Processing passive sentences in English: A cross-methodological study

Caterina Laura Paolazzi

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

UCL, Department of Linguistics

May 2018

Declaration

I, Caterina Laura Paolazzi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgments

This project would not have been possible without the guidance and support of my supervisor, Andrea Santi. Andrea is the person I have always looked up to throughout this journey, she has been for me a female role model, an example of strength and determination.

Thanks to the people who offered ideas, help and advice on this project: Wing Yee Chow, Nino Grillo, Claudia Cera, Nathan Klinedinst.

Thanks to my friends in London and Milano, always ready to celebrate successes or offer support in tougher times, but never without a glass of wine: Francesca Arcostanzo, Zoe Belk, Saverio Blasi, Lucia Camela, Cecilia Casetta, Irene De Pace, Giulio Dulcinati, Bruno Fernandes, Matilde Marangoni, Eleonora Rossi, Margherita Sbanchi, Irini Symeonidou.

Thanks to my family, my root and my rock. All that I am, all that I hope to be, I owe it to you. Although we were far apart, I never felt closer to you all.

Thanks to Giacomo for your unwavering faith in me and in us, for the endless smiles and kisses, eight days a week. Nothing would be half as beautiful without you.

Abstract

Passives are considered more difficult to process than actives. The existing literature presents contrasting findings: passives are read faster than actives in online data, but more errorful in offline ones. A potential source for this contrast was previously overlooked: passivization and predicate semantics interact, in the sense that passivized statives are more difficult to process than actives, while no difference should be found across eventives. Evidence for this interaction was reported in the acquisition and aphasia literature, but not yet explored in healthy adults.

Experiments 1-4 coupled self-paced reading with a verification task to compare offline and online measures, while manipulating predicate semantics, and replicated the contrast, additionally showing that variability in working memory is partially responsible for the offline complexity. No evidence for an interaction was found. Experiments 5-7 used acceptability judgments together with a verification task to investigate the causes of the offline difficulty and found that passives are susceptible to characteristics of the offline tasks, which increase memory-base interference. Once these characteristics were controlled for, an interaction emerged in the offline data, in the expected direction. Finally, Experiment 8 investigated the lack of evidence for an interaction in online data. The main causes for the complexity in passivized statives are coercion and revision, whose effects have only been detected in late, and not early, measures of processing. Eye-tracking while reading, which allows to collect early and late measures of processing, was employed. Evidence for an interaction between passivization and predicate semantics was found in the expected direction.

Overall, passives are not more difficult to process than active ones. Rather they are susceptible to characteristics of the offline task employed and of the predicate semantics, which do not affect processing of active sentences. Future studies in psycholinguistics (e.g., aphasia, language acquisition) could be informed by these findings.

Table of Contents

List of Figures

List of Tables

1 Passives in psycholinguistics: Introduction.

1.1 Introduction to passives.

Passive sentences have been the centre of linguistic research for the past 40 years across various sub-disciplines and populations (healthy adults – e.g., Ferreira, 2003 –, aphasic patients – e.g., Grodzinsky, 1990 –, children acquiring their first language – e.g., Maratsos, Fox, Becker & Chalkley, 1985). The general agreement across these fields is that passives are more complex to process than actives. This increased complexity has been attributed to: (1) an inverse argument order (patient-verb- agent, see (1)) with respect to the canonical, active one (agent-verb-patient, see (2); Christianson, Hollingworth, Halliwell & Ferreira, 2001; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016), (2) an additional syntactic operation that derives passive sentences from active ones (Chomsky, 1981), or (3) the lower frequency of passive sentences with respect to active ones in language use (Johns & Jones, 2015).

(1) The girl $_{\text{PATIENT}}$ was pushed by the boy $_{\text{AGENT}}$.
(2) The girl $_{\text{AGENT}}$ pushed the boy $_{\text{PATIENT}}$.

The general consensus on passivization being complex, along with the underlying morpho-syntactic (rather than lexical) differences between actives and passives has resulted in the passive-active contrast often being employed as a metric of syntactic complexity to assess syntactic processing abilities across various populations. This contrast has in fact been used to identify areas of syntactic processing in the brain of healthy adults (e.g., Mack, Meltzer-Asscher, Barbieri & Thompson, 2013), to establish patterns of impairment in syntactic comprehension of aphasic patients (e.g., Caplan, Waters, DeDe, Michaud & Reddy, 2007; Thothathiri, Kim, Trueswell & Thompson-Schill, 2012) and to investigate the development of semantic and syntactic abilities in children acquiring their first language (e.g., Maratsos et al., 1985).

Generally, it is desirable to understand any processing complexity in healthy adults, before using it as a test of particular processing abilities in other populations. The debate surrounding the source of its complexity makes evident that this is not the case with passive sentences. Moreover, psycholinguistic studies of passive sentence

processing present a heterogeneous picture: they show no evidence of processing difficulty in online reading tasks (Carrithers, 1989; Traxler, Corina, Morford, Hafer & Hoversten, 2014), but they are more errorful in offline comprehension tasks (Ferreira, 2003). These results thus provide mixed support to the accounts of passive sentence processing: the *Good Enough* model, the syntactic complexity approach and the frequentist/usage-based approach. The next section will be devoted to a discussion of these accounts, thereafter the psycholinguistic data will be discussed in detail in the context of the predictions of these three theories. Given that all three models fail to account for the full set of data, two further theoretical frameworks – surprisal-based and expectation-based models, and verbal Working Memory – that have never been applied to the processing of passivization will be considered in explaining the full set of data. Finally, the relevance of predicate semantics for the observed passivization complexity effects will be discussed with predictions for the healthy adult population, based on the results collected in studies on children and people with aphasia. The chapter ends by outlining the predictions that emanate from these theories and briefly summarizing the experiments that test them in Chapters 2-4.

1.2 Approaches to passive sentences processing.

As briefly mentioned in the previous section, passives are considered more complex to process than actives. This additional complexity is generally attributed to (1) revision of a syntactic heuristic that produces an inaccurate interpretation for passive sentences, as per the *Good Enough* model, (2) resolution of a syntactic dependency not present in active sentences, as per the syntactic complexity approach, or (3) lower frequency in language use. This section will now consider these three approaches to passive sentences processing in more detail.

1.2.1 Heuristics revision and the Good Enough model.

Ferreira (2003) found healthy adults to be more errorful in their comprehension of passives than actives and argued that the results provided evidence in favour of the *Good Enough* model of sentence processing, firstly proposed by Christianson and colleagues (Christianson et al., 2001).

Ferreira's study (2003) aimed at investigating how people assign thematic roles to the various participants in a sentence, and for this reason, the study manipulated the argument order, in terms of their thematic roles. A canonical order presented agent before patient, such as in active and subject-cleft sentences, whereas a non-canonical order presented patient before agent, such as in passive and object-cleft sentences. Ferreira's sentences also varied in semantic plausibility (reversible but highly biased arguments, see (3); non-reversible, see (4); symmetrical, see (5)).

(3) The dog bit the man. / The man bit the dog.
(4) The mouse ate the cheese. / The cheese ate the mouse.
(5) The woman visited the man. / The man visited the woman.

In a series of three experiments, participants listened to sentences and were then asked to identify the doer (i.e., the agent) or the acted-on (i.e., the patient) of the action described in each sentence. The primary dependent measure was accuracy, but reaction times were analysed as well. Across the three conditions, non-canonical sentences, i.e., passives and object-clefts, were found to be more errorful than canonical ones, i.e., actives and subject-clefts (see Table 1). Moreover, the non-canonical sentences elicited longer reaction times when answering comprehension questions, with respect to canonical ones. The same results were replicated by Christianson and colleagues (Christianson, Luke & Ferreira, 2010).

	Active	Passive	Subject-cleft	Object-cleft
Reversible, biased	96.5%	84.75%	94.25%	75.5%
Non-reversible	96.25%	85%	92.75%	74.75%
Symmetrical	93.25%	84.25%	94%	76.5%

Table 1: Average accuracy across experiments in Ferreira (2003).

The model proposed by Christianson and colleagues (2001), and later reformulated by Ferreira and colleagues (Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016) to account for these findings is the *Good Enough* parsing model. This model posits that sentence processing is based on both heuristics and algorithmic processes, which are launched simultaneously and in parallel. Algorithmic processes are syntactic operations (e.g., combining words

according to the rules of syntax, assigning thematic roles), which allow to create a precise and complete representation of the linguistic input. Two different heuristics have insofar been proposed in the *Good Enough* literature: a syntactic heuristic and a semantic/pragmatic heuristic. The heuristic that will be discussed here is the syntactic one, which could account for the passivization difficulty in comprehension, under plausible conditions, as this is the focus of the dissertation. A syntactic-based heuristic in English is the agent-first strategy, i.e., the first NP of every English sentence is initially interpreted as agent (e.g., in (6) and (7) , both subjects would be interpreted as agent under this heuristic). Such a heuristic is considered to be "fast and frugal" with respect to the slower algorithmic processes. This heuristic can generally produce a fast and correct output when a sentence complies with it, i.e., the order of the arguments is agent-patient, as in active sentences. However, when the order of the arguments does not abide by this heuristic, as in a passive sentences where the first NP is the patient and not the agent of the action, the heuristic parse will produce an incorrect interpretation of the sentence. In this second scenario, the parser will need to decide whether to consider the heuristic output "good enough" or to start a time-consuming and costly reanalysis of the heuristic output and integration of the slow but correct algorithmic processes to reach its final interpretation.

(6) The girl was pushed by the boy.
(7) The girl pushed the boy.

Ferreira and colleagues (Karimi & Ferreira, 2016; see Figure 1) further specify that the heuristic output should be available quite early in processing, and long before the algorithmic processes have produced an output. Once it is produced, the heuristic output can, at this point, alter the nature of the algorithmic processes, which would become confirmatory and would not receive further resources to produce an independent output. Alternatively, the heuristic output can be refined by the algorithmic processes, thus leading to revision, reflected by a slow-down in processing.

The *Good Enough* model thus predicts two possible outcomes when processing a passive sentence: (1) inaccurate comprehension of the sentence due to the incorrect output produced by the heuristic, if judged "good enough", or (2) correct comprehension of the sentence due to reanalysis/adjudication between the heuristic

and algorithmic processes. As shown above in Table 1, Ferreira (2003) found passive sentences comprehension to be very high, although lower than active sentences comprehension, with an 11.5% difference between the two averages (average active-average passive). Hence, revision/integration of the heuristic and algorithmic outputs must occur more often than not, as signalled by the high level of accuracy to passive sentences (i.e., much higher than floor levels that would occur if participants solely relied on the heuristic). However, the lower accuracy in passives compared to actives provides evidence in favour of the *Good Enough* parsing model, and the idea that the parser sometimes judges the interpretation provided by the heuristics "good enough" even though it is incorrect. These data were replicated by Christianson and colleagues (2010).

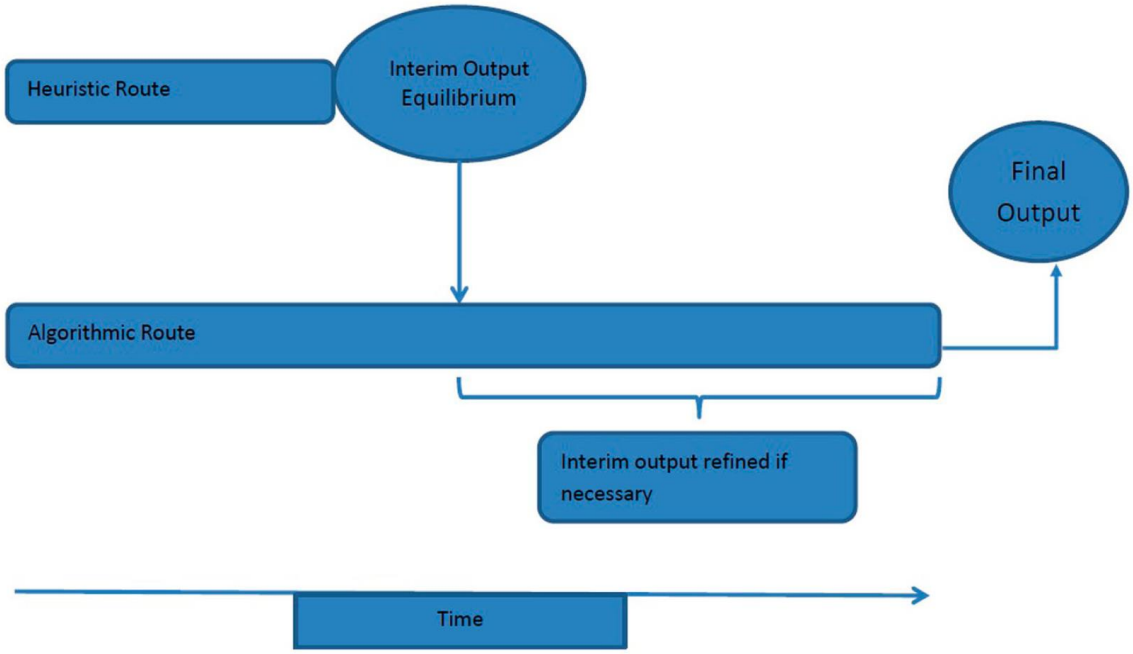

Figure 1: The model of language processing according to the Good Enough model (from Karimi & Ferreira, 2016).

1.2.2 Syntactic approach to passives complexity.

In contrast to Ferreira's model, other mainstream sentence processing models solely rely on algorithmic processes (Kiparsky, 2013). As previously mentioned, algorithmic processes build syntactic representations based on the rules of syntax (i.e., parsing) and simultaneously derive meaning from them (i.e., interpretation). According to these models, difficulties in parsing should hence arise when additional syntactic

operations are involved or, given the limited resources available to the parser, greater demands on syntactic parsing/interpretation are posed.

Passive sentences would present difficulties under this model, as they are thought to require an additional syntactic operation that poses greater demands on the parser. Specifically, passive sentences result from a transformation of the canonical, active structure into a non-canonical one, which is carried out through a series of operations (argument suppression, case absorption, and argument enhancement) mainly realized via movement (e.g., Chomsky 1981). This movement operation would moreover create a syntactic dependency in the passive sentence between the phonologically overt subject (i.e., the sentence initial NP "The girl" in (8)) and the phonologically empty position (i.e., so-called trace) where the subject is interpreted for its thematic role (marked by angled brackets in (8)). In psycholinguistic research on sentence processing, the overt NP is referred to as the "filler" and the thematic role position as "the gap". This filler-gap syntactic dependency (marked in bold in (8)) needs to be resolved by the parser for accurate thematic role assignment.

(8) **The girl** was pushed **<the girl>** by the boy.

An alternative analysis of passivization (e.g., Bresnan, 1982) proposes that passives are derived from lexical rules that either change the grammatical function of the arguments (in the case of verbal passives) or the actual semantics of the lexical form (in the case of adjectival passives). Under this approach, which derives from the functional lexical-grammar framework, both actives and passives are base-generated and stored in the lexicon.

Syntactic complexity has been linked to more processing difficulty in online and offline measures with other movement-derived structures (e.g., object vs. subject relative clauses; Caplan et al., 2007; Garnham & Oakhill, 1987; Gibson, 1998; Grodner & Gibson, 2005; Hakes, Evans & Brannon, 1976; Santi & Grodzinsky, 2010; Staub, 2010; Traxler, Morris & Seely, 2002). For example, relative clauses are thought to be generated via movement of the relative pronoun from the gap position (underlying subject in the case of subject-relatives vs. underlying object in the case of object-relatives) to the SpecCP (filler position), leaving a trace in the original gap position (Haegeman, 1991). While the pronoun in a subject relative clause does not have to cross any similar

constituent on its path to the filler position, the pronoun in an object relative clause will have to cross the subject DP. This movement across a similar constituent is argued to create additional difficulties in processing object with respect to subject-relative clauses (Gordon, Hendrick & Johnson, 2001).

While the transformation approach would predict passives to be more difficult to process than actives, the lexicalist approach to passivization would not predict an increased processing burden on passives with respect to actives, due to the fact that both are base-derived and equally stored in the lexicon. Considering the substantial evidence from language acquisition, aphasia, and offline processing of healthy adults, supporting the idea that passives are more taxing than actives (see Sections 1.2.1, 1.6.2 and 1.6.3), the transformational approach appears to account more adequately for the underlying nature of the passivization process.

The data reported in Ferreira (2003) are hence compatible with the transformational approach. The increased syntactic complexity attributed to passives has indeed been used to account for the greater activation for passives with respect to actives in healthy adults' brain (Mack et al., 2013), the late acquisition of passives compared to actives in children (Borer & Wexler, 1987), and selective difficulties in aphasic patients' production and comprehension of passive sentences with respect to active ones (Dickey & Thompson, 2009; Grodzinsky, 1990, 2000).

1.2.3 Frequentist approach to sentence complexity.

Differently from both the *Good Enough* and the syntactic complexity models, the frequentist approach argues that the processing complexity of a syntactic structure is determined by the amount of exposure to that particular structure (Johns & Jones, 2015). The more frequently a structure is used by native speakers, the easier its parsing will be. Given that passive sentences are less frequent in use than active ones (Gordon & Chafetz, 1990), this model predicts passives to be more complex to process than actives.

Although no specific prediction has been made by supporters of this approach with respect to passive sentences, other studies adopting this approach argue for both online and offline complexity effects for less frequent structures with respect to more frequent ones (e.g., object relative vs. subject relative clauses; MacDonald, 2013; Reali

& Christiansen, 2007). Hence, the frequentist approach would predict passives to be more errorful than actives, thus faring well under Ferreira's (2003) data on passivization. However, this same approach would predict subject-clefts to be more difficult to process than active sentences, as the former are much less frequent than the latter in language use. Contrary to this prediction, Ferreira (2003) found subject-clefts to be processed as accurately as active sentences, thus providing evidence against the frequentist approach to sentence processing.

1.3 Predictions of models of passives processing with respect to existing data.

While existing offline data on passive sentences processing provide support in favour of the *Good Enough* model and the syntactic complexity approach, it is fundamental to consider existing online data on passives processing, to further assess the explanatory adequacy of these models' predictions with respect to how healthy adults process passive sentences.

1.3.1 Online data.

All three approaches to passive sentences processing described in the previous section would predict passives to be processed more slowly than actives online.

According to the Good Enough model, passives should be slower to process than actives due to the necessary revision of the incorrect agent-first heuristic output and/or the integration of the algorithmic processes output into the final interpretation of the sentence. This revision/integration should occur at a point where the parser can no longer entertain the heuristic analysis (i.e., interpret the first NP as agent) as the correct one. The earliest point at which revision/integration should occur is the verb, as prior to this region other agent-first structures would still be possible (e.g., up to the verb, sentence (9) could share the same thematic structure – agent-patient – as sentence (10)). Evidence in favour of a revision/integration effect arising at the verb come from a visual world paradigm study (Kamide, Scheepers & Altmann, 2003), in which participants heard passive and active sentences (e.g., (11) vs. (12)) while inspecting a scene representing the participants in the action described by the sentence (e.g., a hare, a

cabbage, a fox). As soon as participants heard the participial form in the passive condition, which disambiguated between the active and passive condition, they immediately looked at the upcoming argument in the passive condition, i.e., “the fox”. This indicates that revision/integration occurs within the verbal region. However, a Cross Modal Lexical Priming study (CMLP; Osterhout & Swinney, 1993) suggests that revision/integration is delayed. In fact, Osterhout and Swinney found evidence of reactivation of the subject NP 500ms after the verb offset (i.e., post- gap) and up to 1000ms later for a reliable effect. Hence, the revision/integration effect predicted by the Good Enough model should be observed within a few words from the verb. On the contrary, actives would not require revision as their argument structure abides by the agent-first heuristics. Thus, the model predicts longer reading times within a few words post-verb for passives with respect to actives.

(9) The girl was pushed by the boy.

(10) The girl was pushing the boy.

(11) The hare will be eaten by a fox.

(12) The hare will eat the fox.

A similar prediction, i.e., slower processing in passives with respect to actives at or shortly after the verb, is offered by the syntactic complexity approach. According to this approach, passives are considered to be more complex than actives due to a movement operation of the theme/patient from its original position to the subject position (e.g., Chomsky, 1981). A processing complexity should hence arise once the parser resolves the thematic role dependency at the gap/trace left by the moved NP, similarly to what is observed in other movement-derived structures (e.g., object vs. subject relative clauses; Caplan et al., 2007; Garnham & Oakhill, 1987; Gibson, 1998; Grodner & Gibson, 2005; Hakes et al., 1976; Santi & Grodzinsky, 2010; Staub, 2010; Traxler et al., 2002). Several theories exist regarding the number of movements required to generate the passive structure from the active one, but they all minimally propose a movement relation between the subject NP and the verb internal argument position (e.g., after the verb – Chomsky, 1981; after the by-phrase – Collins, 2005; Gehrke & Grillo, 2007, 2009). Hence, it can be assumed that the complexity effect in passives with respect to actives would be expected to arise after the verb. Again, based on the CMLP

data reported above, it may be delayed by 500-1000ms, as it takes time for the parser to identify the gap and resolve its correct argument.

Finally, the frequentist approach would likewise predict passives to be slower to process than actives, although no speculation regarding the precise location of this effect is possible.

Let us turn to existing online data on passive sentence processing to see how these theories fare. In a subject-paced, word-by-word reading task, Carrithers (1989) found that passive sentences were read over 20 msec per word faster than active ones "after the first noun phrase had been processed" (p. 80), and the difference was significant. However, her paper does not specify the precise location of this difference, nor it provides example sentences, thus making the results difficult to interpret.

Traxler and colleagues (2014) used the same methodology, i.e., self-paced reading, but found that passive sentences (see (13) below) were read numerically, not significantly, faster than active ones (see (14) below), at the verb and object NP (e.g., both "tricked" and "cowboy" were read numerically faster in the passive version of the sentence). At the determiner ("the") of the by/direct object phrase the difference was only 1ms.

(13) The farmer was tricked by the cowboy into selling the horse.

(14) The farmer tricked the cowboy into selling the horse.

These data are *prima facie* inconsistent with the predictions for online processing derived from the *Good Enough* model, the syntactic complexity approach and the frequentist approach, given that passive sentences parsing was not overall slower than processing active sentences.

It is worthwhile to note that these studies (Carrithers, 1989; Traxler et al., 2014) did not collect/report comprehension accuracy data, and hence their results cannot represent a complete test of these theories, particularly the *Good-Enough* one. In fact, it is possible that participants relied more on the heuristic output in this online task, without revising its initial correct interpretation, thus speeding up processing times in passives.

Moreover, it is possible that the by-phrases employed in previous online designs (Carrithers, 1989; Traxler et al., 2014) were not sufficiently long to detect a complexity

effect in passives. As mentioned above, cross-modal lexical priming (Osterhout & Swinney, 1993) found a late reactivation of the filler-gap dependency in passives, suggesting that revision/integration of the passive structure might happen up to 1000ms after the verb, hence up to 3/4 words after the verb. If the by-phrases used in previous studies (Carrithers, 1989; Traxler et al., 2014) only contained 3 words (e.g., in (13), "by the cowboy"), it might not have been possible to detect a revision/integration effect in passives.

However, if passives should remain faster to process than actives in online tasks, even after controlling for these factors, then the *Good Enough* model, the syntactic complexity approach and the frequentist approach would be explanatorily inadequate with respect to the online data. In this scenario, it would thus be necessary to turn to alternative models of sentence comprehension, in order to find an adequate explanation.

1.4 An alternative approach to passive sentence processing: Expectation- and surprisal-based models.

Consistent with the online data reported in the literature on passive sentence processing are expectation-based (e.g., Levy, 2007) or surprisal-based models (e.g., Hale, 2001). In these models, the parser considers multiple alternative interpretations in an incremental fashion and will allocate parsing resources according to a ranking based on the plausibility of these interpretations (with the more plausible interpretation receiving more resources and being considered as the most-likely interpretation upon termination of the input). Once a word is encountered, the parser resources will serve the purpose to update the probability distribution over the complete structure to which that word might extend. Difficulty arises when the word encountered does not fit into the interpretation ranked as most probable up to now and hence the various interpretations must be re-ranked in order to accommodate for the new input. Hence, the difficulty in processing each word is a measure of its surprisal with respect to the overall most-probable interpretation built so far, and conversely a measure of its expectancy. It is important to underline that these models describe surprisal (or expectancy) as reflecting upon the processing difficulty of a specific word category rather than of a specific representational content of underlying syntactic structure. In

the ground-breaking paper in which he exposes his theory, Levy (2008) demonstrates the equivalence of the two models.

