2837.

Marinis, T. (2010). Using on-line processing methods in language acquisition research. In E. Blom & S. Unsworth (Eds.). *Experimental Methods in Language Acquisition Research* (pp.139-162). https://doi.org/10.1075/lllt.27.09mar

Marinis, T. (2012). Online psycholinguistic methods in second language acquisition research. In C.A. Chapelle (Ed.) *The Encyclopedia of Applied Linguistics,* https://doi.org/10.1002/9781405198431.wbeal0884

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(1). https://doi.org/10.1017/S0272263105050035

Marinis, T., & Saddy, D. (2013). Parsing the passive: Comparing children with specific language impairment to sequential bilingual children. *Language Acquisition*, *20*(2), 155–179. https://doi.org/10.1080/10489223.2013.766743

Marsden, E. (2006), Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning, 56*(3), 507-566. https://doi.org/10.1111/j.1467-9922.2006.00375.x

Marsden, E., & Chen, H. Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning*, *61*(4), 1058-1098.

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861–904. https://doi.org/10.1017/S0142716418000036

Marsden, E., & Torgerson, C. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education, 5,* 583-616.

Marsden, E., Williams, J., & Liu, X. (2013). Learning novel morphology. *Studies in Second*

Language Acquisition, *35*(4), 619–654. https://doi.org/10.1017/S0272263113000296

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

Master, P. (1991). Active verbs with inanimate subjects in scientific prose. *English for Specific Purposes*, *10*(1), 15–33. https://doi.org/10.1016/0889-4906(91)90013-M

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

McManus, K., & Marsden, E. (2017). L1 explicit instruction can improve L2 online and offline performace. *Studies in Second Language Acquisition*, *39*(3), 459–492. https://doi.org/10.1017/S027226311600022X

McManus, K., & Marsden, E. (2019). Signatures of automaticity during practice: Explicit instruction about L1 processing routines can improve L2 grammatical processing. *Applied Psycholinguistics*, *40*(1), 205–234. https://doi.org/10.1017/S0142716418000553

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Miao, X., Chen, G., & Ying, H. (1986). Sentence comprehension in Chinese. In Z. Man-Shu (Ed.), *Studies in Child Language Development*. Shanghai: East China University Press.

Mitchell, R., Myles, F., & Marsden, E. (2019). Second language learning theories. In *Second Language Learning Theories: Fourth Edition*. https://doi.org/10.4324/9781315617046

Morales, J., Padilla, F., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Simultaneous interpretation selectively influences working memory and attentional networks. *Acta*

Psychologica, *155*, 82–91. https://doi.org/10.1016/j.actpsy.2014.12.004

Morgan-Short, K., & Bowden, H. W. (2006). Processing instruction and meaningful output-based instruction: Effects on second language development. *Studies in Second Language Acquisition*, *28*(1), 31–65. https://doi.org/10.1017/S0272263106060025

Morgan-Short, K., Marsden, E., Heil, J., Issa II, B. I., Leow, R. P., Mikhaylova, A., … Szudarski, P. (2018). Multisite replication in second language acquisition research: Attention to form during listening and reading comprehension. *Language Learning*, *68*(2), 392–437. https://doi.org/10.1111/lang.12292

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, *50*(4), 617–673. https://doi.org/10.1111/0023-8333.00142

Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, *13*(1), 34–65. https://doi.org/10.1191/026765897671676818

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal*, *77*(3), 330–339. https://doi.org/10.1111/j.1540-4781.1993.tb01980.x

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nassaji, H., & Fotos, S. (2004). Current developments in research on the teaching of grammar. *Annual Review of Applied Linguistics*, *24*. https://doi.org/10.1017/S0267190504000066

Niu, X. (in progress). *To what extent can teaching predictive and retrospective parsing strategies benefits L2 morphosyntax learning?* [Unpublished doctoral thesis]. The University of York.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*(3), 417–528.

https://doi.org/10.1111/0023-8333.00136

Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly*, *25*(2), 279. https://doi.org/10.2307/3587464

O'Grady, W. (2005). Syntactic Carpentry. In *Syntactic Carpentry: An Emergentist Approach to Syntax*. https://doi.org/10.4324/9781410612571