A series of data in the psycholinguistic literature appears to be entirely compatible with these models. For example, they adequately capture the degree of difficulty in processing a verb in a verb-final language, such as German, where the position of a pre-verbal object affects processing difficulty, e.g., faster reading times of the verb when the direct object immediately precedes the verb vs. when a different NP intervenes between the 2 constituents (every other characteristic of the experimental items being equal). The original author of the study in which these results were presented (Konieczny, 2000) provides an explanation for these results based on the distance of the dependents from their verb, where the farther the dependent is the more difficult the verb will be to process. On the contrary, Levy claims that processing the verb will be easier as the sentence lengthens. In fact, as soon as the parser realises that the input is part of a verb-final clause it will allocate a certain amount of probability distribution on the next constituent being a verb. This amount will increase after each non-verb constituent is processed as fewer alternatives will be available as possible sentence continuations. Hence, as the sentence lengthens, the probability and expectation of encountering a verb will increase, thus decreasing its surprisal and processing difficulty. The reading times reported in Konieczny (2000) perfectly reflect this hypothesis as the verb was read faster when a long PP intervened between direct object and verb compared to a short PP or no PP.

Moreover, support for these models come from studies on subject vs. object relative clauses (Staub, 2010). According to expectation-based and surprisal-based models, object relative clauses should create parsing difficulty given that they are more uncommon than subject relative clause. Hence, upon encountering *that,* the reader would expect a subject relative clause to continue the sentence. Encountering the subject of the object relative clause would then create a large surprisal effect (within surprisal models) or would be highly unexpected (according to expectation-based models), producing longer reading times, as supported by Staub (2010).

English passive sentences are signalled, at the surface level, by several morphological cues (auxiliary + past participle + “by”) that might increase expectations/reduce surprisal for upcoming grammatical categories in the passive

structure compared to the active one. According to expectation/surprisal models, the verb "pushed" should be more expected/create less surprisal in sentence (15), given that it is a very likely continuation after a subject NP and an auxiliary, than in sentence (16), where the subject NP could actually be followed by other syntactic categories (e.g., an adverb, a preposition, a relative pronoun). Similarly for the post-verbal argument DP, expectation is increased following the "by" in the passive sentence (15) than the verb in the active sentence (16). The same DP ("the boy") in sentence (16) might have to compete with other possible syntactic continuations (e.g., a preposition, an adverb), thus reducing its expectation and increasing surprisal with respect to the same DP in sentence (15).

(15) The girl was pushed by the boy.

(16) The girl pushed the boy.

These increased levels of expectation/reduced levels of surprisal could hence result in facilitation of processing in passive sentences with respect to active ones. According to these models, online processing of passive sentences could actually be faster than for their active counterparts, as was observed by Carrithers (1989) and by Traxler and colleagues (2014).

This account, however, does not predict the offline difficulty observed with passivization. It would thus be necessary to find an alternative explanation to account for the offline complexity effect in passive sentence comprehension reported in the literature. The next section will introduce a possible source of complexity in the offline comprehension tasks reported in the literature: verbal working memory.

1.5 The role of verbal working memory in passives processing.

Verbal working memory (vWM) is a temporally and size limited store for domain specific information within our general memory system. In particular, vWM provides a temporary store for a relatively small amount of phonological information while further linguistic input is processed. Given the nature of language processing which requires resolution of various dependencies either between adjacent words or spanning multiple words (e.g., referent resolution, thematic role assignment), and hence requires a small

amount of information to be temporarily stored while further input is processed, it should not be surprising that vWM has been shown to correlate with subjects' performance on a wide variety of offline linguistic tasks that indirectly tap into the degree of successful sentence processing and interpretation (e.g., acceptability judgments, verification task; Boyle, Lindell & Kidd 2013; Conway et al., 2005; Daneman & Carpenter, 1980; DeDe, Caplan, Kemtes & Waters, 2004; Jaeggi, Buschkuehl, Perrig & Meier, 2010; Kim & Christianson, 2013; Roberts & Gibson, 2002; Sprouse, Wagers, & Phillips, 2012; Swets, Desmet, Hambrick & Ferreira, 2007). Interestingly, it has not been found to correlate with online language processing (Caplan, DeDe, Waters, Michaud & Tripodis, 2011). In a series of experiments, Caplan and colleagues (2011) collected online (self-paced reading) and offline (plausibility judgments, verification task) processing measures, together with vWM measures (alphabet span, subtract 2 span, and sentence span), to investigate the role of working memory on syntactic processing, specifically, relative clauses (subject-clefts, object-clefts, subject relative and object relative clauses). The results of their experiments showed that vWM correlated with subjects' performance on the offline tasks: participants with a lower vWM span were less accurate in both the acceptability judgment (i.e., their judgment with respect to the acceptability of the sentence was less accurate) and the sentence verification tasks, particularly in more complex sentences (i.e., object-clefts and object relative clauses), as these sentences require participants to apply the same demanding operations (e.g., encoding of thematic roles in memory for later comparison to the probe, assessment of thematic roles for plausibility) multiple times. However, vWM did not correlate with reading times. These data thus seem to indicate that vWM correlates with offline but not online linguistic processing.

The above discrepancy between offline and online data bears some semblance to what has been observed in terms of passive difficulty: it is observed offline, but not online. One might expect a correlation between vWM scores and offline performance on passive sentence comprehension for two reasons. Firstly, the *Good Enough* model of passive sentence processing might predict subjects with a lower vWM span to rely more on heuristics, given that algorithmic processes pose greater demands on the parser. These participants should hence make more errors than participants with a greater vWM in an offline verification task. Secondly, subjects with a lower vWM span might

have more difficulties in maintaining a passive representation in memory in order to perform an offline task that targets post-interpretative processes, such as the verification task. Such an offline task may be more demanding on memory mechanisms in the case of noncanonical structures, for example through requiring further manipulations in memory. This would be in line with Caplan and Waters' distinction between parsing and interpretation and post-interpretive processes, where parsing processes, e.g., online reading times, are not dependent on vWM but post-interpretative processes are (Caplan & Waters, 1999). In both cases, we would expect a correlation between vWM span and accuracy to comprehension questions, i.e., the lower the vWM span, the larger the difference between actives and passives, with passives being more errorful than actives.

If indeed such a correlation were to be observed, then we could hypothesize that processing a passive sentence is not difficult *per se* (as suggested by the faster reading times in online tasks). Rather, operating on/maintaining the full representation of these sentences might be more difficult than for active sentences.

1.6 Interaction between passivization and predicate semantics.

1.6.1 The interaction between passivization and predicate semantics in the theoretical literature.

Thus far, I have discussed theoretical frameworks that could account for the pattern of results obtained from healthy adults on processing passive sentences, generally. However, the healthy adult literature has overlooked the fact that there is reason to believe that passivization complexity depends on the predicate semantics[1]. The theoretical literature on passivization describes an interesting interaction between passivization and predicate type across several languages (e.g., French, Italian, Hebrew, Finnish, Dutch, German; Belletti & Rizzi, 1988; Gehrke & Grillo, 2007, 2009; Grillo, 2008).

[1] At the outset of this study, in 2014, no healthy adult study had taken the interaction between passivization and predicate type into consideration. During completion of the present dissertation, a study was published by Ambridge and colleagues (Ambridge, Bidgood, Pine, Rowland & Freudenthal, 2016), investigating this interaction.

In particular, events and states exhibit different properties/restrictions with respect to passivization.

Eventive predicates, or events, are activities, accomplishments or achievements (e.g., “to push”); they describe processes going on in time, which consist of phases developing one after the other (Vendler, 1957). Eventive predicates can be identified through a series of linguistic tests (Vendler, 1957): (1) they should accept the progressive form (see (17) for an example); (2) they can represent a possible answer to the question “what are you doing?” (see (18) for an example); (3) it should be possible to ask “at what time were you X-ing?” and “for how long did you X?”, where X stands for the verb we wish to test (see (19) for an example).

(17) John is pushing the cart.

(18) What is John doing? John is pushing the cart.

(19) For how long did John push the cart?

Stative predicates, or states, represent properties that can be predicated of a subject for a certain period of time, with truth or falsity (e.g., “to love”, “to see”; Vendler, 1957). As eventives, states can be identified by employing the following tests (Vendler, 1957): (1) they cannot be modified by volitional adverbs, i.e., adverbs that express intentions, such as “carefully” or “deliberately” (see (20) for an example); (2) it should be possible to ask “how long did you X?”, where X stands for the verb we wish to test (see (21) for an example); (3) the sentence “I X”, where X stands for the verb we wish to test, can be replaced by “I can X” (see (22) for an example).

(20) *John carefully loved Mary.

(21) How long did John love Mary?

(22) I see = I can see.

Events and states behave differently with respect to passivization: a language that presents passivized states always allows passivized events, while the contrary is not true (Keenan & Dryer, 2007).

Moreover, while eventive predicates appear to consistently deliver verbal passives, stative predicates are temporarily ambiguous between an adjectival and a verbal interpretation of the passive structure (see (23), whose interpretation is

ambiguous between verbal and adjectival interpretation). In languages in which the two interpretations are homophonous (such as English), they can only be disambiguated once a by-phrase is encountered (compare (23) with (24)).

Other languages (e.g., German, Italian; Belletti & Rizzi, 1988) distinguish the two interpretations via the use of dedicated auxiliaries (e.g., in Italian *venire* – to come – cannot be used with adjectival passives, as in (25), while *essere* – to be – can be used with both types of passives, adjectival and verbal, as in (26)). These languages present restrictions on the type of states that can passivize (e.g., object experiencers cannot be passivized in Italian, see (27); Belletti & Rizzi, 1988), as well as on the type of by-phrases that can follows passivized states (e.g., Italian only allows generic by-phrases, see (28) vs. (29); Belletti & Rizzi, 1988).

Furthermore, when a stative predicate requires a verbal interpretation, the meaning of the verb must be coerced into an eventive interpretation, an additional operation with respect to passivizing an eventive predicate. Specifically, in order to be passivized, the basic meaning of stative verbs needs to be reconceptualised into the consequent state of an event, i.e., a state that came into existence and that is predicated over the internal argument. This operation is considered to be a type shift from states to achievements via the addition of a BECOME operator to the semantic structure of the verb (see McIntyre, 2006, for a detailed description of this operator). This addition allows the state described by the verb to be transformed into an "event" type of predicate, thus permitting passivization (Gehrke & Grillo, 2007, 2009; Snyder & Hyams, 2015). If we take sentence (30), the passivized stative verb "own" acquires an additional temporal qualification otherwise not present in the active version of the verb, i.e., the starting point of the ownership state which took place in the past. Support for this additional semantic operation comes from the fact that only coercible statives can be passivized, while predicates that cannot be transformed into a consequent state of an event (e.g., "escape") cannot (see (31)).

(23) Mary was worried.

(24) Mary was worried by John.

(25) *Gianni viene amato.
John comes loved.

(26) Gianni è stato amato. / Gianni è stato colpito.

John was loved. / John was hit.

(27) *Gianni viene preoccupato da Maria.

Gianni comes worried by Maria.

(28) ?Gianni viene apprezzato da me.

Gianni comes appreciated by me.

(29) Gianni viene apprezzato da tutti.

Gianni comes appreciated by everyone.

(30) The house is owned by John. = John has gained the ownership of that house. (from Gehrke & Grillo, 2009)

(31) *John got into a state of escape.

*John was escaped by the solution.

Hence, English (as well as other languages) presents an interaction between passivization and predicate semantics in terms of availability and restrictions on passivization for states, which do not apply to events. Eventive predicates consistently form verbal passives, whereas stative predicates are preferentially interpreted as adjectival, and this preferential interpretation needs to be revised into a verbal one once a by-phrase is encountered, thus creating a potential revision cost. Furthermore, those states that can form verbal passives require coercion of the verb meaning into a consequent state of an event, thus further increasing the complexity of passivizing statives with respect to eventives.

We might hence expect this difference in passivization across events and states to determine a similar difference in processing. In particular, revision and coercion, required to passivize states, but not events, might determine an additional complexity effect in processing passivized states with respect to passivized eventives. Indeed, such a difference in processing was reported both in the acquisition and aphasia literature (Demuth, Moloi & Machobane, 2010; Grodzinsky, 1995; Manetti, 2013; Maratsos et al., 1985; Messenger, 2009; O'Brien, Grolla & Lillo-Martin, 2006; Volpato, Verin & Cardinaletti, 2015). During completion of the present dissertation, new evidence in support of the existence of such a difference was observed also in healthy adults' processing (Ambridge, Bidgood, Pine, Rowland & Freudenthal, 2016). These relevant studies will be discussed in turn below.

1.6.2 Evidence from the acquisition literature.

While previous work on the acquisition of passive sentences reported that passives are acquired later than actives (e.g., Bever, 1970; Maratsos & Abramovitch, 1975), thus implying an additional difficulty in processing passives across the board, recent studies that controlled for the properties of the tested predicates found a distinction between age of acquisition of passivized events vs. passivized states. It is important to note that these studies discuss an actional/agent-patient vs. non-actional/experiencer-theme predicate distinction, while I have discussed an events vs. states distinction. Actional predicates are defined as verbs involving an agent and a patient as subject and object respectively, while non-actional verbs are defined as having experiencers or stimuli as subjects and objects. In line with Gehrke and Grillo (2009), I consider passivization to be an operation on predicate event structure, rather than argument structure. In spite of any difference between these distinctions, there is overlap across these categories, i.e., most actional verbs are events, and most non-actional verbs are states. Hence, findings of a difference in passivization between actional and non-actional verbs should extend to events vs. states predicates.

Three-year old children were found to comprehend and produce passives of actional[2] predicates, short and long, as well as actives, across several languages (e.g., Demuth et al., 2010; Manetti, 2013; Messenger, 2009). On the other hand, a consistent difficulty in production and comprehension of non-actional passives was reported across studies, tasks and languages. In a truth-value judgment task (O'Brien et al., 2006), English-speaking children as young as 3-year old performed as accurately on long actional (e.g., "to chase") passives (91%), as on short ones (88%), while their performance on long non-actional (e.g., "to like") passives was significantly worse (82%). This difference seems to persist up until the age of 9 (Maratsos et al., 1985), as shown by the results of a "who did it task", where children aged 4- to 9-year old performed at chance on non-actional (e.g., "to like", "to love") passives, while they answered

[2] As discussed in the previous paragraph, most actional predicates are events, while most non-actional predicates are statives. Any finding on the processing of actional and non-actional predicates should hence extend to eventives and statives respectively.

accurately when actional (e.g., “to hold”, “to wash”) passives were tested. The same results were replicated in different languages (e.g., Italian; Volpato et al., 2015).

An important exception in this panorama is a recent study published by Messenger and colleagues (Messenger, Branigan, McLean & Sorace, 2012), in which children’s performance in a priming task targeting passive and active sentences was not modulated by the predicates’ characteristics, i.e., the amount of passive sentences produced by children after hearing a passive prime was not affected by verb type, whether the prime contained an agent-patient verb (e.g., “to push”) or an experiencer-theme one[3] (e.g., “to love”). However, data in favour of an interaction were reported within the same study: in a subsequent picture-sentence matching task, children’s accuracy was lower for passives of experiencer-theme verbs than agent-patient ones. Overall, these data suggest that the emergence of an interaction between passivization and predicate semantics might be modulated by tasks characteristics. Messenger and colleagues argue that the interaction only emerges as an artefact of the picture-sentence matching task which is inherently more difficult to perform for the non-actional predicates, as these verbs are difficult to represent and recognize in pictures. While this is a possible interpretation of the data, it could also be the case that the picture-sentence matching task is simply more sensitive to differences in interpretation, as it taps into interpretative processes, while the priming task might simply reflect whether children have access to a certain structure in more general terms.

1.6.3 Evidence from the aphasia literature.

Similarly to the case in acquisition, a general difficulty in processing passive sentences was reported in the aphasia literature (e.g., Caramazza & Zurif, 1976; Grodzinsky, 1990, 2000; Saffran, Schwartz & Linebarger, 1998). However, recent online data, collected using eye-tracking (Dickey & Thomspon, 2009) and self-paced listening (Caplan & Waters, 2003), suggest that aphasic patients process passive sentences in the same time-frame as healthy subjects, and following somewhat similar stages. On the other hand, their gaze patterns on later regions (as well as their performance on offline tasks,

[3] As discussed above, most agent-patient (or actional) predicates are events, while most experiencer-theme (or non-actional) predicates are statives. Any finding on the processing of agent-patient and experiencer-theme predicates should hence extend to eventives and statives respectively.

as described in earlier studies) indicate that passives processing is affected at later stages, when competition from alternative parses is stronger (Avrutin, 2006; Den Ouden, 2006). The inability to suppress these alternative irrelevant interpretations (e.g., heuristic parse of the sentence) would thus lie at the heart of aphasic patients' selective deficit with passive sentences.

Interestingly, a variation in processing difficulties across predicate types was reported in aphasic patients, parallel to the acquisition findings described above. In an anagram task, which aimed at testing the comprehension of active and passive sentences in aphasic patients across predicate types (agent-patient verbs, e.g. "to stop", vs. experiencer-theme verbs, e.g., "to admire"), Grodzinsky (1995) observed that subjects' performance was above chance across all active verbs, independent of predicate type (percentage of errors on agent-patient verbs: 14.28%; percentage of errors on experiencer-theme verbs: 19.63%). While patients' performance on passive sentences was generally worse than on active sentences, in line with the previous literature, it also varied depending on the predicate tested: participants accuracy was at chance with agent-patient verbs, while below chance with experiencer-theme (percentage of errors on agent-patient verbs: 41.97%; percentage of errors on experiencer-theme verbs: 67.83%).

Despite drawing on the same distinction used in the acquisition literature, i.e., a distinction based on the argument rather than event structure of predicates, these results once again confirm the existence of an interaction between passivization and predicate type.

1.6.4 Evidence from the healthy adults' literature.

The results reported in the aphasia literature, together with the acquisition data and theoretical data discussed above, provide support to the idea that passivization difficulty is modulated by extra-syntactic factors, such as predicate semantics. These data should lead to expect the same interaction between passivization and predicate type, i.e., increased difficulty in passivized states, in healthy adults' processing of passives.

Indeed, a recent study found evidence of such an interaction in offline processing of passives in healthy adults (Ambridge et al., 2016). In a series of offline tasks (grammaticality judgments and forced-choice picture-matching comprehension task),

Ambridge and colleagues manipulated both the syntax of the tested sentences (active vs. passive), together with the predicate type of the verbs used (predicates requiring agent-patient vs. theme-experiencer vs. experiencer-theme as subject and object respectively). In the grammaticality judgment task, subjects found passives less acceptable than actives overall, and passives of theme-experiencer and experiencer-theme predicates less acceptable than passives of agent-patient predicates. Similarly, in the forced-choice picture-matching comprehension task, participants responded faster to agent-patient predicates than theme-experiencer and experiencer-theme predicates across syntactic structures, and the difference was larger within passive sentences with respect to active ones. It is important to note that only accurate trials were analysed, thus excluding a possible speed-accuracy trade-off in explaining these results.

The predicate type distinction employed by Ambridge and colleagues is based on a different semantic property of predicates with respect to the event vs. state one, i.e., semantic affectedness, which denotes a "happening or situation" delimited to the internal argument of the verb (Tenny, 1987:75). Pinker and colleagues (Pinker, Lebeaux & Frost, 1987; Pinker, 1989) propose that passivization interacts with this semantic property: specifically, predicates whose internal argument is affected by the action described by the verb (e.g., "push") can be easily passivized, while those that are less consistent with this property (e.g., "admire") might resist passivization to a lesser or greater extent. Although employing a different distinction (agent-patient verbs vs. theme-experiencer/experiencer-theme verbs, based on semantic affectedness) from the eventive vs. stative one described in the theoretical literature, these results suggest that the difficulty in healthy adults' offline comprehension of passive sentences with respect to active ones is at least partially modulated by the semantics of the predicates involved. Furthermore, the difference across predicates is observed in the expected direction: passives of statives, especially subject-experiencer predicates, are more difficult to process than passives of eventives. These results complement nicely with previous data reported in the acquisition, aphasia and theoretical literature described above.

Future studies on healthy adults' processing of passive sentences should hence include a manipulation of predicate type in order to replicate these offline findings, thus strengthening the support to an interaction between passivization and predicate type.

Furthermore, it would also be interesting to investigate possible online correlates of this interaction, by employing different methodologies, such as self-paced reading or eye-tracking.

1.7 The present study.

Although passives are generally considered more complex than actives, and the passive-active contrast has often been used as a metric of complexity to assess processing abilities across populations, the existing literature does not present a homogeneous picture: while offline tasks (requiring a sentence judgment post sentence processing) elicited worse performance on passives than actives (Ferreira, 2003), online data (moment-to-moment processing of sentences) indicate that passives are processed at the same speed, if not faster, than actives (Carrithers, 1989; Traxler et al., 2014). Hence, while the offline data are consistent with the general tenet that passives are more complex than actives, the online data undermine this general assumption.

Given the widespread use of the passive-active contrast as an assessment tool across various populations, it is thus fundamental to establish how healthy adults process this structure. Moreover, the results would allow to assess the predictions made by the models of passives processing described in a previous section and to determine which model most adequately accounts for the data. The existing literature suggests two possible scenarios: (1) passives are more complex than actives, but previous online studies could not detect a complexity effect due to shortcomings in their designs (e.g., no simultaneous collection of offline data, short by-phrases); (2) passives are not more complex than actives, but rather inter-individual differences (e.g., vWM span) affect performance on offline tasks.

The present dissertation aims at investigating the processes engaged when healthy adults parse and interpret passive sentences, while improving on possible shortcomings of previous studies by: (1) simultaneously collecting both online and offline measures of processing; (2) employing long by-phrases to allow sufficient time to detect an online complexity effect; (3) controlling for predicate semantics; (4) collecting measures of vWM to investigate a possible correlation between vWM span and subjects' processing of passive sentences (both online and offline).

Chapter 2 will hence present the first set of 4 experiments, which collected both online (self-paced reading) and offline (accuracy to comprehension questions) measures and focused on the interaction between passivization and predicate structure. Additionally, Experiment 4 investigated the role of vWM in participants' offline performance in passives comprehension.

Results collected in Experiments 1-4 replicated the same contrast between online and offline performance reported in the previous literature: participants were faster at reading passive sentences, but less accurate in answering comprehension questions about them. Moreover, no evidence for an interaction between passivization and predicate properties was found, either offline or online. Crucially, vWM span correlated with the difference in accuracy between passive and active sentences, thus indicating that differences in working memory could partially be responsible for the worse offline performance on passives than actives. Moreover, vWM span did not correlate with subjects' performance on the online task, indicating that performance on the online vs. offline task stems from (at least partially) independent processes. I thus decided to separate the investigation of online vs. offline processing into two different investigations in Chapter 3 and 4, respectively.

Chapter 3 presents Experiment 5-7, which paired two different offline tasks, acceptability judgment and comprehension questions. The goal of these experiments was to investigate a further potential extra-syntactic explanation for subjects' worse performance in the offline verification task, i.e., difficulty in memory encoding or retrieval due to task characteristics. Once these task characteristics were controlled for, the overall main effect of passivization on difficulty in comprehension disappeared, while an interaction between passivization and predicate semantics was observed, in the expected direction. Passivized states were more complex to process (more errorful) than their active counterparts, while no difference was found across eventives.