O'Reilly, D., & Marsden, E. (2020). Eliciting and measuring L2 metaphoric competence: Three decades on from Low (1988). *Applied Linguistics*. 1-37. https://doi.org/10.1093/applin/amz066

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*(S1), 199–230. https://doi.org/10.1111/j.1467-9922.2006.00361.x

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., … Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, *21*(6), 509–521. https://doi.org/10.1016/j.jneuroling.2008.01.001

OUP Oxford (2014). In *Concise Oxford Dictionary of Linguistics*. Retrieved April 12, 2020, from https://www.oxfordreference.com/view/10.1093/acref/9780199675128.001.0001/acref-9780199675128

Patson, N. D., Darowski, E. S., Moon, N., & Ferreira, F. (2009). Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 280–285. https://doi.org/10.1037/a0014276

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*(4), 398–422. https://doi.org/10.3758/BF03202613

Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, *5*(4), 409–453.

https://doi.org/10.1075/lab.5.4.01phi

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pienemann M. (2010) A cognitive view of language acquisition: Processability theory and beyond. In P. Seedhouse, S. Walsh, & C. Jenks (Eds.) *Conceptualising 'Learning' in Applied Linguistics*, (pp.69-88). Palgrave Macmillan. https://doi.org/10.1057/9780230289772_5

Plonsky, L. (2013). Study quality in SLA. *Studies in Second Language Acquisition*, *35*(4), 655–687. https://doi.org/10.1017/S0272263113000399

Plonsky, L. and Derrick, D. J. (2016), A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, *100*(2), 538-553. https://doi.org/10.1111/modl.12335

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, *102*(4), 713–731. https://doi.org/10.1111/modl.12509

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2019). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*. https://doi.org/10.1177/0267658319828413

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Primus, B. (1998). The relative order of recipient and patient in the languages of Europe. In A. Siewierska (Ed.), *Constituent order in the languages of Europe.* Mouton de Gruyter.

Primus, B. (2010). Case-marking typology. In J. J., Song (Ed.) *The Oxford Handbook of Linguistic Typology*. https://doi.org/10.1093/oxfordhb/9780199281251.013.0016

Qualtrics (2005) Qualtrics survey software. *Qualtrics.* https://www.qualtrics.com

Qin, J. (2008). The effect of processing instruction and dictogloss tasks on acquisition of the English passive voice. *Language teaching research*, *12*(1), 61-82. https://doi.org/10.1177/1362168807084494

Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced past passive errors in English.* [Doctoral thesis]. University of Toronto.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). A grammar of contemport English. Seminar Press.

R Core Team (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reber, A. S., Allen, R., & Reber, P. (1999). Implicit vs. explicit learning. In R. J. Sternberg (Ed.), *The Nature of Cognition* (pp. 475-514). MIT Press.

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness. *Studies in Second Language Acquisition*, *37*(2), 299–334. https://doi.org/10.1017/S0272263115000145

Rebuschat, P., & Williams, J. N. (2012). Implicit learning in second language acquisition. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0529

Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, *7*, 361. https://doi.org/10.3389/fnhum.2013.00361

Revelle, W. (2018). *Using R and the psych package to find ω*. http://personality-project.org/r/psych/HowTo/omega.pdf

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1), 31–40. https://doi.org/10.1016/0028-3932(87)90041-8

Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, *28*(1), 113–127. https://doi.org/10.1177/0267658311418416

Roberts, L. (2013). Sentence processing in bilinguals. In R. P. G. van Gompel (Ed.), *Sentence Processing* (pp. 221–246). Psychology Press.

Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, *32*(2), 299–331. https://doi.org/10.1017/S0142716410000421

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, *30*(3), 333–357. https://doi.org/10.1017/S0272263108080480

Roberts, L., & Liszka, S. A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, *29*(4), 413–439. https://doi.org/10.1177/0267658313503171

Roberts, L., & Liszka, S. A. (2019). Grammatical aspect and L2 learners' online processing of temporarily ambiguous sentences in English: A self-paced reading study with German, Dutch and French L2 learners. *Second Language Research*. https://doi.org/10.1177/0267658319895551

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, *45*(2), 283–331. https://doi.org/10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, *19*(2), 223–247. https://doi.org/10.1017/S0272263197002052