Chapter 4 presents Experiment 8, which was motivated by the offline interaction observed through Experiments 5-7. Previously, no effect of an interaction was observed in the online self-paced reading studies (Experiments 1-4). Hence, I decided to turn to a different methodology, i.e., eye-tracking, which allows to collect multiple measures of processing and, as I will argue in Chapter 4, is better suited to detect an interaction effect. Passivization and predicate properties were found to interact online in late

measures of processing in the expected direction in this last experiment. In particular, revision and/or coercion intervened at the verb and at the head of the by-phrase in passivized statives creating a slow-down with respect to their active counterparts, while no difference was found across voice in eventive predicates, thus nicely reconciling the online and offline findings with respect to the predicted interaction between passivization and predicate type.

Finally, Chapter 5 will discuss possible implications of the results reported in the previous chapters for models of passives processing and future experiments to further test these interpretations.

2 A simultaneous investigation of passives processing in online (self-paced reading) and offline (verification task) tasks

2.1 Introduction.

As discussed in Chapter 1 there is a contrast between online and offline processing of passive sentences: facilitation for passives online but difficulty offline. Here I consider potential explanations for this contrast before discussing experiments designed to investigate them.

Previous offline studies (Ferreira, 2003) found passives to be more errorful than actives, but did not simultaneously collect online data. Likewise, previous studies that collected online measures of passive sentence processing and found passives to be read faster than actives (Carrithers, 1989; Traxler et al., 2014), did not collect offline data, e.g., accuracy to comprehension questions. This limits any conclusions that can be drawn from this contrast. For example, it is possible that subjects tested in these previous online studies depended more on a shallow processing (e.g., word-order heuristic) of passive sentences, thus leading to faster reading times in passives and also (although not assessed) to an incorrect understanding of the sentences on a majority of trials. In order to better establish the reliability of this contrast, the simultaneous collection of online and offline data is required.

Another potential concern regarding the online reading time data stems from previous cross-modal lexical priming studies (Osterhout & Swinney, 1993). These studies found that in passive sentences the filler might be reactivated quite late, up to 1000ms after the verb (or gap), i.e., up to 3/4 words after the verb. This suggests that detecting an online complexity effect related to dependency resolution in passives might depend on the presence of a long enough by-phrase. Crucially, previous online studies (Carrithers, 1989; Traxler et al., 2014) employed short by-phrases, which only contained 3 words (e.g., “by the cowboy”).

Given these considerations, Experiment 1-4 were designed to simultaneously collect online (self-paced reading) and offline (verification task) data for passivization, while also manipulating the predicate type contained in the stimuli to test for a possible interaction between passivization and predicate semantics (see Chapter 1 for more

details). The results across experiments replicated the separate findings of previous studies: passives were found to be faster than actives online, while being more errorful offline. Exp. 4 additionally collected measures of vWM to investigate whether variability in vWM across subjects might contribute to the increased number of errors on passives with respect to actives, suggesting that passives are not more difficult than actives across the board, but rather encoding/maintaining the representation of passive sentences, as required by the verification task, might be.

Exp. 1-4 will be discussed in turn, followed by a general discussion section that will recap the findings across experiments and discuss their interpretation as well as the next steps in the study.

2.2 Experiment 1.

Experiment 1 was designed to simultaneously collect online (self-paced reading) and offline (verification task) data for passivization with stimuli containing long by-phrases. The items were designed to only contain eventive verbs. The rationale was that eventive predicates consistently deliver verbal passives, thus providing the best testing ground for a complexity effect due to movement, as hypothesized by the syntactic complexity approach. Finally, a further control condition, beside the simple active one, containing a progressive verb, was added. Previous studies (Carrithers, 1989; Traxler et al., 2014) only presented an auxiliary in the passive condition; this might have cued participants to specific grammatical categories contained in the upcoming passive structure, thus facilitating processing of passive sentences. Hence, previously observed faster reading times might simply represent an artefact of this confound. The progressive condition included an identical auxiliary to that in the passive (i.e., “was”), thus eliminating the auxiliary as a cue to the passive structure, cue that was possibly present in previous designs.

If passives are found to be harder to process than actives in both online (slower to read) and offline (more errorful) data, than theories of passives complexity (the *Good Enough* model, the syntactic complexity approach and the frequentist approach) would adequately account for passives processing.

On the other hand, if passives are found to be read faster and understood more accurately than actives, then it would be necessary to consider alternative models of

passive sentence processing, e.g., expectation-based or surprisal-based models of sentence processing (see Chapter 1 for a discussion of these models with respect to passive sentences processing), as theories of passives complexity would not adequately explain such findings.

Finally, a third scenario consists in finding passives to be processed faster than actives, while being more errorful than actives offline. This would replicate the separate results of previous studies (Carrithers, 1989; Ferreira, 2003; Traxler et al., 2014) and would imply that performance on online and offline tasks on passive sentences stems from different processes, thus leading to different outcomes (facilitation online, complexity offline).

2.2.1 Methods and Design.

2.2.1.1 Participants.

Thirty-five British English native speakers were recruited to participate in the study (24 females; average age: 28.6). They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants were recruited through the UCL Sona System, the online subject pool operated by the UCL Division of Psychology and Language Sciences, and received either payment or course credits for their participation. All the participants were informed of the aims and procedures of the experiment and provided informed consent, approved by UCL ethics.

2.2.1.2 Stimuli.

There were three conditions (see Table 2 for example sentences): (1) passive, (2) active progressive, and (3) simple active. The progressive was included to act as an additional control, which is matched to the passive for the auxiliary “was”. Thirty sentence sets were generated. The sentences all contained eventive predicates. The by-phrase/direct object noun was modified (pre-nominally) by two conjoined adjectives in order to allow sufficient time to detect an online complexity effect in passives. The sentence final prepositions phrases were included to avoid any end-of-sentence effects (Just, Carpenter & Woolley, 1982).

The length (i.e., number of words per sentence) was kept constant within each condition: given the syntactic differences across conditions, passive sentences always had 2 more words than simple actives (i.e., auxiliary and "by"), and progressive sentences always had one more word than simple actives (i.e., auxiliary; overall: 19 words for all 30 passives, 18 words for all 30 progressives, 17 words for all 30 actives; see Table 2 for example sentences; for the complete list of items, see Appendix 7.1).

Additionally, 60 filler sentences, with varying complexity (15 actives; 15 passives; 15 sentences with negation; 15 garden-path "while…" constructions), were created to mask the purposes of the experiment (for the list of fillers, see Appendix 7.2).

Passive	The guitarist was rejected by the attractive and talented singer in the concert hall next to the Irish pub.
Progressive	The guitarist was rejecting the attractive and talented singer in the concert hall next to the Irish pub.
Simple active	The guitarist rejected the attractive and talented singer in the concert hall next to the Irish pub.

Table 2: Example of experimental items in Exp. 1.

2.2.1.2.1 Pre-Norming: Plausibility.

The order of the NPs was not reversed across active and passive items. This means that the NPs in passive sentences were assigned the reversed thematic roles with respect to active sentences. To ensure that plausibility was not affected by reversing the theta-role-NP combinations across conditions, the experimental items were normed in 2 plausibility tasks administered via online questionnaires designed in Qualtrics (https://www.qualtrics.com/). 72 participants (recruited online through the UCL Subjects Pool https://uclpsychology.sona-systems.com/Default.aspx?ReturnUrl=/; 41 females; mean age: 26.84) rated, on a 7-points scale from highly implausible (e.g., 1) to highly plausible (e.g., 7), the plausibility of the experimental items and of implausible items created ad hoc.

The first questionnaire targeted the plausibility of thematic role reversals, i.e., whether reversing the order of the arguments affected the plausibility of the sentence. Both orders were tested in the active form without the PPs. Experimental items were thus NP-VP-NP active sentences, and the manipulation consisted in reversing the NPs.

The second questionnaire targeted the plausibility of the entire sentence across the progressive and passive conditions. Implausible items had the same structure as the experimental sentences, and changed accordingly from the first to the second questionnaire. The implausible filler items resulted from argument role reversals (see (32)), or general semantic/pragmatic oddness (see (33); for the complete list of implausible items, see Appendix 7.3).

(32) The law abiding police man was arrested by the criminal last Saturday on the town's busy high street.

(33) Santa Claus gave coal to all the children on the nice list and presents to those on the naughty list.

Data collected in the first questionnaire were analysed using a linear mixed effects model, containing the order of theta-roles as fixed effect and both subjects and items as random effects (including both intercepts and slopes). The contrasts used were passive order (e.g., "The attractive and talented singer rejected the guitarist") vs. active order (e.g., "The guitarist rejected the attractive and talented singer"; [0.5, -0.5]). P-values were determined through treating the t-value as a z-statistic (Barr, Levy, Scheepers, & Tily, 2013). The analysis revealed that there was no significant difference between the two possible orders of arguments (average active: 4.84; average passive: 5.04; β=-.2, t=-.16, p= .25).

Data collected in the second questionnaire were analysed similarly to data collected in the first questionnaire, but the fixed effect was syntax and the contrast used was passive vs. progressive [0.5, -0.5]. No significant difference was found between the two conditions (average progressive: 4.83; average passive: 4.94; β=-.19, t=-.88, p= .38).

Hence, the items did not significantly differ either in the plausibility of thematic role order across conditions, or in the overall plausibility of the sentences across the progressive and passive condition.

2.2.1.2.2 BNC Corpus Analysis: Structural Frequency of the Verbs.

An analysis of all the verbs' entries contained within the British National Corpus was conducted (BNC; http://www.natcorp.ox.ac.uk/) to exclude the possibility that the predicates we selected could be generally biased towards the passive structure in the

language use. Using the "Phrases in English" tool provided by the BNC, the first 100 instances of each verb in its verbal past-tense form (i.e., as used in the experiment, not as an adjective) were selected. The verbs were analysed in their original sentential context in order to reliably categorize each instance as being in the active or passive voice. An analysis on the frequency of their surface form found the verbs to be more frequent in the active form (on average 64.15% of 100 occurrences appeared in the active form). The results are summarized in Table 3.

Verb	Freq. passive form (%)	Verb	Freq. passive form (%)	Verb	Freq. passive form (%)
abandon	35	hire	25	rescue	32
attack	27	injure	77	save	24
bribe	20	kidnap	48	seduce	32
corrupt	38	mug	62	shock	57
fire	57	pay	19	slap	14
harass	41	push	14	stab	48
		reject	27	sue	20

Table 3: Frequency of surface forms of the predicates employed in Exp. 1.

2.2.1.3 Procedure.

The normed items (90 experimental sentences) were included with the fillers to be used in self-paced reading using Linger 2.88 software *(*http://tedlab.mit.edu/~dr/Linger/readme.html*)*. The chosen regions of interests were: (1) auxiliary, (2) verb, (3) determiner of the by-phrase, (4) first adjective, (5) conjunction, (6) second adjective, (7) object NP, (8) first word after the object NP, (9) second word after the object NP.

Verification questions requiring a "yes" or "no" button response followed each item, both experimental and filler, to ensure participants' active comprehension during the task (for the list of comprehension questions, see Appendix 7.4). Questions were designed to be half correct and half incorrect. Questions could either be simple (e.g., focusing on details within the sentence, for example, a simple incorrect question to the sentences presented in Table 2 could be "Did the musician play the piano?") or complex

(e.g., targeting theta-role assignment). The rationale was to try to target all aspects of the sentence, so as to avoid strategic processing.

Sentences were presented with the words masked by dashes. An empty space appeared between words. A press of the space bar unmasked one word at a time. On the contrary, comprehension questions were presented unmasked on the screen and participants pressed either “j” for “no”, or “f” for “yes”. A gaming keyboard (Razer® Blackwindow) was used for accurate button response timing. Participants were provided with feedback if they chose the wrong answer. Practice trials (6 in total) allowed them to familiarize with the task prior to the testing session.

In total, 90 experimental stimuli (30 experimental items and 60 fillers) were presented to each participant. The task lasted approximately 30 minutes and was administered in a soundproof room.

2.2.1.4 Data analysis.

Due to low accuracy on fillers (lower than 75% overall), 5 participants were excluded from analysis. Hence, data from 30 participants were analysed.

The outcome measures were: accuracy and reaction times to verification questions and reading times.

2.2.1.4.1 Reading times and Question Response times data:

The analysis was run using RStudio, an application for data analysis (https://www.rstudio.com/). Unreasonably high (>2500ms) and low (<100ms) raw reading times were excluded. Residual logRTs were calculated based on word length and the restricted cubic spline of word position (Hofmeister, 2011) considering all sentences (experimental and filler). The residuals were then analysed to identify further possible outliers: data above or below 2.5 times the standard deviation from the mean (by subject, condition, and region) were excluded. This resulted in 0.24% of the original data being removed. In terms of verification response times, unreasonably high (>12000ms) and low (<700ms) values were removed. This resulted in 0.14% of the original response time data being removed.

The cleaned residual logRT reading and logRT reaction time data were analysed using a linear mixed effects model including syntax as a fixed effect and both subjects

and items as random effects (including both intercepts and slopes). The contrasts used were passive vs. actives (both progressive and simple active) [2/3, -1/3, -1/3], and progressive vs. simple active [0.5, -0.5]. P-values were determined through treating the t-value as a z-statistic (Barr et al., 2013).

Given that only accurate trials can be thought to reflect an accurate and complete processing of the sentences, the analysis of reading times in accurate trials will be reported and discussed in the Results section. Moreover, this is critical for the predictions of the *Good Enough* theory, as it is only on accurate trials where revision would need to take place. However, analyses of all trials (accurate and inaccurate) were also performed and no significant difference was found between the two analyses.

2.2.1.4.2 Comprehension Accuracy data:

Accuracy data were analysed using a mixed effects logistic regression with a binomial distribution including the same effects and contrasts as the analysis of the online data.

The full model (containing both intercepts and slopes for random effects) was always run first. The Results section will specify where convergence could not be met and how the model was modified to meet convergence.

2.2.2 Results.

2.2.2.1 Comprehension Question Results:

Offline measures did not reveal significant differences across conditions. Accuracy to comprehension questions was almost identical across conditions (passive: 85.71%; progressive: 83.67%; simple active: 83.3%; passive vs. actives: β=.04, z=.2, p=.84; simple active vs. progressive: β=.03, z=-.11, p=.9; see Table 4[4]), and, similarly, RTs did not significantly differ across conditions (passive vs. actives: β=.004, t=.27, p=.79; simple active vs. progressive: β=.008, t=.52, p=.6; see Table 4[5]).

[4] Model did not converge with random slopes. The same model was used for the analysis of theta-role questions only and for first- vs. second-half analysis, given that more complex models did not converge.

[5] Model only converged with random slope for subjects but not for items. The same model was used for the analysis of theta-role questions only and for first- vs. second-half analysis, given that more complex models did not converge.

Given that Ferreira (2003) argues that the comprehension of passive sentences is selectively impaired with questions targeting theta-role assignment, participants' performance on theta-roles questions only was separately analysed. The analysis did not reveal any significant difference with respect to the overall analysis: participants' performance was generally lower on theta role questions than on the other question types (passive: 78.5%; progressive: 81.12%; simple active: 79.4%), but accuracy did not significantly differ across conditions (passive vs. actives: β=-.11, z=-.45, p=.66; simple active vs. progressive: β=-.05, z=-.15, p=.88). Similarly, RTs to theta-role comprehension questions did not differ from RTs to other question types and there was no significant difference across conditions (passive vs. actives: β=.01, t=.57, p=.57; simple active vs. progressive: β=.02, t=.77, p=.44).

In order to exclude possible learning effects that might have confounded our results, 2 separate analyses on the data collected in the first vs. second half of the sessions[6] were run and compared. There was no significant difference between the results of the two (first vs. second) analyses.

The active sentences demonstrate a lower accuracy with respect to results reported by previous studies (Ferreira, 2003). This is not surprising given that our sentences were significantly longer (long PPs and prenominal modification) and our comprehension questions assessed all aspects of the sentence. In combination, they provided a more demanding task for the participants, which is bound to reduce accuracy.

Overall, the offline data provide no evidence for passive sentences being more difficult to understand.

	Passive	Progressive	Simple active
Accuracy	85.71% (± 6.39)	83.67% (± 6.76)	83.3% (± 6.82)
Reaction Times	3.39 (± .04)	3.37 (± .03)	3.38 (± .04)

Table 4: Accuracy and Reaction Times to comprehension questions in Exp. 1.

[6] Data from trials 7-51 (the first 6 items were practice items) were analysed as "first-half" and data from trials 51-96 as "second-half".

2.2.2.2 Reading Time Results:

Analysis of reading times on correct trials revealed that the verb was read significantly faster in passive sentences, with respect to active sentences (β=-.03, t=-2.42, p=.01) and in progressive sentences with respect to simple active ones (β=.05, t=2.54, p=.01). This outcome is likely related to the presence of the auxiliary in both the passive and the progressive conditions, which eases processing at the following verb due to a smaller *surprisal* effect, i.e., a smaller cognitive load in processing the subsequent word, given its high probability to follow in the sentence (Hale, 2001). In fact, the auxiliary both creates a strong expectation for a verb to follow and determines tense (past participle morphology). In the active, tense must be computed at the verb, thus additionally slowing processing with respect to the other two conditions.

Up to 4 regions after the verb, passive sentences were read numerically faster than actives and significantly so at 3 of them (at the determiner: β=-.04, t=-2.75, p=.006; at the conjunction: β=-.04, t=-3.43, p=<.001; at the second adjective: β=-.03, t=-2.07, p=.04). No significant difference was found at the first adjective or after the second adjective. Results are presented in Figure 2. Finally, the same first vs. second half analysis described for our offline data was conducted on the online data. Again, no significant difference was observed with respect to the overall results.

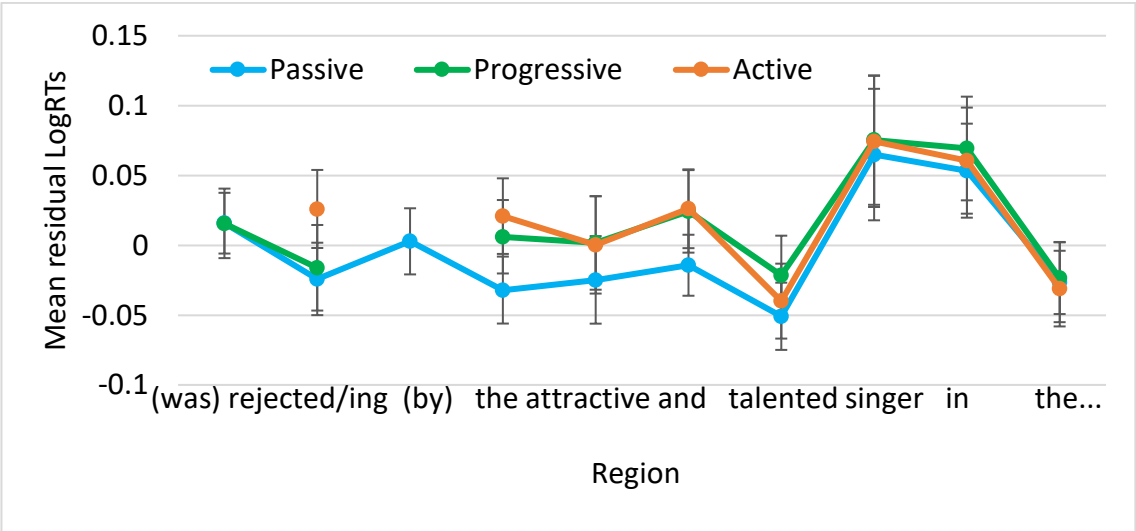

Figure 2: Self-paced reading results in Exp. 1. Region by region residual LogRTs according to verb syntax. Error bars indicate standard error of the mean.

2.2.3 Discussion.

The first experiment had two main results: (1) passive sentences were consistently read faster than actives until the noun of the by-phrase/direct object phrase, indicating that they are not more difficult to process; (2) passive sentences were comprehended as accurately as active ones, indicating that they are not harder to understand, and that the results were consistent both offline and online.

Overall, the combination of findings presented in point (1) and (2) are inconsistent with an agent-first heuristic processing of passive sentences, at least in terms of a commitment to the agent-first strategy that would later require revision through algorithmic processes (Christianson et al., 2001; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016). Given that accuracy was equivalent across actives and passives, the *Good Enough* account would predict passives to be processed more slowly than actives in order for the algorithmic processes to revise the incorrect interpretation on the majority of trials.

Moreover, the present results suggest that non-canonical structures are not more difficult than canonical ones across the board, contrary to what has been generally argued in the literature (see Chapter 1 for a review).

Finally, if indeed passivization is obtained via movement, the results suggest it does not necessarily imply an additional processing cost with respect to the active structure. This contrasts with what is generally assumed in the literature and with what is found in other structures involving movement resulting in a noncanonical argument order (e.g., object relatives are always more taxing than subject relatives; Caplan et al., 2007; Garnham & Oakhill, 1987; Gibson, 1998; Grodner & Gibson, 2005; Hakes, et al., 1976; Santi & Grodzinsky, 2010; Staub, 2010; Traxler, et al., 2002). These results seem to be in line with other studies which report that movement generating a noncanonical order does not produce difficulty on its own, but rather in combination with other factors affecting memory-access. For example, processing difficulty in object- vs. subject-extracted relative clauses was found to be modulated by the syntactic/semantic similarity between the NPs in the sentence that generates interference in memory (Gordon, Hendrick & Johnson, 2001).

The faster reading times for passives, described in point (1) could be explained by surprisal or expectation effects (Hale, 2001; Levy, 2007). As already mentioned in

Chapter 1, the English passive structure is visually signalled, at the surface level, by a higher number of morphological cues (e.g. auxiliary, past participle morphology, "by"), with respect to the active structure, which is not introduced by dedicated morphology in English. These cues create a higher expectation/lower surprisal for a verb to follow a subject NP and auxiliary in the passive structure, with respect to a simple subject NP in the active structure. In fact, the subject NP in the active condition could potentially be followed by other grammatical structures (e.g., adverb or preposition), thus lowering levels of expectation/increasing surprisal for a verb to follow with respect to the passive context. Moreover, a high expectation/low surprisal is set for an NP after the past participle and "by" have been processed. On the other hand, the same NP will have to compete with alternative possible continuations (e.g., preposition) after a verb has been processed in the active structure, thus lowering expectation/increasing surprisal with respect to the passive context.

Given that thematic role assignment in our stimuli was equally plausible in both orders (e.g., the order NP^X – VERB – NP^Y was equally plausible to the order NP^Y – VERB – NP^X) as established by our previous norming tests, the observation made in point (2) cannot be explained by a preference for theta role assignment in the passive order with respect to the active one.

Another possibility is that the predicates we selected are generally biased towards the passive structure in the language use. This alternative explanation is ruled out by the post-hoc analysis run on the frequency of use of our predicates in either passive or active voice. The wide range of frequencies and the on average lower frequency of our predicates appearing in the passive compared to the active form (on average 35.85% of 100 occurrences appeared in the passive form) is not compatible with frequency of surface form accounting for the faster reading times in the passive. A Pearson's product-moment correlation analysis performed in R (https://www.rstudio.com/), between the frequency of the predicates surface structure and the residual logRTs at the determiner (the region that most robustly showed faster reading times in passives), confirmed that frequency did not play a role in shaping our results (t=-0.65, p=.5). This result is in line with previous studies, which did not find a significant role of surface frequency in determining facilitation/difficulty in processing structures of varying complexity (Street & Dąbrowska, 2010).

The data summarised in the second point, i.e., the offline results, seem to be in line with the explanation proposed for the faster reading times observed in passive compared to active sentences. Moreover, the consistency between offline and online data in the results suggests an identical underlying explanation. The lack of ambiguity in the type of structure signalled by complex event predicates in our experimental sentences, i.e., a passive structure, could have indeed eased comprehension, thus resulting in equal accuracy across conditions.

Hence, passives processing does not seem to be adequately described or accounted for by passive sentence complexity theories (the *Good Enough* model, the syntactic complexity approach and the frequentist approach). A different explanation must hence be found in order to account for the offline difficulty effect of passive sentence comprehension, reported in previous studies (Ferreira, 2003).

As specified in the methods and design section, this experiment only contained event predicates, which consistently deliver verbal passives cross-linguistically. On the contrary, states often generate ambiguous adjectival/verbal interpretations of the passive structure. Moreover, verbal passivization of stative predicates has been argued to require coercion to an eventive interpretation (Gehrke & Grillo, 2009; Snyder & Hyams, 2015). Hence, passivized states might increase comprehension difficulty in passives with respect to actives. Indeed, the use of mixed predicate types in previous studies (e.g., Ferreira, 2003) might have partially contributed to the comprehension difficulty in passive sentences with respect to the control active condition.

A second experiment was thus designed, similar to the first one but testing only stative predicates. The coercion required by stative passives and their temporary ambiguity may introduce a complexity in processing (e.g., slower reading times and more errors in comprehension), thus explaining previous findings of a difficulty effect with passives.

2.3 Experiment 2.

The second experiment aimed at further exploring the relationship between event structure and passivization (see Chapter 1 for more details). In fact, while complex events consistently deliver verbal passives across languages, states are ambiguous between an adjectival and a verbal interpretation and the predicate's meaning must be

coerced into an eventive reading for the verb to passivize. Hence, from a processing point of view, passives of statives could potentially create more complexity than passives of eventives. The design of this experiment thus resulted from a single but fundamental modification of Exp. 1: the predicates employed were always stative, and particularly subject-experiencer psych predicates.

If passivization interacts with the predicate event structure in determining processing complexity, and passives of statives are more difficult to produce than passives of eventives for the reasons outlined in Chapter 1, passive sentences in Exp. 2 should be more difficult to process than their active counterparts. In particular, passive sentences should be processed more slowly (i.e., slower reading times with respect to actives) and accuracy in comprehension should be lower compared to actives.

2.3.1 Methods and Design.

2.3.1.1 Participants.

Twenty-six British English native speakers were recruited to participate in the study (21 females; average age: 23.4). They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants were recruited through the UCL Sona System and received either payment or credits for their participation. All the participants were informed of the aims and procedures of the experiment and provided informed consent, approved by UCL ethics. None of the participants had previously participated in any of the experiments related to this study (including the pre-norming tests).

2.3.1.2 Stimuli.

Three modifications to the experimental manipulations of Exp. 1 had to take place: (1) complex event predicates were substituted with subject experiencer predicates; (2) the locative PPs used in the previous experiment are not acceptable continuations with subject experiencer predicates, and were substituted with implicit causal clauses; (3) moreover, subject experiencer predicates do not generally allow the progressive form, hence only two experimental conditions were included: passive and active. To avoid the confound of an auxiliary bias, (i.e., only encounter a passive following the auxiliary

"was"), 10 of the filler sentences of Exp. 1 were modified to include "was" with the verb in the progressive form. Finally, some of the argument pairs were slightly modified to adjust to the new verbs and create plausible experimental items. Everything else (types of filler sentences; overall sentence length) was kept identical to Exp. 1. 30 sentence sets were generated as in Exp. 1. Examples of experimental items are presented in Table 5 (for the complete list of items, see Appendix 7.5).

Passive	The guitarist was admired by the attractive and talented singer for keeping the band focused through the whole tour.
Active	The guitarist admired the attractive and talented singer for keeping the band focused through the whole tour.

Table 5: Example of experimental items in Exp. 2.

2.3.1.2.1 Pre-Norming: Plausibility.

The experimental items were then normed in a plausibility study administered via an online questionnaire designed in Qualtrics (https://www.qualtrics.com/). Participants for this study were recruited via Prolific Academic, a platform for online research (https://prolific.ac/). Filler implausible items, rating scale, structure of experimental items (only active sentences; manipulation: order of arguments) were identical to the first plausibility questionnaire run in Exp. 1. 66 participants (37 females; average age: 31.5) were recruited to participate in the online questionnaire.

Data were analysed using a linear mixed effects model, containing the order of theta-roles as fixed effect and both subjects and items as random effects (including both intercepts and slopes). The contrasts used were passive order vs. active order [0.5, -0.5]. P-values were determined through treating the t-value as a z-statistic (Barr et al., 2013).

The results of the plausibility task revealed that there was no significant difference between the two possible orders of arguments (average active: 5.89; average passive: 5.76; β=-.02, t=.71, p=.48). Hence, the items did not significantly differ in the plausibility of thematic role order.

2.3.1.2.2 BNC Corpus Analysis: Structural Frequency of the Verbs.

An analysis of all the verbs' entries contained within the British National Corpus was conducted (BNC; http://www.natcorp.ox.ac.uk/) to exclude the possibility that the

predicates we selected could be generally biased towards the passive structure in the language use. Using the "Phrases in English" tool provided by the BNC, the first 100 instances of each verb in its verbal past-tense form (i.e., as used in the experiment, not as an adjective) were selected. The verbs were analysed in their original sentential context in order to reliably categorize each instance as being in the active or passive voice. An analysis on the frequency of their surface form found the verbs to be more frequent in the active form (on average 69.54% of 100 occurrences appeared in the active form). The results are summarized in Table 6.

Verb	Freq. passive form (%)	Verb	Freq. passive form (%)	Verb	Freq. passive form (%)
admire	25	envy	63	pity	26
adore	11	fear	14	resent	19
cherish	45	hate	5	respect	63
despise	35	loath	13	trust	49
detest	14	love	6	value	69

Table 6: Frequency of surface forms of the predicates employed in Exp. 2.

2.3.1.3 Procedure.

The normed items (60 experimental sentences) were then coded for regions of interests and inserted in a text file (together with the fillers) to be used in Linger 2.88 (http://tedlab.mit.edu/~dr/Linger/readme.html). The regions of interests were identical to Exp. 1. Moreover, comprehension questions to be presented after each item to ensure participants' engagement in the task were generated following the same rationale as in Exp. 1 (for the list of comprehension questions, see Appendix 7.6). Finally, testing was carried out following the same procedure and in the same environment as in Exp. 1. In total, 90 experimental stimuli (30 experimental items and 60 fillers) were presented to each participant.

2.3.1.4 Data analysis.

Due to low accuracy results on fillers (lower than 75% overall), 2 participants were excluded from the final analysis. Hence, data from 24 participants were analysed.

The outcome measures were: accuracy and reaction times to comprehension questions and reading times. The analysis was run using RStudio, an application for data analysis (https://www.rstudio.com/).

2.3.1.4.1 Reading times and Question Response times data:

The same outlier analysis was run as in Experiment 1: 0.54% of the original residual reading time and 0.31% of the original response times data were removed.

The same model was used to analyse the data as in Experiment 1, although the contrast used was passive vs. active [-0.5, 0.5]. Again, both accurate trials only and accurate and inaccurate trials analyses were performed and compared, and no significant difference was found between the two analyses.

2.3.1.4.2 Comprehension Accuracy data:

The same model was used to analyse the data as in Experiment 1 (although the contrast was again passive vs. active [-0.5, 0.5]). The full model (containing both intercepts and slopes for random effects) was always run first. The Results section will specify where convergence could not be met and how the model was modified to meet convergence.

2.3.2 Results.

2.3.2.1 Comprehension Question Results:

In contrast to Exp. 1, comprehension accuracy differed across the active and passive conditions. Questions following passives were found to be responded to significantly slower (passive vs. active: β=.04, t=2.49, p=.01; see Table 7) and significantly less accurately (passive: 78.3%; active: 86.1%; passive vs. active β=-.61, z=-2.941, p=.003; see Table 7[7]) than those following actives. Given that Ferreira (2003) argues that comprehension of passive sentences is selectively impaired when questions target the theta-roles of the sentence, the participants' performance on theta-role questions only was separately analysed. The analysis revealed the same significant effect: participants' performance on theta role questions was generally lower than for other question types

[7] Model did not converge with random slopes. The same model was used for the analysis of theta-role questions only and for first- vs. second-half analysis, given that more complex models did not converge.

(passive: 72.2%; active: 82.87%), and accuracy significantly differed across conditions (passive vs. actives: *β*=-.58, *z*=-2.4, *p*=.02). Similarly, there was a significant difference across conditions in Reaction Times to theta-role comprehension questions only (passive vs. actives: *β*=.06, *t*=2.9, *p*=.004). In order to exclude possible learning effects that might have confounded our results, 2 separate analyses on the data collected in the first vs. second half of the sessions[8] were run and compared. The analysis did not reveal any significant difference with respect to the overall results. Again, the lower accuracy to active sentences, with respect to previous studies (Ferreira, 2003), was likely due to the length of our stimuli. This is discussed in more details in the Results section of Experiment 1.

These data indicate that passives were indeed harder to understand than actives with stative predicates.

	Passive	Active
Accuracy	78.3% (± 8.4)	86.1% (± 7.06)
Reaction Times	3.34 (± .04)	3.3 (± .04)

Table 7: Accuracy and Reaction Times to comprehension questions in Exp. 2.

2.3.2.2 Reading Time Results:

Analysis of reading times[9] on correct trials revealed that passive sentences were read numerically faster than active ones up to the 2nd adjective, but only significantly so at the determiner (*β*=-.04, *t*=-3.18, *p* =.001). The reverse effect was observed after the second adjective (a marginally significant difference was observed at the head of the by-phrase: *β*=.04, *t*=1.72, *p*=.08). No numerical trend could be observed after the 4th region following the verb. Results are presented in Figure 3. Finally, the same first vs. second half analysis described for our offline data was conducted on the online data. Again, no significant difference was observed with respect to the overall results.

[8] The data were divided in first vs. second half as per Experiment 1.

[9] For all the reading times analyses, the most complex model, including both intercept and random slope for both subjects and items, always converged. The only exception was the first adjective region, where the model only converged with intercept for both subjects and items and slope only for subjects and not items.

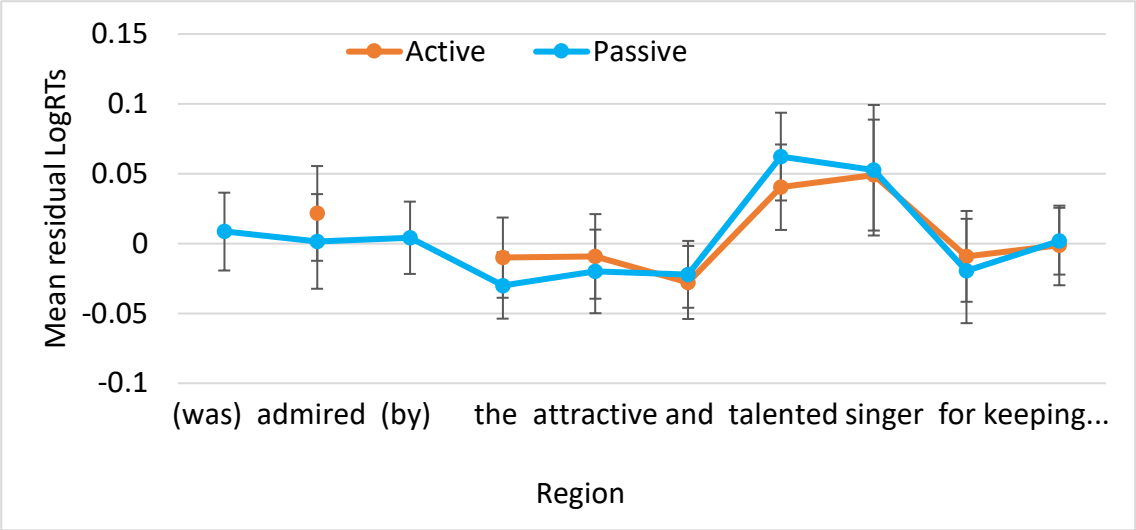

Figure 3: Self-paced reading results in Exp. 2. Region by region residual LogRTs according to verb syntax. Error bars indicate standard error of the mean.

2.3.3 Discussion.

The results of the investigation into passivization processing with stative predicates can be summarised in the following points: (1) passive sentences were comprehended less accurately and more slowly than active ones offline, indicating that they were more difficult to understand; (2) passive sentences were read faster than active ones, but the difference was significant only at the determiner of the by-phrase, differently from Exp. 1, where we found a significant difference between reading times of passives and actives in multiple regions (i.e., at the determiner, conjunction and second adjective). At the noun of the by-phrase/direct object passive sentences were read numerically slower than active ones. These results cannot be explained by a speed-accuracy trade-off, given that (1) we analysed accurate trials only and still found passives to be read faster than actives; (2) Reaction Times to comprehension questions were longer for passive than active sentences, and yet subjects made more errors on the former than the latter.

Considered together with the data obtained in Exp. 1, the results form a very interesting picture: passives sentences are processed faster than active sentences online, regardless of the predicate type. However, during offline processing, passivization does seem to interact with the predicate type: more specifically, passivizing stative predicates creates more difficulty than passivizing eventive predicates. The present results then raise at least two questions: (1) why passives are

processed faster than actives online; (2) what causes the difficulty in passivizing a stative predicate.

At present, the only processing model that is compatible with the online results are expectation-based or surprisal-based accounts of syntactic comprehension (e.g., Hale, 2001; Levy, 2007). As for the first experiment, neither plausibility nor frequency of surface form could offer an adequate answer to either questions. In fact, a frequency-based explanation is excluded by our pre-norming study, while a plausibility-based one from a post-experiment analysis, identical to the one performed in the first experiment, of the BNC instances of the tested predicates (average frequency of passive form: 30.46%).

On the other hand, the offline findings could be explained, as I initially hypothesized, by the existence of an interaction between the predicate properties and passivization. In fact, states seem to interact with passivization in the opposite direction with respect to complex event predicates, i.e., disrupting the facilitation observed in eventive processing. The underlying cause of the difficulty in processing passives of states, which is the focus of the second question raised by our data, could result from several possible factors. From a theoretical linguistic perspective, states may pose a difficulty due to any combination of the following: (1) the temporary ambiguity generated in English between adjectival and verbal passive; (2) the required coercion of the verb meaning to allow passivization. Moreover, in the case of subject-experiencer predicates, a third possibility exists. An unaffected internal argument is promoted to the subject position when a subject-experiencer predicate is passivized, thus possibly reducing the validity of employing the passive structure (see Pinker, 1989, on semantic constraints on passivization in the healthy adult grammar).

Whatever the underlying reason, interpreting a passive of a stative predicate seems to require additional computational effort with respect to interpreting a passive of a complex event and this appears to be the fundamental factor shaping the present findings. However, the presence vs. absence of a significant difference in Exp. 1 and 2 does not necessarily represent a significant difference overall. Hence, in order to strengthen my claim, i.e., that passivization and predicate structure interact in processing, I decided to run a third experiment containing the same stimuli as Exp. 1 and 2, but with a within-subject, 2 (active, passive) x 2 (event, state) design. This modification

in the original design would allow to statistically measure the hypothesized interaction between passivization and predicate semantics. Moreover, the new experiment should also allow to test the replicability of the findings of Exp. 1 and 2.

Finally, the experimental items in the new experiment will only be followed by comprehension questions targeting theta-roles, unlike Exp. 1 and 2. This should allow to further test the predictions made by the *Good Enough* model of sentence processing (Christianson et al., 2001; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016). According to the model, an incorrect application of the agent-first heuristic to passive sentence comprehension should only affect theta-roles assignment, and not other sentential aspects (e.g., location or time of the event described in a sentence), as confirmed by Ferreira's results (2003). Using questions that only target theta-roles should increase the chances of finding a complexity effect in passives, independent of predicate type, and falsifying the hypothesis of an interaction between passivization and predicate type.

2.4 Experiment 3.

The third experiment consisted in a direct replication of Exp. 1 and 2 in a within-subject, 2x2 design, manipulating both the syntax of the sentence and the predicate type as independent variables. The experiment had two main goals: (1) to replicate the results obtained in Exp. 1 and 2; (2) to directly measure an interaction between passivization and predicate type.

In addition to this modification, I decided to implement a further change. In Exp. 1 and 2, experimental items were followed by two types of questions: complex, which targeted the theta-roles of the sentence (i.e., patient or agent of the action), and simple questions, which targeted other sentential aspects (e.g., location or time of the action). In the current experiment, I decided to employ only complex questions. This modification was implemented to increase the chances of observing a difficulty effect following a passive rather than an active sentence, regardless of the predicate type, and hence to falsify my hypothesis. Ferreira (2003) found that the difficulty in processing passives could only be observed when participants were asked to answer questions regarding the thematic roles of the sentence. If indeed the difficulty in processing passives arises from the incorrect application of a heuristic strategy that interprets the

first NP as agent, then the interpretation of thematic roles should be selectively disrupted. In Exp. 1 and 2, accuracy to comprehension questions targeting theta-role questions alone was analysed, but given that they represented a small subset of items there might not have been sufficient power to observe a difference.

Given the results of the previous experiments and the aims of the present one, the predictions of Exp. 3 were: (1) passives should be processed significantly faster than actives online at several points within the by-phrase (i.e., faster reading times in passives with respect to actives), but possibly slower at the head of the by-phrase in statives only, signalling revision of the initial incorrect adjectival interpretation, as indicated by the marginally significant difference in Exp. 2; (2) an interaction between the syntax of the sentence and the predicate type should emerge in the offline data (i.e., in accuracy and reaction times to comprehension questions): in other words, passives of eventives should be as easy to understand as their active counterparts, while passives of stative predicates should be more difficult to understand than their active counterparts.

Since our previous offline task might not have been sensitive enough to detect an effect of passivization, as Ferreira (2003) did, using only theta-role targeted questions might lead to a passivization effect (i.e., passives to be more errorful than actives) across predicate type.

2.4.1 Methods and Design.

2.4.1.1 Participants.

Sixty-five British English native speakers were recruited to participate in the study (46 females; average age: 23). They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants were recruited through the UCL Sona System and received either payment or credits for their participation. All the participants were informed of the aims and procedures of the experiment and provided informed consent, approved by UCL ethics. None of the participants had previously participated in any of the experiments related to this study (including the pre-norming tests).

2.4.1.2 Stimuli.

Twenty-eight sentence sets were chosen among the sets used in Exp. 1 and 2. Each sentence set contained 4 sentences, 1 per condition. The sentences followed a 2 syntax (active vs. passive) by 2 predicate type (eventive vs. stative) design.

As in Exp. 2, to avoid the confound of an auxiliary bias, (i.e., only encounter a passive following the auxiliary "was"), 10 of the filler sentences included an auxiliary and a verb in the progressive form.

Everything else (types of filler sentences; overall sentence length) was kept identical to Exp. 1 and 2. Examples of experimental items are presented in Table 8 (for the complete list of items, see Appendix 7.7).

Eventive	Passive	The guitarist was rejected by the attractive and talented singer in the concert hall next to the Irish pub.
Eventive	Active	The guitarist rejected the attractive and talented singer in the concert hall next to the Irish pub.
Stative	Passive	The guitarist was admired by the attractive and talented singer for keeping the band focused through the whole tour.
Stative	Active	The guitarist admired the attractive and talented singer for keeping the band focused through the whole tour.

Table 8: Example of experimental items in Exp. 3.

2.4.1.2.1 Pre-Norming: Plausibility.

A pre-norming plausibility test was administered via an online questionnaire designed in Qualtrics (https://www.qualtrics.com/). Filler implausible items, rating scale, structure of experimental items (only active sentences; manipulation: order of arguments) were identical to the plausibility questionnaires run in Exp. 1 and 2.

Participants for this study were recruited via Prolific Academic, a platform for online research (https://prolific.ac/). Forty-four participants (25 females; average age: 30.5) were recruited to participate in the online questionnaire.

Data were analysed using a linear mixed effects model, containing the order of theta-roles as fixed effect and both subjects and items as random effects (including both

intercepts and slopes). The contrast used was passive order vs. active order [0.5, -0.5]. P-values were determined through treating the t-value as a z-statistic (Barr et al., 2013). Based on the plausibility ratings, 28 sets that displayed no significant difference between the two possible orders of arguments (average active: 5.41; average passive: 5.40; β=.01, t=.09, p=.92) were chosen. Hence, the items did not significantly differ in the plausibility of thematic roles order. However, the items in the stative conditions were rated significantly more plausible (β=.31, t=2.46, p=.01) than the ones in the eventive conditions. Plausibility did not differ according to an interaction between syntax and predicate type. These results allow the possibility of analysing syntax (active vs. passive) and its interaction with verb type, but not the predicate type (event vs. state) effect alone.

2.4.1.2.2 SUBTLEX Corpus Analysis: Frequency analysis across predicates.

A frequency analysis, run on data taken from SUBTLEX-UK (which contains word frequencies based on subtitles of British television programmes; van Heuven, Mandera, Keuleers & Brysbaert, 2014), revealed that our eventive verbs are twice as frequent as our stative ones (eventives Zipf frequency: 4.65; statives Zipf frequency: 2.48).

Hence, any effect of predicate type would be hard to interpret, given that the two predicate groups differed both in terms of frequency and plausibility.

2.4.1.3 Procedure.

The normed items (112 experimental sentences) were then run in a self-paced reading paradigm using Linger 2.88 (http://tedlab.mit.edu/~dr/Linger/readme.html). The regions of interests were identical to Exp. 1. Moreover, comprehension questions followed each sentence, but only targeted theta-roles in the experimental items (e.g., “Did the musician reject the guitarist?”; for the list of comprehension questions, see Appendix 7.8). Fillers were followed by complex and simple questions, as in previous experiments, to avoid creating a bias in attention towards specific parts of the sentence (in this case, the NP-VP-NP part). Finally, testing was carried out following the same procedure and in the same environment as in Exp. 1 and 2. In total, 88 experimental stimuli (28 experimental items and 60 fillers) were presented to each participant.

2.4.1.4 Data analysis.

Due to low accuracy results on fillers (lower than 75% overall), 5 participants were excluded from the final analysis. Hence, data from 60 participants were analysed.

The outcome measures were: accuracy and reaction times to comprehension questions and reading times. The analysis was run using RStudio (https://www.rstudio.com/).

2.4.1.4.1 Reading times and Question Response times data:

The same outlier analysis was run as in Experiment 1 and 2: 0% of the original residual reading time and 0% of the original response times data were removed.

The residual logRT reading times and logRT reaction times were analysed using a linear mixed effects model including syntax and predicate type along with their interaction as fixed effects and both subjects and items as random effects (including both intercepts and slopes). The contrasts used were passive vs. active [-0.5, 0.5] and eventive vs. stative [-0.5, 0.5]. Again, both accurate trials only and accurate and inaccurate trials analyses were performed and compared, and no significant difference was found between the two analyses.

2.4.1.4.2 Comprehension Accuracy data:

Offline data (accuracy) were analysed using a mixed effects logistic regression with a binomial distribution including the same effects and contrasts as the analysis of the online data. The full model (containing both intercepts and slopes for random effects) was always run first. The Results section will specify where convergence could not be met and how the model was modified to meet convergence.

2.4.2 Results.

2.4.2.1 Comprehension Question Results:

The offline results did not replicate the results of Exp. 1 and 2. In comprehension accuracy we found an effect of syntax (β=-.44, z=-2.08, p=.04) due to accuracy being lower following a passive rather than active sentence and an effect of predicate type (β=-.29, z=-2.09, p=.04), due to lower accuracy following a stative rather than an

eventive predicate (see Table 9[10]). The interaction between predicate and syntax was not significant, contrary to the predictions based on Exp. 1 and 2. Reaction Times (RTs) to comprehension questions showed a significant effect of syntax (β=.06, t=4.26, p<.001), due to RTs being longer following a passive sentence than its active counterpart (see Table 9). In RTs, the predicate type effect was not significant nor was the interaction. Overall, the data indicate that passive sentences were more difficult to interpret than active ones, independent of predicate type, and sentences containing a stative predicate were more difficult than sentences containing an eventive one, independent of syntax. We did not find an interaction between syntax and predicate type, contrary to what we expected.

		Passive	Active
Eventive	Accuracy	82.1% (± 4.95)	85.9% (± 4.49)
	Reaction Times	3.36 (± .03)	3.29 (± .02)
Stative	Accuracy	79.1% (± 5.26)	82.1% (± 4.95)
	Reaction Times	3.34 (± .03)	3.29 (± .02)

Table 9: Accuracy and Reaction Times to comprehension questions in Exp. 3.