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. Long (Eds.). *Handbook of Second Language Acquisition* (pp.631-678). Blackwell

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language 57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

Rosa, E. M., & Leow, R. P. (2004). Computerized task-based exposure, explicitness, type of feedback, and Spanish L2 development. *The Modern Language Journal*, *88*(2), 192–216. https://doi.org/10.1111/j.0026-7902.2004.00225.x

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Russell, M. (2014). *A comparison of linguistic features in the academic writing of advanced English language learner and English first language university students.* [Unpublished master's thesis]. Portland State University.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: when are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430. https://doi.org/10.1177/0267658308090186

Sagarra, N., & Herschensohn, J. (2011). Proficiency and animacy effects on L2 gender agreement processes during comprehension. *Language Learning*, *61*(1), 80–116. https://doi.org/10.1111/j.1467-9922.2010.00588.x

Sanz, C. (2003). Computer delivered implicit vs. explicit feedback in processing instruction. In B. VanPatten (Ed.) *Processing Instruction* (pp.241-256). Erlbaum.

Sanz, C. & Leow, R. (2011). *Implicit and Explicit Language Learning*. Georgetown University Press. Retrieved from http://www.jstor.org/stable/j.ctt2tt7k0

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning*, *54*(1), 35– 78.

Sanz, C., & Morgan-Short, K. (2005). Explicitness in pedagogical interventions: Input, practice, and feedback. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice* (pp. 234–263). Georgetown University Press.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.). *Attention and Awareness in Foreign Language Learning*. (pp.1-63). University of Hawai'i.

Schmidt, R. (2001). Cognition and second language instruction. In P. Robinson (Ed.), *Cognition and Second Language Instruction*. (pp. 3-32). https://doi.org/10.1017/CBO9781139524780

Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistics methodology*. The University of Chicago Press.

Schumacher, P. B., Roberts, L., & Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua*, *185*, 25-41. https://doi.org/10.1016/j.lingua.2016.07.004

Schwartz, B. D. (1993). On explicit and negative data effecting and affecting “competence and linguistic behaviour”. *Studies in Second Language Acquisition*, *15*(2), 147–163. Retrieved from http://www.jstor.org/stable/44487615

Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385. https://doi.org/10.1017/S0142716400010845

Seyednejad, H., & Gholami, H. (2017). The effect of structured input, meaningful output and traditional instruction on EFL learners’ productive use of passive voice. *Theory and Practice in Language Studies*, *7*(9), 737. https://doi.org/10.17507/tpls.0709.05

Sharwood-Smith, M. (1981). Consciousness-raising and the second language learner, *Applied Linguistics*, *2*(2), 159–168. https://doi.org/10.1093/applin/II.2.159

Sharwood-Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*(2). 165 - 179. https://doi.org/10.1017/S0272263100011943.

Sheliga, B. M., Craighero, L., Riggio, L., & Rizzolatti, G. (1997). Effects of spatial attention on directional manual and ocular responses. *Experimental Brain Research*, *114*(2), 339–351. https://doi.org/10.1007/PL00005642

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research*, *98*(3). https://doi.org/10.1007/BF00233988

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language learning*, *63*(2), 296-329. https://doi.org/10.1111/lang.12001

Simard, D., & Foucambert, D. (2013). Observing noticing while reading in L2. In J. Bergsleithner, S. Nagem Frota, & J. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 207–227). Manoa, Hawai'i: National Foreign Language Resource Center, University of Hawai'i.

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*(3), 723–751. https://doi.org/10.1111/lang.12129

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*(2), 263. https://doi.org/10.1111/j.1467-9922.2010.00562.x

SR-Research. (2011). SR Research Experiment Builder User Manual. *SR Research Ltd.*

Stafford, C. A., Bowden, H. W., & Sanz, C. (2012). Optimizing language instruction: Matters of explicitness, practice, and cue learning. *Language Learning*, *62*(3), 741–768. https://doi.org/10.1111/j.1467-9922.2011.00648.x

Stromswold, K., Eisenband, J., Norland, E., & Ratzan, J. (2002). Tracking the acquisition & processing of English passives: Using acoustic cues to disambiguate actives & passives. *CUNY Conference on Sentence Processing*.