Given this discrepancy, it is worth noting that unlike Exp. 1 and 2, we observed a large amount of variance in the offline data. The variance was largely affected by 15 (out of 60) participants performing at chance on actives and 21 (out of 60) on passives. Again, this is unlike what we observed in Exp. 1 and 2, where no participant performed at chance on actives and only one participant performed at chance on passives in both Exp. 1 (out of the 30 analysed) and Exp. 2 (out of the 24 analysed). I will return to contemplate the underlying source of these differences in the Discussion section.

Finally, in order to exclude possible learning effects that might have confounded our results, 2 separate analyses on the data collected in the first vs. second half of the sessions[11] were run and compared. In accuracy, differently from the overall analysis, the syntax effect was only observed in the second half, and not first half, of the sessions, which contradicts a learning effect interpretation of this data. Moreover, in RTs to

[10] Model only converged with intercepts.

[11] Data from trials 7-51 (the first 6 items were practice items) were analysed as “first-half” and data from trials 51-96 as “second-half”.

comprehension questions, the syntax effect was significant across both sessions, indicating, once again, that the absence of a syntax effect in accuracy in the first half of the session cannot be due to learning effects.

2.4.2.2 Reading Time Results:

The analyses of reading times on correct trials only replicated the results of Exp. 1 and 2. In fact, we found a significant effect of syntax at the verb (β=-.06, t=-2.68, p=.007[12]), determiner (β=-.05, t=-6.07, p<.001), first adjective (β=-.02, t=-2.08, p=.04), and conjunction (β=-.02, t=-2.56, p=.01; see Figure 4), indicating that passives were read faster than actives in these regions. No complexity effect was found at the head of the by-/object phrase in passive with respect to active sentences, contrary to what was reported in Exp. 2. No effect of predicate type or interaction between syntax and predicate type were found. Overall, the data further confirm that passive sentences are processed faster than active ones, independent of predicate type. Finally, the same first vs. second half analysis described for the offline data was conducted on the online data. No significant difference was observed with respect to the overall results.

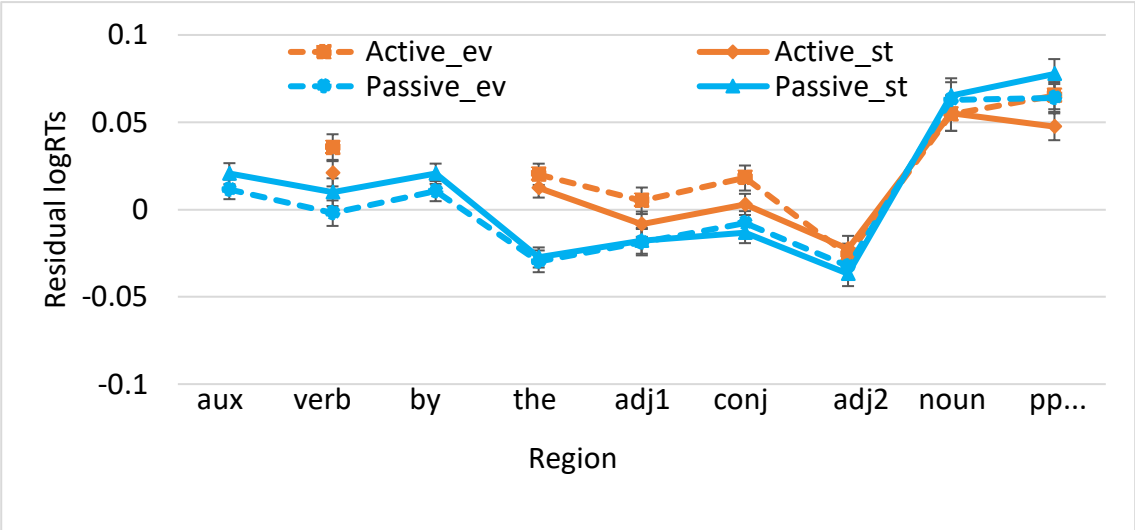

Figure 4: Self-paced reading results in Exp. 3. Region by region residual LogRTs according to verb syntax. Error bars indicate standard error of the mean.

[12] Model for the verb region analysis only included Syntax in the structure of the random effect of the items, as the verb differed across predicate type.

2.4.3 Discussion.

The results of the within subject syntax (active, passive) by predicate type (event, state) experiment can be summarised in the following points: (1) passives were read significantly faster than actives at several regions, independent of predicate type (verb, determiner, adjective and conjunction); (2) passives were more errorful than actives, independent of predicate type; (3) there was no interaction between passivization and predicate type in the online or offline data. As per Exp. 2, the results presented in points (1) and (2) cannot be explained by a speed-accuracy trade-off, given that (1) we analysed accurate trials only and still found passives to be read faster than actives; (2) Reaction Times to comprehension questions were longer for passive than active sentences, and yet subjects made more errors on the former than the latter.

The results of the third experiment were hence mixed with respect to my predictions. In terms of the online reading data, passives were found to be read faster than actives, thus replicating the results of Exp. 1 and Exp. 2. However, regions that were only numerically faster in Exp. 2 were found to be significantly faster in this experiment, just as they were in Exp. 1. There was no significant difference at the head of the by-phrase, contrary to what was suggested by the results of Exp. 2.

In line with the previous literature on passive sentence processing (Carrithers, 1989; Traxler et al., 2014), Exp. 1-3 found that passives are processed faster than actives. This implies that they are not more complex to process than actives. These results are compatible with surprisal-based or expectation-based accounts of sentence comprehension, but not with the *Good Enough* model, syntactic approach or frequentist approach to passive processing.

The offline accuracy results completely veered from the predictions based on the results from Exp. 1 and 2. Comprehension accuracy was lower in passives with respect to actives, regardless of the predicate type. More specifically, no significant interaction between passivization and predicate type was found, contrary to the independent results from Exp. 1-2 and results reported in the relevant literature (healthy adults – Ambridge et al., 2016; children acquiring their first language – Maratsos et al., 1985; O'Brien et al., 2006; Volpato et al., 2015; aphasic patients – Grodzinsky, 1995; see Chapter 1 for more details). In fact, in Exp. 2 with stative predicates, passivization had a significant offline effect, which was not observed in Exp. 1 that only contained eventive

predicates. However, the verification questions in Exp. 1 and 2 targeted theta-roles and other sentential aspects, while questions in Exp. 3 only targeted theta-role assignment, thus providing a more powerful test of interpretation difficulty. Thus, it may be that an effect on accuracy was not observed in Exp. 1 because the test was not sufficiently sensitive across an adequate number of trials (although this effect was clearly seen with the same test with stative predicates in Exp. 2, indicating some further variability, possibly in predicate characteristics, as will be further discussed in Chapter 3 and 4).

The finding that passives were more errorful than actives in the offline verification task contrasts with the online data, as they indicate that passives are more difficult than actives. While these offline results are predicted by the *Good Enough* model, the syntactic complexity approach and the frequentist approach, they are not predicted by surprisal- or expectation-based accounts of sentence comprehension. Collectively, no model would adequately account for the results observed in this and previous experiments. First, however, I need to establish which data across Exp. 1-3 are most reliable to build a model from.

There are several differences across Exp. 1-3 that could be affecting the offline results and explain the difference in results across experiments. Some of these may also help in establishing a model(s) commensurate with all the data. As suggested in the Results section of Exp. 3, the large variance present in our dataset draws concerns about different characteristics between Exp. 1&2 vs. 3. Firstly, the large variance observed in Exp. 3 might come from real variability amongst the participants. The most notable inter-subject variability measure in sentence comprehension research is verbal Working Memory (vWM) span, which was found to correlate with linguistic tasks across several studies (e.g., Boyle et al., 2013; Conway et al., 2005; Daneman & Carpenter, 1980; Dede et al., 2004; Jaeggi et al., 2010; Kim & Christianson, 2013; Roberts & Gibson, 2002; Sprouse et al., 2012; Swets et al., 2007; see Chapter 1 for a summary of these studies). Moreover, the correlation between vWM span and subjects' performance on a linguistic task was found to positively correlate with task difficulty (e.g., Just & Carpenter, 1992). Our previous experiments might not have captured this inter-subject variability possibly due to a smaller sample size (30 participants were recruited in Exp. 1, 24 in Exp. 2 and 60 in Exp. 3), or due to the type of comprehension questions asked in Exp. 1&2 vs 3. Experimental sentences in the third experiment were only followed by complex

questions, targeting the thematic-role assignment, while Exp. 1 and 2 used a mixture of simple (targeting other sentential aspects, such as time or location) and complex questions. If the variability observed in Exp. 3 was indeed determined by different vWM spans among subjects, increasing the difficulty of the task might have increased the observed variance in our samples, from Exp. 1&2 to Exp. 3. Furthermore, the difference in accuracy between active and passive sentences observed in Exp. 3 might emanate from demands on vWM in completing the comprehension task. A review of the literature on sentence processing and vWM (see Chapter 1 for more details) revealed that vWM correlates with offline but not online linguistic processing. This discrepancy in the correlation between vWM and offline data, but not online ones, bears some semblance to what we have observed in terms of passive difficulty: passives show a facilitation in processing online, while a difficulty offline. This would indicate that, while variability in vWM does not affect participants' performance on the self-paced reading task, variability in vWM might contribute to the complexity effect in performing the verification task on passive sentences. In particular, such an offline task may be more demanding on memory mechanisms in the case of passive sentences, for example through requiring further manipulations in memory. If indeed a correlation between vWM span and accuracy to comprehension questions (i.e., the lower the vWM span, the larger the difference between actives and passives, with passives being more errorful than actives) were to be found, then I could hypothesize that processing a passive sentence is not difficult *per se* (as suggested by the faster reading times in online tasks). Rather, encoding/retrieving the full representation of these sentences might be more difficult than for active sentences. This would provide an adequate explanation for the contrast between our online and offline findings.

Secondly, the variance might be caused by differences in the number of conditions across experiments. Given that the third experiment was intended to be a replication of Exp. 1 and 2, the same items were employed. However, the new experiment had 4 conditions instead of 2 or 3, as in the previous experiments, thus greatly reducing the number of items per condition (from 10 in Exp. 1, to 14 in Exp. 2, to only 7 in Exp. 3). This modification to the original design might have underpowered Exp. 3 with respect to Exp. 1 and 2, and introduced a large amount of noise in the

dichotomous offline data, which could represent the underlying cause of the variance observed across results.

Finally, the difference in the offline results of Exp. 3 with respect to Exp. 1 and 2 might derive from the different proportion of eventive/stative predicates across experiments. In fact, eventives were the prevalent predicate type in Exp. 1, given that they were contained both in experimental and filler items. Despite the presence of eventive predicates in the filler items of Exp. 2, the nature of the experimental stimuli (i.e., only containing states) created an equal balance between the two predicate types. On the contrary, Exp. 3 contained more eventive than stative predicates, since the former were present both in the experimental and in the filler items. Given the unambiguous verbal nature of the passivized eventives, participants in Exp. 3 might have developed a biasing strategy towards the verbal interpretation of the passive structure, independent of predicate type, thus levelling the performance across sentences.

In order to control for factors that might have affected the offline results in Exp. 3 compared to Exp. 1&2 and that could possibly explain the discrepancy between online and offline results, Exp. 4 was designed. Beside the self-paced reading and verification tasks, two additional memory tasks, which are generally considered to be a reliable measure of working memory, were added: the sentence reading span task and the N-back task (a comprehensive description of the tasks, together with experimental stimuli, are given in the Introduction and Methods and Design sections of Experiment 4). In order to control for the possible presence of noise due to a small number of items per condition, the items of Exp. 3 were split in a between-subject design, according to the predicate type. In doing so, the number of items per condition were doubled while keeping every other aspect identical across experiments, thus allowing to perform between-subject analyses on a possible interaction between passivization and predicate type. Moreover, by splitting the predicate types between subjects, the same conditions of Exp. 1 and 2 with respect to the proportion of eventive/state predicates were recreated, thus avoiding a possible bias towards a verbal interpretation of stative predicates.

2.5 Experiment 4.

Experiment 4 primarily aimed at investigating the contribution of memory resources to the difference in comprehension accuracy between passives and actives, but also controlled for the other design issues that varied across Exp. 1&2 vs. 3 (see Exp. 3 Discussion). A correlation between vWM span and the difference in accuracy on actives vs. passives in offline verification questions might be expected for two reasons. Firstly, the *Good Enough* model (Christianson et al., 2001; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016) would predict subjects with a lower WM span to rely more on heuristics than algorithmic processes, given that the latter pose greater demands on the parser than the former and might thus require a larger vWM span. Secondly, the syntactic approach to passive processing considers passives to be derived from actives, via a movement-like operation. It might hence be more difficult for participants with a lower vWM span to maintain a derived representation in memory in order to answer a comprehension question. In both scenarios, we would expect vWM span to negatively correlate with the difference in accuracy between actives and passives (i.e., the lower the vWM span, the larger the difference between actives and passives, with passives being more errorful than actives).

Two vWM tasks were selected based on a survey of the vast literature on the relation between vWM capacity and linguistic tasks: the reading span task and the N-back task (e.g., Boyle et al., 2013; Conway et al., 2005; Daneman & Carpenter, 1980; Dede et al., 2004; Jaeggi et al., 2010; Kim & Christianson, 2013; Roberts & Gibson, 2002; Sprouse et al., 2012; Swets et al., 2007). The reading span task requires participants to perform two tasks at the same time: semantic acceptability judgment and serial recall letter. The two tasks are interleaved. Given the nature of the tasks and the fact that the two tasks interfere with one another, the reading span task is thought to tap into the coordination of storage, processing and retrieval of linguistic input. Because these functions are necessary to perform reading comprehension tasks, it is not surprising that several studies reported a strong correlation between performance on the reading span task and sentence comprehension tasks (Daneman & Carpenter, 1980; Kim & Christianson, 2013; Roberts & Gibson, 2002; Swets et al., 2007). In the ground-breaking paper that described the reading span task for the first time, Daneman & Carpenter

(1980) showed that the subjects' score on the reading span task correlated with three measures of reading comprehension: the word span test, the Verbal/Critical Reading SAT score, and a comprehension test targeting fact retrieval and pronominal reference.

On the other hand, the N-back task is based on externally triggered recognition. During the task, subjects are required to encode new items and their position into a pre-existing list and to compare the new items with items stored in the list, in order to indicate whether the target appeared 1, 2 or 3 positions back in the list, or to inhibit incorrect responses to lures (Kane & Engle, 2002). A consistent correlation was also observed between the N-back task and sentence comprehension tasks (Roberts & Gibson, 2002). Roberts and Gibson (2002) designed a sentence memory task in which participants had to listen to a series of sentences (from 2 to 5), which contained structures that varied in syntactic complexity (relative clauses, sentential complements, double object relative clauses), and were then asked to answer a question about one of the sentences presented in the series. They found that subjects' performance on the N-back task strongly correlated with subjects' accuracy on the sentence memory task.

Interestingly, Roberts and Gibson (2002) also tested the correlation between participants' performance on the sentence memory task and their performance on the reading span task, further confirming the results reported by Daneman and Carpenter (1980). Roberts and Gibson (2002) additionally tested the correlation between the scores obtained by subjects in the N-back task and those obtained in the reading span task and found no correlation between the two memory scores. This is in line with the existing literature on the reliability and validity of working memory tasks, which consistently found no correlation between the scores as reported by the two tasks (Jaeggi, Buschkuehl, Jonides & Perrig, 2008; Jaeggi et al., 2010; Kane, 2005), thus confirming that the two tasks tap into different working memory constructs.

Beside investigating the contribution of memory resources to subjects' performance on passives, Exp. 4 aimed at controlling for factors that differed between Experiment 1&2 and Experiment 3, in order to allow a fairer comparison across the results obtained in the previous experiments. As a reminder, Exp. 3 contained a fewer number of items per condition and differed in the proportion of eventive/stative predicates. The larger proportion of eventive passives, which can only be interpreted

verbally, in Exp. 3 might have biased participants to interpret any passive as verbal, independent of the predicate characteristics, thus limiting the role of revision.

To address these potential issues and establish which results are reliable between Experiment 1&2 and Experiment 3, the same items from Exp. 3 were used with predicate type (event vs. state), as a between-subjects factor, while syntax (active vs. passive) remained a within-subjects factor.

If part of the variance observed in offline performance in Exp. 3 was caused by inter-subject variability in working memory span, we should find: (1) a significant effect of vWM in the mixed effects logistic regression analysis on accuracy data; (2) a correlation between the results of the offline comprehension task and the participants' vWM span, i.e., participants with a lower vWM span should have a larger difference in accuracy between passives and actives. This would indicate that subjects with a lower vWM span do rely more on heuristics than algorithmic process to parse sentences, as per the *Good Enough* model (Christianson et al., 2001; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016), or that maintaining the representation of a derived sentence, in this case, a passive sentence, is taxing, as per the syntactic approach. However, the faster reading times of passive sentences consistently observed in accurate trials indicate that real-time parsing of passives is not difficult. Rather, operating on the full representation of the passive sentence for the verification task may be, and this difficulty could increase in subjects with a lower vWM span.

Finally, Exp. 4 also offers the chance to replicate the online findings of Exp. 1, 2 and 3, namely passives should be read faster than actives across several regions, independent of predicate type. Replicating the online findings is crucial, given the replication crisis recently observed in Psychology (e.g., Ito, Martin & Nieuwland, 2016; Shanks et al., 2015).

2.5.1 Methods and Design.

2.5.1.1 Participants.

A hundred and one British English native speakers were recruited to participate in the study (68 females; average age: 24). A larger sample of participants was used in order

to ensure sufficient variability in vWM to test our hypotheses. They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants were recruited through the UCL Sona System and received either payment or credits for their participation. All the participants were informed of the aims and procedures of the experiment and provided informed consent, approved by UCL ethics. None of the participants had previously participated in any of the experiments related to this study (including the pre-norming tests).

2.5.1.2 Stimuli.

As mentioned in the Discussion section of Exp. 3, to control for noise associated with a small number of items per condition, we split the items of Exp. 3 in a between-subject design, according to predicate type, while syntax of the sentence (active vs. passive) remained a within-subject factor.

The experiment contained 2 sentence sets, each with 28 items taken from Exp. 3.

Everything else (types of filler sentences; overall sentence length) was identical to Exp. 3. Examples of experimental items are presented in Table 10 (for the complete list of items, see Appendix 7.7).

Eventive	Passive	The guitarist was rejected by the attractive and talented singer in the concert hall next to the Irish pub.
Eventive	Active	The guitarist rejected the attractive and talented singer in the concert hall next to the Irish pub.
Stative	Passive	The guitarist was admired by the attractive and talented singer for keeping the band focused through the whole tour.
Stative	Active	The guitarist admired the attractive and talented singer for keeping the band focused through the whole tour.

Table 10: Example of experimental items in Exp. 4.

2.5.1.3 Memory tasks.

2.5.1.3.1 Reading span task.

The reading span task consisted of 2 separate, but intermixed tasks. The first task was a sentence judgment task: participants were asked to decide whether the sentences (10 to 15 words long) presented to them on a PC screen were correct or incorrect. As specified in the instructions, their decision had to be based on semantic/pragmatic, rather than grammatical, considerations (see Table 11 for example sentences). Each item was presented at the centre of the screen and would automatically disappear after 1000ms, regardless of whether the participant had made a decision or not. Participants were instructed to be as accurate as possible (maintain the overall accuracy in the judgment task higher than 85%) and received feedback on their accuracy after each trial.

1a	The host greeted all the guests and asked them to sit at the table.	Correct
1b	The host greeted all the guests and asked them to sit at the sky.	Incorrect
2a	John never liked crowds and that is why he now lives in the country.	Correct
2b	John never liked chocolate and that is why he now lives in the country.	Incorrect

Table 11: Example of stimuli in the sentence judgment task.

The second task consisted in a letter recall task. After each judgment, a letter (only uppercase consonants: F, H, J, K, L, N, P, Q, R, S, T, Y) would appear on the screen for 1000ms. At the end of each block of trials, participants were asked to recall all the letters in the correct order and received feedback on their performance (see Figure 5). Each block could contain between 2 and 5 sentences.

Participants completed 6 practice trials of the first task, 3 practice trials of the second task and 3 practice trials of the two tasks interleaved. Each trial contained 2 to 5 items (i.e., sentences plus letter) that were presented in a random order, rather than ascending order, to avoid the possibility that participants would rely on strategies (e.g., proactive interference) that might come from anticipating the number of items to recall per trial. Each set of items (2, 3, 4, 5) was presented 3 times, hence participants completed 12 trials. Feedback was given on both tasks to ensure participants' engagement, which implied that both the storage and processing functions of working memory were actually at work.

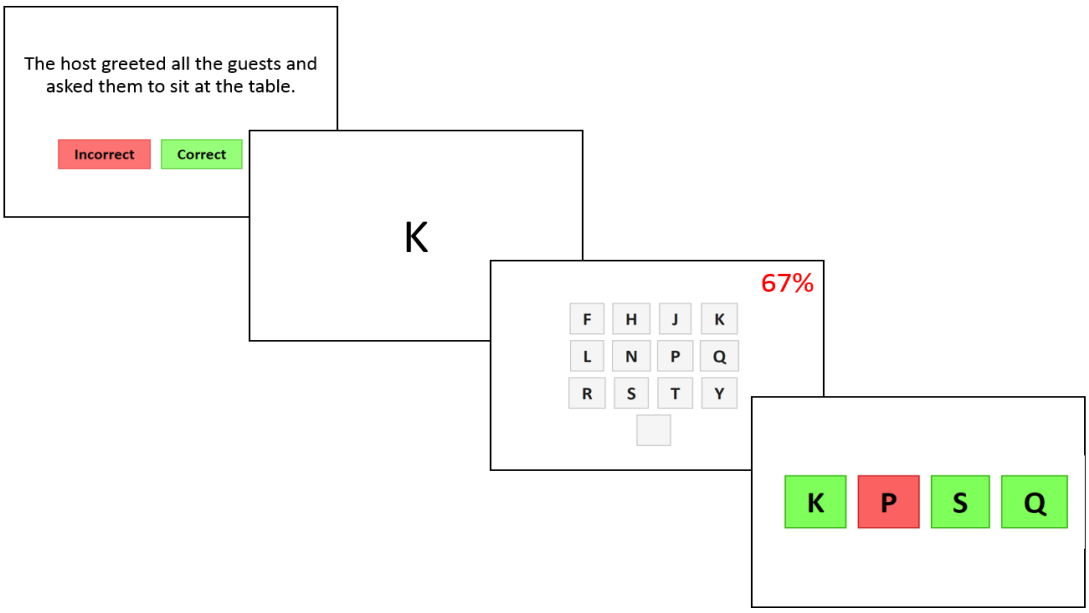

Figure 5: Example of reading span task. Experiment flow.

2.5.1.3.2 N-back task.

In the N-back task participants were presented with a series of letters (only uppercase consonants) on a PC screen and had to respond (by pressing the "A" key on the keyboard) when the letter currently presented was identical to the one presented n-positions back (e.g., 1-, 2-, 3-positions back; see Figure 6 for an example). Participants received 9 trials of practice per level tested, and were tested on 3 blocks of 15 trials (i.e., letters) per level. In each block, 5 trials were targets and 10 were not. Each letter was presented for 500ms. When the letter disappeared, a blank screen appeared for 2000ms. Hence, each trial lasted 2500ms.

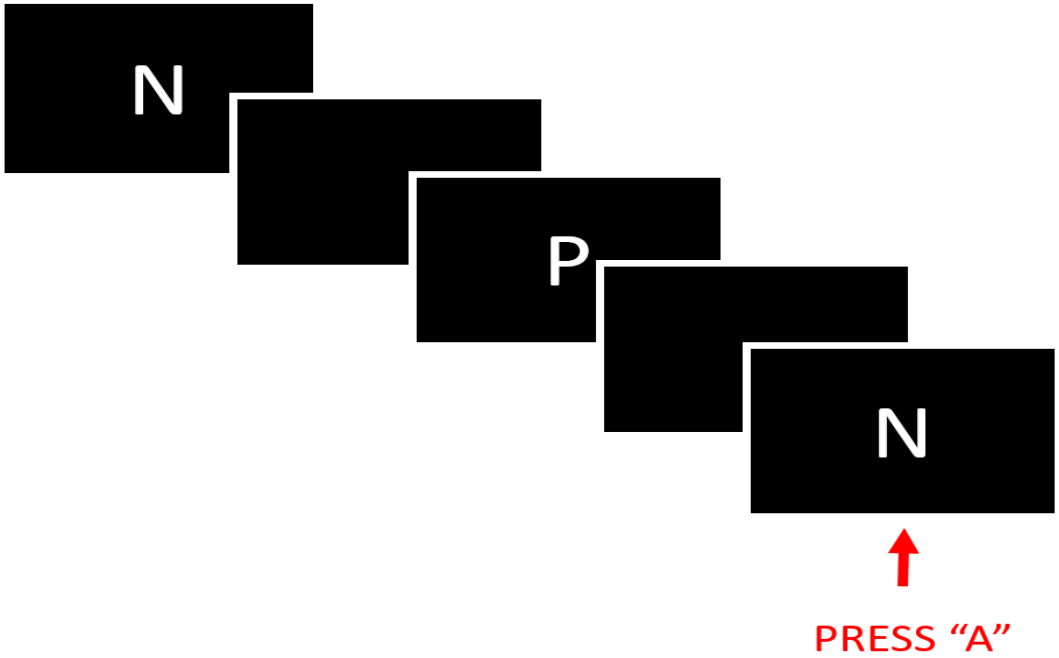

Figure 6: Example of 2-back task. Experiment flow.

2.5.1.4 Procedure.

For the self-paced reading task, the same procedure as in Exp. 3 was used.

In order to maintain the same procedure as in Exp. 1, 2 and 3, the self-paced reading task was always administered first. To avoid possible order effects, the 2 memory tasks were then presented in a counterbalanced order, i.e., half of the participants performed the reading span task first, and half the N-back task first.

2.5.1.5 Data analysis.

In the self-paced reading task, 5 participants scored lower than the exclusion threshold in accuracy on fillers (75%), hence they were excluded from the final analysis. Data from 96 participants were analysed.

The outcome measures were: accuracy and reaction times to comprehension questions and reading times. The analysis was run using RStudio, an application for data analysis (https://www.rstudio.com/).

2.5.1.5.1 Reading times and Question Response times data:

The same outlier analysis was run as in Exp. 3: 1.5% of the original residual reading time and 0% of the original response times data were removed.

The residual logRT reading times and logRT reaction times were analysed using the same model and contrasts as in Exp. 3. Again, both accurate trials only and accurate and inaccurate trials analyses were performed and compared, and no significant difference was found between the two analyses.

2.5.1.5.2 Comprehension Accuracy data:

Offline data (accuracy) were analysed using a mixed effects logistic regression with a binomial distribution including the same effects and contrasts as the analysis of the online data. The full model (containing both intercepts and slopes for random effects) was always run first. The Results section will specify where convergence could not be met and how the model was modified to meet convergence.

2.5.1.5.3 Correlation analysis:

2.5.1.5.3.1 Reading Span Task. Data from the reading span task were scored using the following procedure. One point was given to each trial if the subject had responded correctly to both tasks, i.e., if the participant had correctly judged the sentence and correctly recalled the letter in its serial position. Hence, each participant obtained an absolute score out of 42 (total number of trials; Kim & Christianson, 2013; Swets, et al. 2007). This score was then standardized to avoid collinearity (Belsley, 1991).

2.5.1.5.3.2 N-back Task. Data from the N-back task were scored by subtracting the total number of false hits across blocks (when the participant pressed "A" even if the letter currently presented was not a target) from the total number of correct hits across blocks. The score was then divided by the total number of blocks (Jaeggi et al., 2010) and the results standardized to avoid collinearity (Belsley, 1991).

The standardized scores of the reading span task and the N-back task were then inserted in the fixed effects structure of both the linear mixed effects model used to analyse reaction and response times of the self-paced reading task, the mixed effects logistic regression used to analyse accuracy data collected in the self-paced reading task and our correlation analyses. In our correlation analyses, the following measures were considered: accuracy (mean overall; difference between actives and passives), Reaction Times (mean overall; difference between actives and passives), Reading Times (entire

sentence and critical regions only; mean overall; difference between actives and passives).

The analyses were performed in R (https://www.rstudio.com/).

2.5.2 Results.

2.5.2.1 Comprehension Question Results:

Offline accuracy and response time results replicated the results of Exp. 2 and 3. There was a significant difference in comprehension accuracy following a passive with respect to an active sentence (active: 85.42%; passive: 79.76%; β=-.59, z=-3.38, p<.001[13]; see Table 12). No other effect was significant.

Response times to comprehension questions following passive sentences were significantly longer than following their active counterparts (β=.07, t=6.2, p<.001; see Table 12), similarly to what we found in our previous experiments. Moreover, we found a significant effect of memory span as measured by the sentence span task (β=-.03, t=-2.79; p=.005), indicating that people with a larger memory span were faster in answering comprehension questions. No other effect was significant.

In order to exclude possible learning effects that might have confounded our results, we ran and compared 2 separate analyses on the data collected in the first vs. second half of the sessions[14]. The analysis did not reveal any significant difference with respect to the overall results. Finally, as previously, the lower accuracy to active sentences compared to results reported by Ferreira (2003) is likely due to our longer and more complex stimuli, as explained in more details in the Results section of Experiment 1.

[13] For all accuracy analyses, the model only converged with random intercepts and not slopes.

[14] The data were divided in first vs. second half as per Experiment 1, 2 and 3.

		Passive	Active
Eventive	Accuracy	81.3% (± 4)	87.2% (± 3.4)
	Reaction Times	3.36 (± .02)	3.29 (± .02)
Stative	Accuracy	78.3% (± 4.2)	83.6% (± 3.8)
	Reaction Times	3.38 (± .02)	3.32 (± .02)

Table 12: Accuracy and Reaction Times to comprehension questions in Exp. 4.

2.5.2.2 Reading Time Results:

The analysis of the reading times for correct trials only replicated the results of our previous experiments, but more robustly. Passives were read significantly faster than actives at the verb (β=-.03, t=-3.47, p<.001), at the determiner of the by-phrase (β=-.05, t=-8.79, p<.001), at the first adjective (β=-.02, t=-3.16, p=.002), at the conjunction (β=-.03, t=-3.78, p<.001), and at the second adjective[15] (β=-.02, t=-2.6, p=.009, see Figure 7). However, as reported in Exp. 2, this trend reversed at the head of the by-phrase, where passive sentences were read marginally slower than active sentences (β=1.84, t=1.89, p=.059), signalling a possible point of revision or difficulty in integration.

No effect of predicate type on reading time data was found.

An effect of vWM span as measured by the N-back task was only found at the conjunction (β=-.01, t=-2.1, p=.04), indicating that people with a larger memory span were faster at reading this region. No other effect was significant.

Finally, to examine possible learning effects, we ran and compared 2 separate analyses on the data collected in the first vs. second half of the sessions. The analysis did not reveal any significant difference with respect to the overall results.

[15] The model only converged with syntax and not predicate type in the structure of the Item random effect.

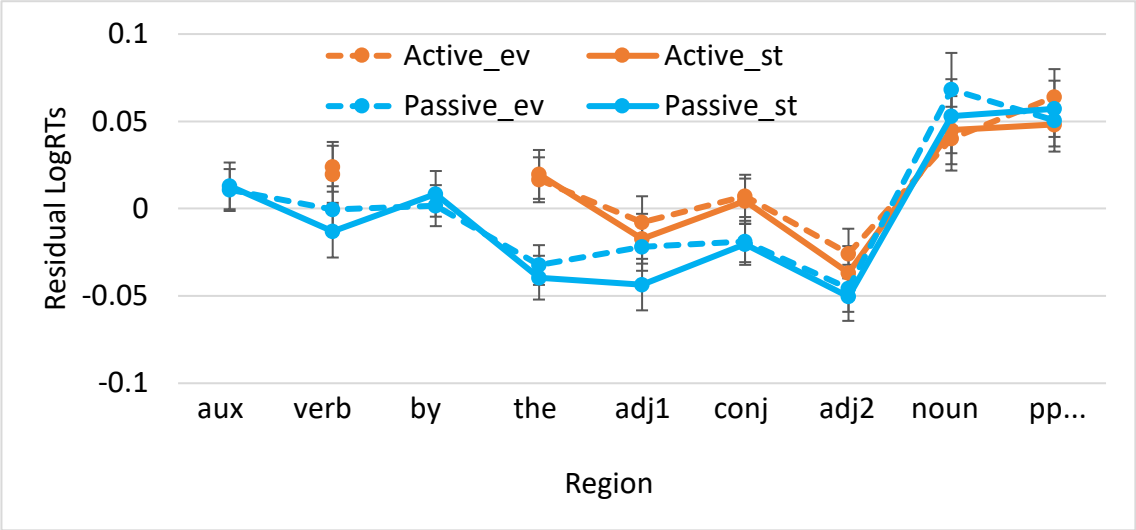

Figure 7: Self-paced reading results in Exp. 4. Region by region residual LogRTs according to verb syntax. Error bars indicate standard error of the mean.

2.5.2.3 WM Correlations:

With respect to vWM measures, our correlation analyses revealed a significant correlation between Response Times to comprehension questions and vWM span as measured by both the sentence span task (r=-.22, t=-2.19, p=.03) and the N-back task (r=-.23, t=-2.28, p=.03). The greater the vWM span the faster the response time.

More importantly, we found a correlation between the difference in accuracy in active vs. passive sentences and vWM span as measured by the sentence span task (r=-.2, t=-2.01, p=.047; see Figure 8), due to the fact that participants with a lower vWM presented a larger difference between active and passive sentences. Moreover, vWM span as measured by the sentence span task correlated with accuracy on passive sentences (r=.22, t=2.2; p=.03; see Figure 9), due to the fact that participants with a lower vWM span performed worse on passive sentences. Sentence span scores did not correlate with accuracy on active sentences, or accuracy overall. The N-back task scores did not correlate with any accuracy measure (average accuracy, difference in accuracy between actives and passives, accuracy on actives only, accuracy on passives only).

Finally, there was no correlation between online data (reading times of the entire sentence and critical regions only; mean overall reading times; difference between reading times on active and passive) and vWM span. Hence, both the sentence reading span and the N-back span appear to be good predictors of offline rather than online processing. These results are consistent with what is reported in the literature on

working memory, which found effects of memory span on offline rather than online processing (Caplan et al., 2011; Evans et al., 2014).

Overall, the data confirm that passive sentences are processed faster than active ones, up to the head of the by-phrase where participants are marginally slower in passives with respect to actives, which could possibly indicate the presence of revision or integration cost.

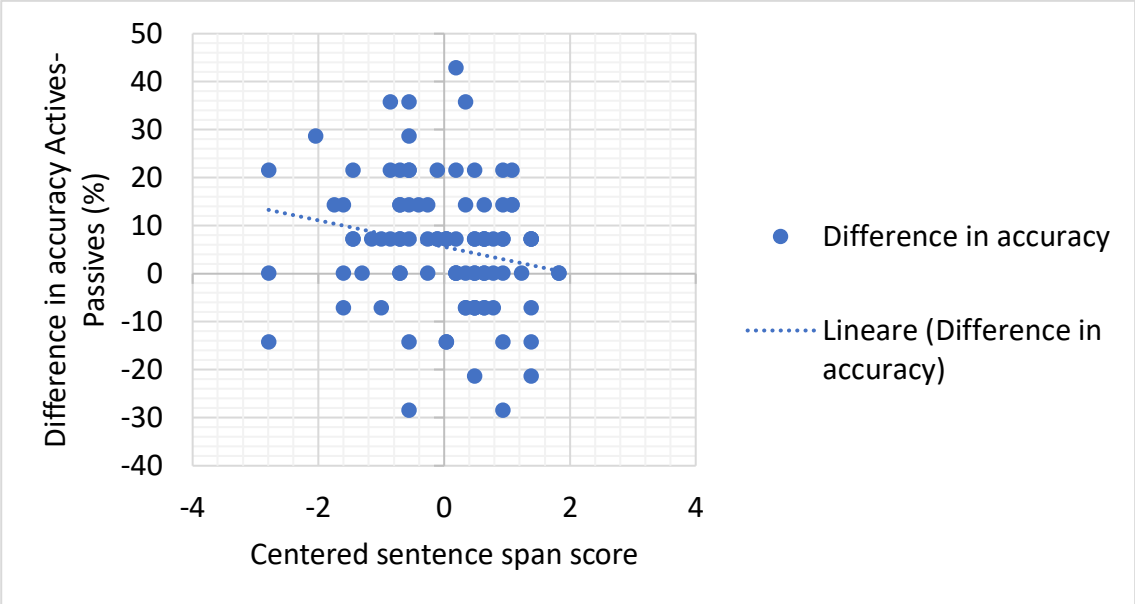

Figure 8: Correlation between Sentence Span score and difference in accuracy to comprehension questions (actives-passives) in Exp. 4.

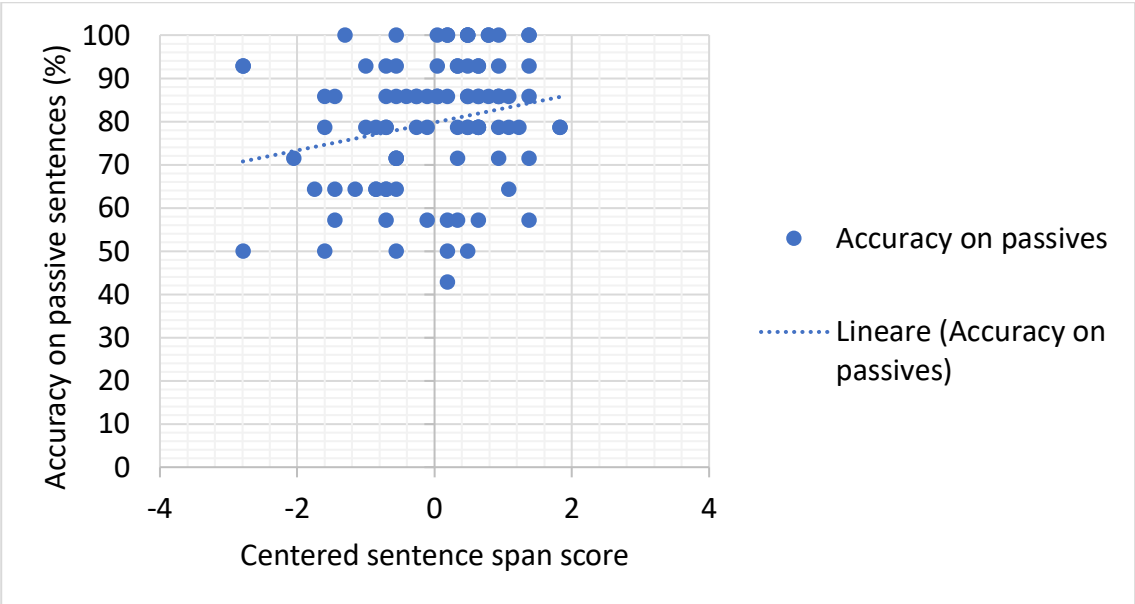

Figure 9: Correlation between Sentence span score and accuracy to comprehension questions on passive sentences in Exp. 4.

2.5.3 Discussion.

The results of the investigation into vWM as a possible explanation for the contrast in offline and online data for passivization can be summarised in the following points: (1) vWM span correlated with the difference in accuracy between actives and passives, and with participants' accuracy on passive sentences, but not active sentences, or with reading times; (2) passives were processed faster than actives up to the by-phrase, where the pattern reversed and passives were read marginally slower than actives, independent of predicate type; (3) passives were more errorful than actives, independent of predicate type; consequently, no interaction was observed between passivization and predicate type in the online or offline tasks. As per Exp. 2 and 3, the results presented in points (2) and (3) cannot be explained by a speed-accuracy trade-off, given that (1) we analysed accurate trials only and still found passives to be read faster than actives; (2) Reaction Times to comprehension questions were longer for passive than active sentences, and yet subjects made more errors on the former than the latter.

Despite the results of the correlation analysis presented in point (1), the difference in vWM span seems weakly related to the difference in accuracy between passives and actives, given that there was no effect of vWM or interaction between syntax and vWM span in the regression analysis. In line with the literature on vWM and its correlation with linguistic performance (e.g., Caplan et al., 2011; Evans et al., 2014), vWM only correlated with the subjects' offline accuracy, but not online reading time data. This indicates that online and offline tasks depend on partially different processes.

The reading time data presented in point (2) replicate the results observed across Exp. 1, 2 and 3, and thus represent a robust and reliable finding. Passive sentences are read faster than active ones, even when considering accurate trials only. This is incompatible with the *Good Enough* model, the syntactic complexity approach and the frequentist approach to passive sentence processing, as not only we did not find passives to be read slower than actives, we actually found them to be faster than actives. At present the only models compatible with these results are expectation-based and surprisal-based accounts of sentence processing (Hale, 2001; Levy, 2007).

However, these models do not account for the accuracy results from point (3) above. Replicating the results of Exp. 2 and 3 (but not Exp. 1), passives were found to be

more errorful than actives, showing that they are complex in offline comprehension tasks. This result also demonstrates that the difference between Exp. 1&2 and 3 was not simply a matter of fewer items per condition or the higher proportion of eventive predicates. Rather, Exp. 1 and 2 may not have been sufficiently sensitive to show as reliable an effect of passivization on accuracy, since they did not contain only questions targeting theta-roles, which allow for greater sensitivity to the difficulty of passivization in offline accuracy.

While the accuracy results could be accounted for by the *Good Enough* model, the syntactic complexity approach or the frequentist approach to passives processing, these models are insufficient when considering the online reading data reported in the previous paragraph. The data, collectively, point to a need for distinct interpretations of the offline and online data. Some possible interpretations are offered in the General Discussion below.

Finally, both the reading time and accuracy results presented in point (2)-(3) above seem to indicate that passivization and predicate type do not interact in sentence processing, as found in Exp. 3, but contrary to the independent findings of Exp. 1 and 2. This is also in contrast with the data reported in the theoretical literature (e.g., Belletti & Rizzi, 1988; Gehrke & Grillo, 2009) and the findings presented in the acquisition, aphasia and, recently, healthy adults' literature (e.g., Ambridge et al., 2016; Grodzinsky, 1985; Maratsos et al., 1985; O'Brien et al., 2006; Volpato et al., 2015), which found that passivization complexity was modulated by predicate semantics. Some possible accounts for the discrepancy will be considered in the General Discussion below.

Overall, the data collected in Exp. 4 return the same contrasting picture of passives processing that was described in Chapter 1: passives are processed faster than actives online, thus suggesting that they are not more complex to parse or interpret. However, they are more errorful than actives in offline verification tasks, which seem to indicate that they create an additional difficulty with respect to active sentences. These diverging results are mirrored by an overall independence of accuracy and reading time measures as they differentially relate to vWM span. The correlation between vWM span and the difference in accuracy between active and passive sentences further suggests that the difficulty in passive sentences might arise at some stage in post-interpretative processes, possibly related to a memory component (either encoding or retrieval)

required by the task. Additional methodologies, which might help to further understand these results, are considered in more detail in the General discussion below.

2.6 General discussion.

Overall, passives were found to be processed faster than actives across four experiments, thus suggesting that they are not more complex to parse than actives, while they were more errorful than actives in three out of four experiments (Exp. 2-4), indicating that difficulty arises in the offline verification task. Moreover, the passivization complexity in the offline task (i.e., the difference in accuracy between actives and passives, as well as accuracy on passive sentences only) correlated with vWM span.

While the *Good Enough*, syntactic complexity and frequency approaches can account for the offline accuracy results, they cannot account for our robust online findings. Conversely, expectation- and surprisal-based models can offer an adequate explanation for the online, but not offline findings. Collectively, these results suggest that passives are *not* more difficult to parse or interpret than actives, but a difficulty might arise post-interpretatively due to memory demands (either in encoding or retrieval) that specific tasks place on passives. It is thus necessary to consider these two sets of (online and offline) findings separately, in order to reach a comprehensive interpretation of these results.

2.6.1 Passives facilitate processing.

Passives seem to be easier to process than actives, as suggested by the self-paced reading data, in which passives were read significantly faster than actives across multiple regions. These results are incompatible with processing models that predict passives to be more difficult to process than actives do to their non-canonical nature, e.g. the *Good Enough* processing model (Christianson et al., 2001; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016). These results are also incompatible with theories that consider passives to be derived from actives due to an additional syntactic transformation (e.g., movement) and that deem this operation taxing on the parser. Moreover, any account that resorts to frequency of a specific structure in language use to predict its processing complexity, is also incompatible with

these results. Our BNC corpus analysis determined that the predicates tested in our experiments are less frequent in the passive rather than active form in language use, and yet passives were processed faster than actives.

These results are however compatible with surprisal-based and expectation-based models (Hale, 2001; Levy, 2007). These models predict passives to be processed more easily than actives due to their dedicated morphology, which creates expectation for upcoming syntactic categories (i.e., auxiliary, past participle, "by"). After encountering an auxiliary, a verb is more likely to occur than other syntactic categories. Conversely, the subject NP in the active version of the sentence could be followed by several types of syntactic categories (e.g., PP, subordinate clause, AdvP), and hence the verb must compete with these alternative continuations, thus increasing surprisal and lowering expectation with respect to the same verb in the passive version of the sentence. Similarly, after encountering the past participle and "by" in a passive sentence, the only possible continuation is a DP. At this point, the expectation for this specific grammatical structure is so high (and, conversely, surprisal so low once it is actually encountered) that the processing of upcoming words is facilitated relative to the same region in the corresponding active sentence. In the active, the verb can be followed by several sentence continuations (e.g. PP, AdvP), with which the DP must hence compete. The upcoming DP in the active sentence is thus processed with a lower level of expectation and higher level of surprisal with respect to the corresponding DP in the passive sentence.

The results of Exp. 1 further support this interpretation, given that the verb region was read significantly faster in progressive sentences with respect to simple active ones. The presence of an additional morphological cue in the progressive condition with respect to the simple active one (i.e., the auxiliary) increases expectation and lowers surprisal for a verb to follow with respect to any other grammatical category, thus facilitating processing once the verb is encountered.

Marginal evidence for a difficulty effect was found in the self-paced reading data at the post-verbal NP. However, this marginal effect was only found in two out of four experiments (in the other two experiments, the effect at the head of the by-phrase did not even approach significance, p-value>.7), and it is thus not reliable.

Overall, the self-paced reading data consistently suggest that passives are not more difficult to parse or interpret than active sentences.

2.6.2 Passives impair comprehension.

On the other hand, passives generate more difficulty than actives in offline comprehension accuracy. How can we reconcile these online and offline findings? There are two possible answers to this question. The first possibility is that the complexity effect in processing passives occurs in later regions than expected. For example, if the complexity effect arises at the post-verbal NP (or thereafter), it may not be detectable in self-paced reading, given its susceptibility to spill-over effects (Mitchell, 1984). Spill-over effects refer to the fact that the subject presses the button for the next word before having fully interpreted (i.e., integrated) the current word. This seems an unlikely explanation of the data, given the following PP region shows no such effect. However, this region is also contaminated by end-of-sentence effects (Just et al., 1982), so it might be problematic to detect any effect due to the experimental manipulation here. Additionally, some proponents of the *Good Enough* model advocate that revision/integration of the heuristic and algorithmic output should happen quite late in processing, after the interpretation of the entire sentence has been built (comment made by Kiel Christianson to a paper under review). In this case, it might not be possible to detect a revision/integration effect in online measures relying on first reading pass, such as self-paced reading.

Secondly, the online and offline results might contrast due to their recruitment of different processes. While the self-paced reading task exclusively targets sentence parsing and interpretation, the verification task also includes additional post-interpretative processes, e.g., encoding of the target sentence and retrieval of its interpretation in order to answer a comprehension question. The offline accuracy effect could hence derive from these post-interpretative operations interacting with the passivization manipulation, and not necessarily from a complexity in parsing and/or interpreting passives. In support of this second interpretation comes the correlation found between the difference in accuracy between actives and passives and vWM span, and between accuracy on passives only and vWM. This indicates that a selective difficulty in passives could arise at some stage in their post-interpretative processing.