Su, I.-R. (2001). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied Psycholinguistics*, *22*(1), 83–112. https://doi.org/10.1017/S0142716401001059

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*(5), 1229–1261. https://doi.org/10.1017/S014271641700011X

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*(4), 860–895. https://doi.org/10.1111/lang.12138

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747–790. https://doi.org/10.1111/lang.12241

Tamminen, J., Payne, J.D., Stickgold, R., Wamsley, E.J., & Gaskell, M.G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, *30*, 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

The Concise Oxford Dictionary of Linguistics. (2014). In *The Concise Oxford Dictionary of Linguistics*. https://doi.org/10.1093/acref/9780199675128.001.0001

Tight, D. G. (2012). The first noun principle and ambitransitive verbs. *Hispania*, *95*(1), 103–115.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204. https://doi.org/10.1017/S0272263105050102

Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, *22*(7), 1092–1106. https://doi.org/10.1080/09541440903325178

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 183–203. https://doi.org/10.1017/S0272263100012870

Torgerson, C. J., & Torgerson, D. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Education Studies*, *49*(3), 316-328. https://doi.org/10.1111/1467-8527.t01-1-00178

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47*, 69–90. https://doi.org/10.1006/jmla.2001.2836

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, andverb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204–224. https://doi.org/10.1016/j.jml.2005.02.010

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition*, *33*(3), 339–372. https://doi.org/10.1017/S0272263111000015

Trenkic, D., Mirkovic, J., & Altmann, Gerry, T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, *17*(2), 237–257. https://doi.org/10.1017/S1366728913000321

UK Council for International Affairs (2019). *International student statistics.* https://www.ukcisa.org.uk/

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in* *adult second language acquisition* (pp. 141– 178). Georgetown University Press.

Ullman, M., & Lovelett, J. (2016). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research. 34*(1), 39-65. https://doi.org/10.1177/0267658316675195.

Uludag, O., & VanPatten, B. (2012). The comparative effects of processing instruction and dictogloss on the acquisition of the English passive by speakers of Turkish. *International Review of Applied Linguistics in Language Teaching*, *50*(3). https://doi.org/10.1515/iral-2012-0008

Van Heugten, M., & Johnson, E. K. (2011). Gender-marked determiners help Dutch learners' word recognition when gender information itself does not. *Journal of Child Language*, *38*(1), 87–100. https://doi.org/10.1017/S0305000909990146

van Ek, J. A., & Trim, J. L. M. (1999). Waystage 1990. In *Waystage 1990*. https://doi.org/10.1017/CBO9780511667107

van Ek,.J.A. & .Trim, J.L.M. (2001). *Vantage*. Cambridge University.Press.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Ablex Publishing Corporation. https://doi.org/10.1191/026765898677993396

VanPatten, B. (2000). Processing instruction as form-meaning connections: Issues in theory and research. In J. F. Lee Ed & A. Valdman Ed (Eds.), *Form and Meaning: Multiple Perspectives. Issues in Language Program Direction: A Series of Annual Volumes* (pp. 43–68).

VanPatten, B. (2002). Processing instruction, prior awareness and the nature of second language acquisition: A (partial) response to Batstone. *Language Awareness*, *11*(4), 240–258. https://doi.org/10.1080/09658410208667059

VanPatten, B. (2004). Input processing in SLA. In B. VanPatten (Ed.). *Processing Instruction: Theory, Research, and Commentary*. (pp.5-33). Routledge. https://doi.org/10.4324/9781410610195

VanPatten, B. (2018). Processing instruction. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–7). Wiley. https://doi.org/10.1002/9781118784235.eelt0094

Vanpatten, B., & Cadierno, T. (1993). Input processing and second language acquisition: A role for instruction. *The Modern Language Journal*, *77*(1), 45–57. https://doi.org/10.2307/329557

VanPatten, B, Collopy, E., Price, J. E., Borst, S., & Qualin, A. (2013). Explicit information, grammatical sensitivity, and the first-noun principle: A cross-linguistic study in processing instruction. *The Modern Language Journal*, *97*(2), 506–527. https://doi.org/10.1111/j.1540-4781.2013.12007.x

VanPatten, B., Collopy, E., & Qualin, A. (2012). Explicit information and processing instruction with nominative and accusative case in Russian as a second language: Just how important is explanation? *Slavic and East European Journal*, *56*(2), 256–276.