There are several reasons why passives might be selectively affected with respect to actives in these processes: for example, encoding passive sentences might be more demanding than encoding active ones. In particular, it has been proposed that participants create an "active-like" representation of the target sentence (e.g., agent-action-patient; Anderson, 1974) for certain post-interpretative tasks, such as comprehension questions. This strategy may be adopted to help avoid memory decay. However, generating such a representation for passive sentences would require manipulating the representation in memory, thus making the passive more prone to errors than actives. Alternately, retrieval of the interpretation of the target passive sentence might be complicated by the active syntax of the comprehension questions. Given that participants answered questions without the possibility of re-reading the sentence, our offline task involved retrieving a memory of the parse and checking its contents against the interpretation proposed by the parse of the comprehension question. However, in our task, comprehension questions always contained an active form of the verb. Hence, the syntactic form of the experimental sentence and of the comprehension question never coincided. This might have introduced greater interference in passives than actives.

Further support to the idea that complexity in passives might arise from an interaction with the task characteristics comes from studies on aphasic patients, which found that difficulty in processing non-canonical structures (i.e., object relatives, passive sentences) varied with the task used (verification task, self-paced listening; Caplan, Michaud, Hufford and Makris, 2016). New studies are necessary to test this hypothesis, as will be discussed in the next section.

2.6.3 Interaction between passivization and predicate type.

While the independent results of Exp. 1 and 2 suggested that passivization interacts with predicate semantics, the results of Exp. 3 and 4 contradict this hypothesis.

There are several possible explanations for the absence of an effect in Exp. 3 and 4. One might argue that the interaction between predicate and passivization availability has no psycholinguistic counterparts. However, this hypothesis would be refuted by psycholinguistic evidence of such an interaction reported in the aphasia and acquisition

literature (e.g., Grodzinsky, 1985; Maratsos et al., 1985; O'Brien et al., 2006; Volpato et al., 2015).

Alternatively, it might be possible that healthy adults are not sensitive to a complexity in passivizing states. However, this is in contrast with the recent results reported by Ambridge and colleagues (2016), which found an interaction between passivization and predicate type in the healthy adults' performance on offline tasks. However, Ambridge and colleagues (2016) used a different type of offline task, namely grammaticality judgments and forced-choice picture-matching comprehension task, which might be more sensitive to this particular interaction. This possibility will be further investigated in Chapter 3.

2.6.4 Moving beyond the current results.

The next set of experiments will tackle two different problems: (1) identifying the source of the offline accuracy difficulty in passives; (2) investigating whether a possible interaction between passivization and predicate type could be found with a different methodology.

The hypothesis that the verification task poses processing demands that may have interacted with passivization will be investigated in Exp. 5-7, presented in Chapter 3. These three experiments collected an additional offline measure of passive processing complexity – acceptability judgments – which do not involve encoding/retrieval of the target sentence parse in memory and could provide evidence of a late complexity effect in passivization not detectable in self-paced reading. Additionally, Exp. 7 balanced the syntactic form of the comprehension question between active and passive forms to equate the demands on memory across the voice manipulation.

Moreover, acceptability judgments might be more sensitive to detect an interaction between passivization and predicate type, as Ambridge and colleagues (2016) found evidence for such an interaction using this methodology.

The results of the next three experiments (Exp. 5-7) should then allow to ascertain (1) whether the difficulty observed in the offline data of our comprehension tasks was due to a complexity in processing passives or to a difficulty introduced by the nature of the task, and (2) whether passivization and predicate type interact in processing. Consequently, this should help reconciling the online and offline results

towards a more comprehensive and uniform understanding of passive sentences processing.

3 Investigating the offline complexity effect in passive sentence comprehension.

3.1 Introduction.

In the four self-paced reading experiments reported in the previous chapter, passive sentences were consistently found to be read faster and to result in more errors in response to comprehension questions than active ones (Exp. 2-4). Additionally, the final experiment discussed in the previous chapter (Exp. 4) found a correlation between vWM span and the difference in accuracy between actives and passives, as well as a correlation between vWM span and accuracy on passive sentences only (but vWM did not correlate with accuracy on active sentences or accuracy overall). Finally, vWM did not correlate with online processing measures (i.e., reading times). These results indicate that *post-interpretive* demands on memory are greater on passive sentences. They also further confirm that the online facilitation effect and the offline complexity effect found in our previous studies stem from different sources, thus warranting a separate investigation of the two effects.

This chapter aims at investigating the source of the offline difficulty effect of passivization. The investigation across Exp. 5-7 is based on the following reasoning: the consistently faster reading times suggest that passives are not more difficult to parse and interpret than their active counterparts. The offline comprehension verification task employed in the previous experiments does not exclusively target sentence parsing and interpretation, but rather it involves additional post-interpretive processes that may interact with passivization. Additional operations necessary to perform the verification task, include: (1) storing the memory of the parse of the original sentence, (2) retrieving it in order to compare it with the interpretation provided by the comprehension question, (3) computing the truth value of whether the two meanings are equivalent. I will call these operations post-interpretative processes, to indicate that they take place after the initial sentence is parsed and interpreted.

The accuracy effect observed across passivization in the verification task (Exp. 2-4) could hence derive from a difficulty in performing any of the operations described above interacting with passivization, and not necessarily from a complexity in parsing

and/or interpreting passives. There are two reasons why the above processes may generate more difficulty with passives.

Firstly, it has been proposed that participants create an "active-like" representation of the target sentence (e.g., agent-action-patient; Anderson, 1974) for certain post-interpretative tasks, such as comprehension questions. This strategy may be adopted to help avoid memory decay. However, generating such a representation for passive sentences would require manipulating the representation in memory, thus making the passive more prone to errors than actives. This would be consistent with the results of Exp. 4 (presented in Chapter 2).

Secondly, the comprehension questions were always presented in the active form in our previous experiments, which might have facilitated the active condition through priming. The active syntax of the target sentence could have primed participants' processing of the active comprehension question, thus facilitating the task. Conversely, in the case of passive sentences, this might have introduced interference between the syntax of the comprehension question, i.e., active, and the structure of the target sentence stored in memory, i.e., passive. A possible interference in memory retrieval would once more be consistent with the findings reported in Exp. 4 (see Chapter 2).

In alternative to this hypothesis, it is possible that passives are more difficult to parse and interpret than actives, but this complexity arises later than it would be possible for self-paced reading to detect (see General Discussion of Chapter 2 for more details). For example, if the complexity effect arises at the post-verbal NP (or thereafter), it may not be detectable in self-paced reading, given its susceptibility to spill-over effects (Mitchell, 1984). This seems an unlikely explanation of the data, given the following PP region shows no such effect. However, this region is also contaminated by end-of-sentence effects (Just et al., 1982), so it might be problematic to detect any effect due to the experimental manipulation here. Additionally, some proponents of the *Good Enough* model advocate that revision/integration of the heuristic and algorithmic output should happen quite late in processing, after the interpretation of the entire sentence has been built (comment made by Kiel Christianson to a paper under review). In this case, it might not be possible to detect a revision/integration effect in online measures relying on first-pass reading measures, such as self-paced reading.

In order to test whether the passive complexity effect observed in previous verification tasks is due to the fact that passives are more taxing on post-interpretative processes, rather than being more difficult to parse and interpret than active ones, I used an offline task that would not rely on memory encoding or retrieval and that is sensitive to syntactic complexity effects (including possible late effects) – acceptability judgments. In the following, I will describe the value of acceptability judgment to this purpose.

3.1.1 Acceptability judgment task.

Acceptability judgments have been used in theoretical and experimental syntactic research for decades and are generally considered a reliable measure of syntactic difficulty (Chomsky, 1965; Schütze, 1996; Sprouse, 2008). This property derives from the main assumptions of acceptability judgments: acceptability judgments are influenced by the grammaticality of the sentence; the acceptability of a sentence can be consciously perceived and reported by native speakers (Sprouse, 2008). From a psycholinguistic perspective, acceptability judgments are thought to arise as a result of an attempt to understand the sentence presented to the hearer/reader. As such, they are considered to be affected by parsing strategies (Schütze, 1996), as well as by grammar competence. Finally, cognitive resources (e.g., working memory) normally employed in sentence comprehension are involved in the judgment process, thus implying that grammaticality judgments are subject to the same cognitive limitations as other comprehension tasks. However, they are different from comprehension questions in the exact way that we would wish them to be, namely that they do not rely on memory retrieval of the parse of the sentence given that the judgment applies to the presented sentence and does not require processing another one and comparing them.

Given the similarity between acceptability judgments and the verification task from the perspective of processing demands, coupling the two tasks together allows a comparison of two reliable offline measures of processing difficulty. While the latter involves recovering the memory of the parse of the sentence, the former does not. Moreover, the judgment considers the entire sentence parse and interpretation, which could go beyond a possible limitation of self-paced reading in not being able to detect a late complexity effect of passivization.

Hence, by studying passivization with an acceptability judgment task, it should be possible to distinguish 'pure' effects of parsing and/or interpretation difficulty from demands on memory in retrieving the outcome of those processes for the comprehension task.

3.1.2 Interaction between passivization and predicate type.

Our previous experiments found no evidence of an interaction between passivization and predicate type. However, evidence in favour of an interaction, such that processing a verbal passive of states is complex, while parsing a passive of eventive is not, was reported by several psycholinguistic studies (e.g., Ambridge et al., 2016; Grodzinsky, 1995; Maratsos et al., 1985; O'Brien et al., 2006; Volpato et al., 2015), which targeted various populations (healthy adults, children acquiring their first language, aphasic patients) and employed different methodologies (acceptability judgment task, forced-choice picture-matching comprehension task, truth-valued judgment task, picture-sentence matching task, anagram task).

Most relevant to this work is the fact that the acceptability judgment task was previously used in studying the interaction between passivization and predicate semantics in healthy adults (Ambridge et al., 2016). Ambridge and colleagues found that passives were less acceptable than actives overall, and, more importantly, passives of experiencer-theme predicates (i.e., stative) less acceptable than passives of agent-patient predicates (i.e., eventive). However, there are some limitations to this previous study. Firstly, no fillers were used to mask the purpose of the study, which might allow prescriptive grammar to play a larger role in the participants' judgments of passives. Also, there were no ungrammatical sentences, which might artificially inflate differences between sentences when subjects are required to use the whole judgment scale in accordance with the instructions and no anchor for the lowest point (i.e., an ungrammatical item) is provided, thus maximising scale bias across participants (Schütze & Sprouse, 2013).

In improving on these design features (by inserting filler sentences as well as ungrammatical items), the acceptability judgment task is used to further test a possible interaction between passivization and predicate type. In particular, this effect is expected to emerge as passives of states being more difficult to process (i.e., less

acceptable) with respect to their active counterparts but no difference in eventives. Employing this task would also allow to possibly replicate the findings reported by Ambridge and colleagues (2016). Finally, this methodology does not rely on memory retrieval, differently from our previous verification task. This should allow to determine whether the passivization difficulty effect found in the verification task extends to this method – as found by Ambridge and colleagues – or not. This latter option (i.e., passives as acceptable as actives) would indicate that the difficulty effect is due to particular demands on memory components of post-interpretative processes, thus better establishing the source of the offline difficulty effect observed in accuracy.

I will now present Experiments 5-7 which use acceptability judgment alongside comprehension questions to establish that the complexity of passivization is dependent on post-interpretive factors of the verification task. Experiment 7 also balances the syntax of the comprehension question between active and passive to establish that the accuracy effect is not dependent on passivization, but on the priming/interference effect generated by match/mismatch between the syntax of the target sentence and the syntax of the verification question.

3.2 Experiment 5.

The previous experiments (Exp. 2-4, described in Chapter 2) have left two possibilities concerning the complexity of processing passivization: (1) there is an intrinsic complexity in passive sentence processing that is detected with offline comprehension questions, but not with self-paced reading (due to a possible late effect, i.e., arising once the complete interpretation of the VP or of the sentence has been built) or (2) passivization interacts with post-interpretative factors (e.g., passives are more demanding on memory retrieval), and hence the complexity only emerges in comprehension questions (and not self-paced reading). Additionally, the previous experiments have failed to provide evidence in favour of an interaction between predicate type and passivization.

In order to investigate whether passive sentence complexity emerges due to post-interpretive factors rather than simply a late interpretation effect, our 'classic' offline comprehension question task was coupled with an additional offline acceptability judgment task. The acceptability judgement task is affected by the same parsing and interpretation demands that are measured by the verification task. Crucially, however,

the acceptability judgment task differs from the verification task in terms of post-interpretive processes, i.e., subjects are not required to rely on memory retrieval in order to perform the task. Using this task should hence allow to distinguish pure syntactic effects from memory limitations in performing a specific task. Furthermore, the task provides an overall measure of sentence difficulty, which would overcome a possible limitation to identify a complexity effect in self-paced reading.

Finally, given the previous findings of Ambridge et al. (2016), which did find an interaction between predicate type and passivization using the acceptability judgment task, a possible interaction between passivization and predicate type was investigated in Exp. 5 using this same methodology.

The predictions of the present experiment were thus three-fold. If passive sentences are indeed more difficult to process than active ones due to their syntactic structure (i.e., noncanonical argument order), as suggested by the *Good Enough* model, the syntactic complexity approach and the frequentist approach, then passive sentences should be less acceptable than active ones and accuracy to comprehension questions should be lower following a passive rather than an active sentence.

On the other hand, if passive sentences are susceptible to post-interpretive processing demands only (e.g., increased complexity in encoding with respect to actives or memory retrieval failure), passive sentences should be as acceptable as active ones, given that acceptability judgments do not place the same demands on memory as the comprehension verification task. Under this scenario, accuracy to comprehension questions should be lower following a passive rather than an active sentence, given the increased demands on memory processes described above.

Finally, if predicate semantics interacts with passivization, passives of statives should be less acceptable/accurate than their active counterparts, while no effect should be present across voice in eventives.

3.2.1 Methods and Design.

3.2.1.1 Participants.

One-hundred and forty-five British English native speakers were recruited to participate in the study (66 females; average age: 31). A larger sample of participants was recruited given that web-based studies typically generate noisier data, and it is thus necessary to

collect a bigger sample than lab-based studies in order to ensure sufficient power in the analysis (Finley & Penningroth, 2015). They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants were recruited through Prolific Academic (https://www.prolific.ac/) and received payment for their participation. All the participants were informed of the aims and procedures of the experiment and provided informed consent, approved by UCL ethics. None of the participants had previously participated in any of the experiments related to this study (including the pre-norming tests).

3.2.1.2. Stimuli.

The experiment contained 28 sentence sets, identical to the ones used in our previous self-paced reading experiments (Exp. 3 and 4, see Chapter 2) in order to collect comparable data to previous results (see Table 13; for the complete list of items, see Appendix 7.7). Hence, as in our previous experiments, the manipulated factors were syntax (within-subject: active vs. passive) and predicate type (between-subject: eventive vs. statives).

The plausibility of thematic role assignment across syntactic structures was controlled for in a previous pre-norming study (see Chapter 2), which ensured that sentences in the passive form were equally plausible to sentences in the active form.

Eventive	Passive	The guitarist was rejected by the attractive and talented singer in the concert hall next to the Irish pub.
Eventive	Active	The guitarist rejected the attractive and talented singer in the concert hall next to the Irish pub.
Stative	Passive	The guitarist was admired by the attractive and talented singer for keeping the band focused through the whole tour.
Stative	Active	The guitarist admired the attractive and talented singer for keeping the band focused through the whole tour.

Table 13: Example of experimental items contained in Exp. 5.

Half of the sentences were acceptable (i.e., grammatical) and another half unacceptable (i.e., ungrammatical), hence it was necessary to partially modify the fillers

used in previous experiments (Exp. 3-4). Previous experiments contained 60 fillers with various complex syntactic structures (15 actives; 15 passives; 15 sentences with negation; 15 garden-path "while…" constructions). Given that experimental items in Exp. 5 were necessarily always acceptable, we used 14 fillers from our previous experiments (n=14), while the others were made ungrammatical (n=42; see Appendix 7.9 for the complete list of fillers). In order to encourage participants to use the full acceptability scale, we created different levels of ungrammaticality by varying the number of grammatical errors within the fillers (either 1, see (34); or 2, see (35)). Moreover, in order to avoid the possibility that participants could create strategies by only focusing on a specific segment or word, the errors targeted different syntactic rules (agreement, see (36); complementizer, see (37); preposition, see (38)) that were located in different positions within the sentence (subject NP, matrix verb, prepositional phrase).

(34) The ghost haunted the villager who were currently living in the ghost's house on the top of the hill.

(35) The writer spent two year in his chalet in a Alps and wrote his autobiography.

(36) The door-to-door salesperson married his hard-working wife on a rainy days in Portofino during the spring.

(37) The zombie who bit the man was once a compassionate king how had ruled his lands fairly and kindly.

(38) The solo for the symphony orchestra was played by a very talented cellist this year among the St. Pancras church.

3.2.1.3 Procedure.

The items so designed (28 experimental items, 56 fillers) were inserted in an acceptability task implemented in Qualtrics (https://www.qualtrics.com/), a web-based platform. Participants were required to provide demographic data (native language, age, gender), and were then presented with the following instructions:

"[…] you will need to rate the sentence based on how natural it sounds to you in terms of its grammaticality or acceptability. […] So as to be clear, the terms "acceptable"

and “unacceptable” in this case do not have anything to do with the events described in the sentence and how likely these may or may not be to occur in real life. What is important is how natural, well-formed or acceptable you find the sentence itself, i.e., whether you personally think someone could ever say something like that in English. […]”

As repeatedly stressed in the relevant literature (e.g., Schütze, 1996), we tried to delineate a clear difference between plausible and acceptable, as the latter might often be confounded with the former. Participants were presented with a sentence and instructed to rate its acceptability on a scale (in which 1=completely unacceptable, 7=completely acceptable) presented underneath the sentence. After rating the sentence, a new screen appeared containing a comprehension question that required a yes/no selection.

Questions were designed to be half correct and half incorrect. Moreover, questions following both experimental items and fillers could either target theta-roles (see (39)) or other sentential aspects (e.g., time, location, adjective, see (40); for the complete list of comprehension questions, see Appendix 7.10). This is different from Exp. 3 and 4, where questions following experimental items only targeted theta-roles, but it is identical to the procedure of Exp. 1 and 2. Given that Ferreira (2003) argues that passive sentence comprehension is selectively impaired with questions targeting theta-roles, we suggested that the limited number of theta-role questions in Exp. 1 might have prevented us from detecting an offline complexity effect of passivization, which was later detected in Exp. 3 and 4, where questions only targeted theta-roles. Given the increased power of this experiment (128 participants in Exp. 5 vs. 30 in Exp. 1), we decided to go back to our original procedure (questions targeting both theta-roles and other sentential aspects) in order to assess our interpretation of the difference in results between Exp. 1 vs. Exp. 3 and 4.

(39) Did the guitarist cherish the singer?

(40) Did the salesgirl work at H&M?

Participants clicked the preferred answer (Yes or No) presented on the screen. After answering the question, the experiment would go on to the next item.

Randomization was performed by pseudo-randomly dividing the items into 4 lists and then randomly assigning each participant to one list. In total, 84 experimental stimuli (28 experimental items and 56 fillers) were presented to each participant.

3.2.1.4 Data analysis.

Data from 17 participants were excluded due to low overall accuracy to comprehension questions (lower than 75%) and/or because the mean difference in rating between grammatical and ungrammatical sentences was lower than 1 or negative. In other words, we excluded participants with low accuracy and/or who could either not distinguish between grammatical and ungrammatical items or that judged ungrammatical sentences more acceptable than grammatical ones. Such a high exclusion rate (12%) is normal in web-based studies, in which data collection is inherently noisier than in lab experiments. In fact, exclusion rate in web-based studies can vary from 7% to 54% (Finley & Penningroth, 2015). Data from 128 participants were analysed.

The outcome measures were: acceptability rating and accuracy to comprehension questions. The analysis was run using RStudio (https://www.rstudio.com/). Our pre-norming plausibility test (see Chapter 2 for more details) revealed that our eventive items are significantly less plausible than our stative ones (β=.35, t=2.65, p=.008), but critically predicate type does not interact with passivization in plausibility – the contrast of interest. Moreover, a frequency analysis, run on data taken from SUBTLEX-UK (see Chapter 2 for more details), revealed that our eventive verbs are twice as frequent as our stative ones (eventives Zipf frequency: 4.65; statives Zipf frequency: 2.48). Hence, any effect of predicate type would be hard to interpret, given that the two predicate groups differed both in terms of frequency and plausibility, and will not be given further discussion.

3.2.1.4.1 Acceptability data:

The acceptability data were analysed with a linear mixed effects model including syntax and predicate type along with their interaction as fixed effects and both subjects and items as random effects (including both intercepts and slopes). The contrasts used were

passive vs. active [-0.5, 0.5] and eventive vs. stative [-0.5, 0.5]. P-values were determined through treating the t-value as a z-statistic (Barr et al., 2013).

3.2.1.4.2 Comprehension Accuracy data:

The accuracy data were analysed using a mixed effects logistic regression model including the same effects and contrasts used in the analysis of the acceptability data. During discussion of the results, modifications of the random effects structure due to convergence failure will be indicated.

3.2.2 Results.

3.2.2.1 Acceptability data:

Although participants performed the task correctly (average rating of grammatical items: 6.33, average rating of ungrammatical items: 2.92), the analysis of the acceptability data revealed no significant effect. In fact, ratings did not seem to differ across syntax or predicate type, although passives of eventives were rated slightly more acceptable than their active counterparts (passive eventive mean rating: 6.42; active eventive mean rating: 6.29), while no such difference was evident between passives and actives of statives (passive stative mean rating: 6.22; active stative mean rating: 6.23; see Table 14). Hence, passive sentences were not less acceptable than active sentences; on the contrary, a slight trend toward the opposite pattern was found. This suggests that, when using a task that does not require memory retrieval, passives do not demonstrate greater complexity than actives. Finally, in order to exclude possible learning effects that might have confounded our results, we ran and compared two separate analyses on the data collected in the first vs. second half of the sessions[16]. There was no significant difference between the results of the two (first vs. second) analyses.

[16] Data from trials 1-42 were analysed as “first-half” and data from trials 43-84 as “second-half”.

3.2.2.2 Accuracy data:

On the other hand, the analysis of accuracy data revealed a significant effect of syntax (β=-.41, z= -3.59, p<.001[17]) due to lower accuracy in answering questions following a passive than an active sentence, consistent with our previous results (mean passive accuracy: 82.7%; mean active accuracy: 87.6%; see Table 14). Although the difference across statives was larger than across eventives (mean accuracy in eventive active: 86.38%%, eventive passive: 82.7%, stative active: 88.84%, stative passive: 82.59%), no other effect was significant. As reported in the methods section, Ferreira (2003) argues that comprehension of passive sentences is selectively impaired when questions target the theta-roles of the sentence. Our design contained both questions targeting theta-roles and questions targeting other sentential aspect in further verifying this claim. It also provided an opportunity to test whether the difference in results between Exp. 1 vs. 3 and 4 was due to their different proportion of theta-roles question (see Procedure section of this experiment for more details). We separately analysed the participants' performance on theta-roles questions only and the analysis revealed the same significant effect as with the overall analysis: passives were responded to significantly less accurately than active sentences (β=-.83, z=-3.68, p<.001). Moreover, participants' performance on theta-role questions was generally lower than for the other question types (mean accuracy on theta-role questions: 82.2%; mean accuracy on other question types: 87%). On the contrary, syntax did not affect accuracy on questions targeting other sentential aspects. Hence, the difference in accuracy across passive and active sentences was driven by answers to questions targeting theta-roles. In this case, passive sentences were more difficult to understand than active ones, independent of the predicate type. This gives further support to the idea that a complexity effect in passivization could not be observed in Exp. 1, due to the limited number of theta-roles questions, which decreased the power of the passivization complexity effect and hence our ability to detect it.

Finally, the same first- vs. second-half analysis described for the rating data was performed, showing that the syntax effect was only marginal in the first half of the

[17] For all accuracy analyses, the model converged only with the Syntax slope for the subject and both intercepts.

experiment (β=-.31, z=-1.74, p=.08) and became significant in the second half of the experiment (β=-.48, z=-3.03, p=.002), indicating that passives were more errorful than actives. No other effect was significant.