VanPatten, B., & Houston, T. (1998). Contextual effects in processing L2 input sentences. *Spanish Applied Linguistics*, *2*, 53–70.

VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. R. Eckman, J. Mileham, R. Rutkowski Weber, D. Highland, P. W. Lee (Eds.) *Second Language Acquisition Theory and Pedagogy*. (pp.169-187). Taylor & Francis.

VanPatten, B, & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition*, *18*(4), 495–510. https://doi.org/10.1017/S0272263100015394

VanPatten, B, & Uludag, O. (2011). Transfer of training and processing instruction: From input to output. *System*, *39*(1), 44–53. https://doi.org/10.1016/j.system.2011.01.013

VanPatten, B., & Wong, W. (2004). Processing instruction and the French causative: Another replication. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 97-118). New Jersey: Lawrence Erlbaum Associates, Inc.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in pavlovian conditioning: Application of a theory. In M. S. Halliday & R. A. Boakes (Eds.), *Inhibition and Learning* (pp. 301–336). San Diego, CA.: Academic Press.

Wang, P., & Wang, W. (2012). Causes of and remedies for Chinglish in Chinese college students’ writings. *Open Journal of Modern Linguistics, 2*(2), 71–78. https://doi.org/10.4236/ojml.2012.22009

Weckerly, J., Kutas, M. (1999) An electrophysiological analysis of animacy effects in the processing of object-relative sentences. *Psychophysiology, 36,* 559-570

Weighall, A.R., Henderson, L.M., Barr, D.J., Cairney, S.A., & Gaskell, M.G. (2016). Eye-tracking the time- course of novel word learning and lexical competition in adults and children. *Brain and Language, 167,* 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics*, *12*(4), 416–432. https://doi.org/10.1093/applin/12.4.416

Williams, J. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*(2), 269–304. https://doi.org/10.1017/S0272263105050138

Williams, J., & Evans, J. (1998). What kind of focus and on which form? In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 139–155). Cambridge University Press.

Williams, J., Möbius, P., & Kim, C. (2001). Native and non-native processing of English wh - questions: Parsing strategies and plausibility constraints. *Applied Psycholinguistics*, *22*(4), 509–540. https://doi.org/10.1017/S0142716401004027

Williams, J., & Rebuschat, P. (2016). *Implicit learning and second language acquisition*. Routledge.

Wills, A. J. (2009). Prediction errors and attention in the presence and absence of feedback. *Current Directions in Psychological Science*, *18*(2), 95–100.

https://doi.org/10.1111/j.1467-8721.2009.01616.x

Winke, P. M. (2013). The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition*, *35*(2), 323–352. https://doi.org/10.1017/S0272263112000903

Winter, B. (2019). *Applied statistics for linguists: An Introduction Using R.* New York, Routledge.

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition*, *23*(3), 345–368. https://doi.org/10.1017/S0272263101003023

Wong, W. (2004). The nature of processing instruction. In B. VanPatten (Ed.), *Processing instruction* (pp. 33–63). Erlbaum.

Wong, W., & Ito, K. (2018). The effects of processing instruction and traditional instruction on L2 online processing of the causative construction in French: An eye-tracking study. *Studies in Second Language Acquisition*, *40*(2), 241–268. https://doi.org/10.1017/S0272263117000274

Wright, R. D., & Ward, L. M. (2008). *Orienting of attention.* Oxford University Press.

Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in Contrast*, *6*(1), 109–149. https://doi.org/10.1075/lic.6.1.05xia

Yulin, Y. (2012). *Focus structure and its semantic interpretation in Chinese sentences*. Beijing: Commercial Press.

Zarcone, A., van Schijndel, M., Vogels, J., & Demberg, V. (2016). Salience and attention in surprisal-based accounts of language processing. *Frontiers in Psychology*, *7*(844). https://doi.org/10.3389/fpsyg.2016.00844

Zhang, R. (2015). Measuring university level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*(3), 457–486.

https://doi.org/10.1017/S0272263114000370