		Passive	Active
Eventive	Acceptability	6.42 (±.19)	6.29 (±.22)
	Accuracy	82.7% (±7.73)	86.38% (±7)
Stative	Acceptability	6.22 (±.26)	6.23 (±.26)
	Accuracy	82.59% (±7.74)	88.84% (±6.43)

Table 14: Average acceptability and accuracy in Exp. 5.

3.2.3 Discussion.

The present study found passive sentences to be judged as acceptable as active ones. Overall, the data seem to indicate that passives are not inherently more difficult to process than actives. On the other hand, accuracy in comprehension questions following a passive sentence was lower than following an active one, independent of the predicate type, replicating what was found in the previous experiments (Exp. 3-4, described in Chapter 2). Similarly, there was no interaction between passivization and predicate type in the rating data.

The overall pattern of results cannot be accounted for by the *Good Enough* model, the syntactic complexity approach and the frequentist approach. These accounts would predict passive sentences to be judged less acceptable than active ones, which is not what was found. Moreover, the acceptability data also exclude a possible late passivization effect (not detected in previous self-paced reading studies due to limitations in the methodology). Given that acceptability data are based on the entire sentence parse and interpretation, they are sensitive to any complexity effect, independent of whether it occurs during parsing or after the entire sentence interpretation has been built. The absence of such an effect in the acceptability data indicates that passives are not more complex to parse or interpret than actives.

Conversely, the overall results are more compatible with the hypothesis that passives comprehension is selectively affected by memory demands which the verification task places on passive sentences at encoding or retrieval. In particular, the

additional transformation required to store passive sentences in an active-like form (Anderson, 1974), might have made participants more prone to errors in passives than actives due to manipulations in memory. Alternatively, the passive sentence parse and the active comprehension question might have generated interference in memory when establishing whether the meanings of the representations (i.e., the one retrieved from memory and the one given by the comprehension question) map onto one another. The hypothesis that passives are not more difficult to parse and interpret than actives, but their comprehension performance is selectively affected by memory demands in the verification task would explain the contradictory set of results observed in the present experiment. Indeed, these memory demands, selective to passives in the verification task, should not be present in the acceptability task as participants rated the sentence while it was still visible on the screen, and hence did not have to rely on the memory of the parse.

It is important to note that the difference in accuracy between actives and passives only emerged in questions targeting theta-roles and not other sentential aspects. Taken alone, these data would be compatible with the *Good Enough* model predictions regarding passive sentences processing complexity, which argues that comprehension of passive sentences is selectively impaired with questions targeting theta-role assignment. However, other data collected in this experiment, i.e., the acceptability data, as well as in previous experiments (Exp. 1-4, see Chapter 2), i.e., the self-paced reading data, argue against this model. Alternatively, these results suggest that theta-role assignment is the primary locus of encoding or retrieval failure/interference, which is expected given that all other aspects of our experimental sentences were identical across voice (e.g., type of action, location, time).

Our acceptability data contrast with the results obtained by Ambridge and colleagues (2016), which found passives to be less acceptable than actives across predicate types. However, I argue that several potential issues in the design of the study by Ambridge and colleagues (2016) might have selectively affected their results. First of all, their study did not contain any filler items to mask the purpose of the experiment. This additionally implies that the study did not contain any ungrammatical sentences (beyond unpassivizable passives), thus possibly artificially generating a significant effect. Moreover, participants were presented with both the active and passive version of the

same sentence, which shared the same NPs pairs (e.g., the same subject would see both sentence (41) and (42)). Given that items were presented all together in a spreadsheet, subjects could re-read previous sentences and directly compare the active and passive versions of the same item. Consequently, judgments might have been biased by prescriptive grammar rules. The interaction with predicate semantics seems more difficult to account for under these explanations. However, a closer look at this interaction shows that when unpassivizable passives (e.g., to sleep, to cost) are included (Ambridge et al., 2016: Experiment 2) they appear to solely account for the interaction with predicate semantics and there is little if any difference between agent-patient and experiencer-theme predicates. When non-passivizable passives are excluded and a subset of each of the three predicates (agent-patient, e.g., kick; experiencer-theme, e.g., fear; theme-experiencer, e.g., to frighten) included (Ambridge et al., 2016: Experiment 3), there is nonetheless an interaction, but criteria for predicate exclusion are not clear and do not seem accounted for based on predicate class or semantic affectedness of the predicate on the internal argument. Finally, Ambridge and colleagues did not include any participants' exclusion criteria based on participants' understanding of the task or the quality of their answers. Hence, it is possible that the interaction effect reported by Ambridge and colleagues was generated by the absence of ungrammatical fillers, which might have introduced a scale bias, or by the use of specific predicates, whose selection was not specified in Ambridge et al.'s paper, or by a combination of the two factors.

(41) Wendy kicked Bob.

(42) Bob was kicked by Wendy.

Alternatively, an interaction in acceptability judgments might not have emerged in Exp. 5 due to the characteristics of the sentence final phrases, which differed across the two predicates due to the nature of the predicates: the eventives contained locative PPs, while the statives contained for-clauses. It may be that the for-clause completion helps support the interpretation of a stative passive. Moreover, the for-clause introduced a confound across the passivization contrast. In the passive condition, the for-clauses contained a non-local dependency between the subject of the matrix clause and the null subject of the for-clause (e.g., "Mary was admired by John for singing beautifully" contains a dependency between "Mary" and the subject of "singing"), while

in the active condition, a local one, between the object of the matrix clause and the subject of the for-clause (e.g., “Mary admired John for singing beautifully” contains a dependency between “John” and the subject of “singing”).

In the next two experiments (Exp. 6 and 7), these concerns were eliminated by taking away the sentence completions, which are not critical for acceptability judgment as they are for self-paced reading in excluding end-of-sentence effects (Just et al., 1982). This should allow for a fair test of an interaction between predicate type and syntax in the acceptability data.

3.3 Experiment 6.

The previous experiment showed that passives are equally acceptable as actives and passivization does not interact with predicate semantics in acceptability. However, the for-clause may have assisted stative passive interpretation and/or introduced a confound unrelated to the experimental manipulation.

In order to investigate this possibility, a second web-based acceptability study was designed, Exp. 6, identical to the previous one, but eliminating the sentence completions in the experimental stimuli.

3.3.1 Methods and Design.

3.3.1.1 Participants.

Ninety-nine British English native speakers were recruited to participate in the study (52 females; average age: 32). They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants recruitment was the same as in Experiment 5.

3.3.1.2 Stimuli.

The experiment contained 28 sentence sets, identical to the ones used in Exp. 5, but without sentence completions, as specified in the Discussion section of Exp. 5 (see Table 15; for the complete list of items, see Appendix 7.11).

Eventive	Passive	The guitarist was rejected by the attractive and talented singer.
Eventive	Active	The guitarist rejected the attractive and talented singer.
Stative	Passive	The guitarist was admired by the attractive and talented singer.
Stative	Active	The guitarist admired the attractive and talented singer.

Table 15: Example of experimental items in Exp. 6.

Given that our experimental items were shortened going from Exp. 5 to 6, the fillers used in Exp. 5 were shortened in order to reach an equal amount of words per sentence (i.e., between 11 and 13 words per sentence, as in the 2 versions of our experimental items – 11 in the active and 13 in the passive; see Appendix 7.12 for the list of fillers). Everything else was kept identical to Exp. 5.

3.3.1.3 Procedure.

The procedure was identical to Exp. 5, except the comprehension questions. Contrary to Exp. 5, which contained different types of comprehension questions, Exp. 6 only contained comprehension questions targeting theta-roles since this was shown to be the critical type to observe a passivization complexity effect in accuracy. To avoid creating a bias in attention towards specific parts of the sentence (in this case, theta-roles), comprehension questions following fillers targeted other sentential aspects, as well as theta-roles (for the complete list of comprehension questions, see Appendix 7.8). In total, 84 experimental stimuli (28 experimental items and 56 fillers) were presented to each participant.

3.3.1.4 Data analysis.

Data from 3 participants were excluded due to low overall accuracy to comprehension questions (lower than 75%) and/or because the mean difference in rating between

grammatical and ungrammatical sentences was lower than 1 or negative. Hence, data from 96 participants were analysed.

The outcome measures were: acceptability rating and accuracy to comprehension questions. The analysis was run using RStudio (https://www.rstudio.com/).

The data analysis was the same as in Experiment 5.

As mentioned in Exp. 5, any effect of predicate type would be hard to interpret, given that the two predicate groups differed both in terms of frequency and plausibility.

3.3.2 Results.

3.3.2.1 Acceptability data:

As per Exp. 5, participants performed the task correctly (average rating of grammatical items: 6.68; average rating of ungrammatical items: 2.5). The analysis of the rating data (see Table 16Table 16: Average acceptability and accuracy in Exp. 6.) revealed no significant effect of syntax. Overall, only a slight difference was observed in acceptability between active and passive sentences (mean acceptability active: 6.69; mean acceptability passive: 6.66). Finally, a marginal interaction emerged between syntax and predicate type (marginal effect in overall analysis: β=-.13, t=-1.71, p=.09), in the expected direction, i.e., passivized states were less acceptable than their active counterpart while no such difference (actually, the opposite) emerged in eventives (mean acceptability eventive active: 6.62, eventive passive: 6.65; mean acceptability stative active: 6.76, stative passive: 6.67; marginal effect of syntax in statives: β=-.09, t=-1.77, p=.08). In order to exclude possible learning effects that might have confounded our results, we ran and compared two separate analyses on the data collected in the first vs. second half of the sessions[18]. In the first half, the interaction between syntax and predicate type was actually significant (β=-.23, t=-2.4, p=.02), as was significant the syntax effect in stative sentences only (β=-.14, t=-2.3, p=.02), indicating that passives of states were rated less acceptable than actives of states, while no such effect was found

[18] Data from trials 1-42 were analysed as “first-half” and data from trials 43-84 as “second-half”.

in eventive sentences. The interaction disappeared in the second half of the experiment, and no other effect was significant.

3.3.2.2 Accuracy data:

The analysis of the accuracy data[19] revealed a significant effect of syntax (β=-.89, z=-5.77, p<.001) but no interaction or any other effect. Passive sentences were understood less accurately than active sentences, independent of the predicate type (mean accuracy in eventive active: 95.24%, eventive passive: 90.63%, stative active: 94.2%, stative passive: 86.31%). It is nonetheless worth noting that the difference between actives and passives was almost double in statives with respect to eventives (see Table 16). The same first vs. second half analysis reported for the rating data was conducted on the accuracy data. There was no difference between the two analyses or between the overall analysis and the half-analyses.

		Passive	Active
Eventive	Acceptability	6.65 (±.15)	6.62 (±.16)
	Accuracy	90.63% (±5.95)	95.24% (±4.35)
Stative	Acceptability	6.67 (±.17)	6.76 (±.14)
	Accuracy	86.31% (±7.02)	94.2% (±4.78)

Table 16: Average acceptability and accuracy in Exp. 6.

3.3.3 Discussion.

The results collected in this experiment can be summarised in the following two points: (1) passives of eventives were as acceptable as their active counterparts, while passives were found to be less acceptable than actives in statives, albeit marginally; (2) passives were more errorful than actives independent of predicate type.

The acceptability data collected in this experiment, coupled with the self-paced reading data presented in Chapter 2, suggest that passive sentences are not inherently more difficult to process than active sentences. This is not compatible with the *Good Enough* model, the syntactic complexity approach and the frequentist approach,

[19] For all the accuracy analyses, the model only converged with random intercepts and not slopes.

discussed in Chapter 1, which predict passives to be more difficult to process than actives across the board. Moreover, this is in contrast with the results reported by Ambridge and colleagues (2016), who found passives to be less acceptable than actives across predicate types. However, potential issues in the design of Ambridge and colleagues' study, described in more details in the Discussion section of Exp. 5, suggest that any passivization data reported in their study might be conflated by specific strategy/bias in performing the task.

Moreover, a marginal interaction between passivization and predicate type in acceptability emerged in the expected direction: passivization was not significant in eventives, i.e., passives and actives of eventives were equally acceptable, while a marginal difference emerged in statives, with passives being less acceptable than actives. Although the effect was not significant, this is in line with my predictions, and with the existing theoretical and psycholinguistics literature on the topic (Ambridge et al., 2016; Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Grodzinsky, 1995; Maratsos et al., 1985; O'Brien et al., 2006; Snyder & Hyams, 2015; Volpato et al., 2015).

On the other hand, the acceptability results are at odds with the accuracy data (both the ones collected in this experiment and the ones collected in Exp. 2-4, presented in Chapter 2, and Exp. 5) which indicate that participants made more errors when answering questions regarding a passive sentence rather than an active one, independent of predicate type, although a numerically bigger difference was observed across statives than eventives in both Exp. 5 and 6.

As already argued in the Discussion section of Exp. 5, these contrasting results cannot be accounted for by models that predict passive sentences to be inherently more complex than active ones. Instead, these contrasting findings could be reconciled by hypothesizing that passive sentences are not harder to parse and interpret than active ones. Rather, they are selectively susceptible to post-interpretative processing demands, such as encoding and retrieval of the sentence interpretation, to which active sentences are not susceptible. This is supported by finding no effect of passivization in tasks that do not require encoding and retrieving the interpretation of the target sentence (e.g., acceptability judgment). A complexity effect of passivization only arises when the task involves encoding and retrieval of the target sentence interpretation (e.g.,

comprehension questions). This suggests that a possible failure in memory encoding or retrieval does contribute to the complexity found in passive sentences comprehension.

Relevant to post-interpretive memory demands is the fact that the syntax of the comprehension question was always in the active voice (e.g., “Did the guitarist reject the singer?”) independent of whether the target sentence had an active or passive syntax. The active syntax of the target sentence could have primed participants’ processing of the active comprehension question, thus facilitating the task for active sentences. On the other hand, in the case of passive sentences, the active syntax of the question might have introduced interference between the active syntax of the comprehension question and the passive structure of the target sentence stored in memory.

The contribution of the comprehension question syntax to memory retrieval was investigated in Exp. 7. This experiment replicated Exp. 6 with a crucial modification to the comprehension question design. The syntactic form of the comprehension question was balanced with respect to the syntax of the sentence: half of the active sentences were followed by active questions and half by passive questions; half of the passive sentences were followed by passive questions and half by active questions.

This new design should allow to determine whether the difficulty in passive sentence comprehension was caused by the discrepancy between the syntactic form of the retrieved memory of the parse and the syntactic form of the question.

Moreover, by equating across voice the mis/match between syntax of the comprehension question and syntax of the target sentence, thus eliminating the interference bias affecting passives comprehension, it should be possible to determine whether the passivization difficulty observed in previous experiments mainly derived from a difficulty in encoding or retrieval. Given that it must be assumed that participants will encode the target sentences in the next experiment as in previous ones, a difficulty in encoding, if indeed it affects passives comprehension, should appear in the next experiment as well. On the contrary, should the passivization complexity disappear once the interference bias at retrieval is equated across voice, then a failure in retrieval should be seen as the main contributor to the passivization complexity observed in previous studies, and a difficulty in encoding should be discarded as a possible explanation of this complexity.

3.4 Experiment 7.

Both the acceptability data and the self-paced reading data from the previous experiments provide convincing evidence that passive sentences are not harder to parse or interpret than active ones. Rather there is something about the comprehension question task that gives rise to difficulty in the presence of a passive sentence. I have proposed that this difficulty arises either during the encoding or the retrieval phase in post-interpretative processes. In this experiment, I will explore the second possibility, i.e., that passivization complexity derives from memory interference, or alternatively phrased, the lack of syntactic priming when provided with an active comprehension question to a passive sentence. This hypothesis was investigated by balancing the syntactic form of the comprehension questions.

I expect to replicate the finding of a (marginal) interaction in acceptability ratings between passivization and predicate type in the expected direction, i.e., increased complexity in passives of states with respect to their active counterparts, but no difference across voice in eventives, as predicted in the theoretical literature and supported by previous psycholinguistic studies (Ambridge et al., 2016; Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Grodzinsky, 1995; Maratsos et al., 1985; O'Brien et al., 2006; Snyder & Hyams, 2015; Volpato et al., 2015). Additionally, I expect that accuracy will demonstrate no difference across the voice manipulation once the comprehension question voice is balanced across the experimental manipulation. This finding should also show that the passivization complexity in previous studies was not due to a difficulty in encoding the representation of the passive sentence, but rather to interference in memory retrieval due to the mismatch between the syntax of the comprehension question and the syntax of the target sentence.

3.4.1 Methods and Design.

3.4.1.1 Participants.

A hundred and four British English native speakers were recruited to participate in the study (53 females; average age: 30). They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants recruitment was identical to recruitment in Exp. 5 and 6.

3.4.1.2 Stimuli.

The experiment contained 28 sentence sets, identical to the ones used in Exp. 6 (see Table 17; for the complete list of experimental items see Appendix 7.11).

Eventive	Passive	The guitarist was rejected by the attractive and talented singer.
Eventive	Active	The guitarist rejected the attractive and talented singer.
Stative	Passive	The guitarist was admired by the attractive and talented singer.
Stative	Active	The guitarist admired the attractive and talented singer.

Table 17: Example of experimental items contained in Exp. 7.

3.4.1.3 Procedure.

The procedure was identical to Exp. 6, apart from the syntactic form of the comprehension questions. As previously mentioned, questions following experimental items were balanced between an active and passive form (see examples in (43) and (44); for the complete list of comprehension questions see Appendix 7.13). The syntactic form was balanced across active and passive experimental sentences, i.e., half of the active sentences were followed by active questions, half by passive questions, while half of the passive sentences were followed by passive questions, half by active questions. The response types were also balanced between "Yes" and "No". In addition, we inserted 10 passive questions following filler sentences. Everything else was kept identical to Exp. 6.

(43) Did the guitarist reject the singer?

(44) Was the singer rejected by the guitarist?

In total, 84 experimental stimuli (28 experimental items and 56 fillers) were presented to each participant.

3.4.1.4 Data analysis.

Data from 8 participants were excluded due to the same exclusion criteria that were applied in Exp. 5 and 6. Hence, data from 96 participants were analysed.

The outcome measures were: acceptability rating and accuracy to comprehension questions. The analysis was run using RStudio (https://www.rstudio.com/). The acceptability data and accuracy data were analysed using the same procedure as in Exp. 5 and 6.

3.4.2 Results.

3.4.2.1 Acceptability data:

Although participants performed the task correctly (average rating of grammatical items: 6.65; average rating of ungrammatical items: 2.53), the analysis of the rating data revealed no significant effect (mean rating eventive active: 6.74; eventive passive: 6.74; stative active: 6.54; stative passive: 6.55; see Table 18). Not even a numerical difference was found between active and passive sentences (mean acceptability active: 6.64; mean acceptability passive: 6.64). In order to exclude possible learning effects that might have confounded our results, we ran and compared two separate analyses on the data collected in the first vs. second half of the sessions[20]. No effect was found. Overall, these data seem to indicate that passive sentences are not more difficult to process than active sentences, independent of the predicate type.

3.4.2.2 Accuracy data:

The analysis of the accuracy data[21] (see Table 18) revealed no effect of syntax (mean passive accuracy: 88.84%; mean active accuracy: 91.67%; β=-.08, z=-.34, p=.73). There was a marginal interaction between syntax and predicate type (β=-.47, z=-1.74, p=.08). Since this was compatible with our hypothesis (of an interaction between passivization and predicate type), we broke down the interaction, which revealed a significant effect of syntax in statives only (β=-.43, z=-2.12, p=.03) with more errors on passives than actives (mean accuracy eventive active: 92.11%; eventive passive: 91.67%; stative active: 91.22%; stative passive: 86.01%). Finally, the same first vs. second half analysis reported for the rating data was conducted on the accuracy data. In the first half, we

[20] Data from trials 1-42 were analysed as “first-half” and data from trials 43-84 as “second-half”.

[21] For all accuracy analyses, the model only converged with the Syntax slope and intercepts in the random effects structure.

observed a significant effect of syntax, driven by the difference within statives (mean accuracy eventive active: 92.7%; eventive passive: 90.3%; stative active: 91.3%; stative passive: 86.2%; *β*=-.53, *z*=-2.41, *p*=.02), as well as a marginal interaction between syntax and predicate type (*β*=-.8, *z*=-1.81, *p*=.07). When further analysed, the interaction revealed a significant effect of syntax in statives only (*β*=-.93, *z*=-3.1, *p*=.002). No other effect was significant. Finally, no effect was significant in the second half of the sessions. Overall, these data indicate that questions following a passive sentence containing a stative predicate are more difficult to interpret than questions following its active counterpart, while no such difference is present in sentences containing eventive predicates.

		Passive	Active
Eventive	Acceptability	6.74 (±.11)	6.74 (±.11)
	Accuracy	91.67% (±2.82)	92.11% (±2.75)
Stative	Acceptability	6.55 (±.13)	6.54 (±.13)
	Accuracy	86.01% (±3.54)	91.22% (±2.89)

Table 18: Average acceptability and accuracy in Exp. 7.

3.4.2.3 Acceptability Data across Exp. 5, 6 and 7:

Our acceptability results varied across experiments. In order to investigate whether a general trend could be observed across experiments, we decided to collapse all data in a single analysis. This analysis found no significant effect of syntax or predicate type, confirming that there was no difference across voice or across predicate type in acceptability. More interestingly, a marginal interaction between passivization and predicate type emerged (*β*=-.09, *t*=-1.85, *p*=.065). When broken down, this interaction revealed a significant effect of syntax in eventives, due to the fact that passives of eventives were rated more acceptable than their active counterparts (*β*=.06, *t*=2.21, *p*=.03), while there was no significant effect in statives, contrary to our original predictions and to the findings reported in the previous psycholinguistic literature (Ambridge et al., 2016; Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Grodzinsky, 1995; Maratsos et al., 1985; O'Brien et al., 2006; Snyder & Hyams, 2015; Volpato et al., 2015).

3.4.2.4 Accuracy Data across Exp. 5, 6 and 7:

Across Exp. 5, 6 and 7, we consistently observed a larger difference in accuracy across voice in statives with respect to eventives. However, a marginal interaction only emerged in Exp. 7. We thus decided to collapse data across Exp. 5, 6 and 7[22] to determine whether the marginal interaction in Exp. 7 would become significant with sufficient power, in line with the trend across experiments. When collapsing across experiments, we found a significant effect of syntax (β=-.57, z=-4.41, p<.001), due to the fact that passives were less accurate than actives. More interestingly, we found a significant interaction between syntax and predicate type (β=-.31, z=-2.2, p=.03). When broken down, this interaction revealed a significant effect of syntax in both eventives (β=-.41, z=-2.97, p=.003) and statives (β=-.72, z=-5.27, p<.001), although the effect was significantly larger (i.e., double) in statives.

These results confirm that the interaction found in Exp. 7 is not simply a spurious effect, but it is in line with a general trend that was partially masked in previous experiments by a task artefact affecting memory retrieval.

3.4.3 Discussion.

The present study manipulated the sentence syntax (active, passive) and predicate type (event, state) and collected acceptability ratings alongside comprehension accuracy. The aim was to determine whether the low accuracy in passive sentence comprehension was due to a possible task artefact introduced by the syntactic form of the comprehension question, which introduced an additional complexity in retrieval for passives only. Moreover, Exp. 7 aimed at replicating the findings of Exp. 6, where a (marginal) interaction between passivization and predicate type in the acceptability data was observed.

Unlike Experiments 2-6, no effect of syntactic structure on accuracy was observed once the syntax of the comprehension questions was balanced. The absence of a main effect of syntax is consistent with the hypothesis that a mismatching syntax in the comprehension question introduces a task artefact in the accuracy data. In previous

[22] We did not include data from previous experiments (Exp. 1-4), given that we used different methodologies (self-paced reading vs. acceptability) and procedures (lab based vs. web-based).

studies, the active syntax of the sentence might have primed an active structure, thus facilitating processing of the comprehension question in the active condition and/or introduced interference with the syntax of the target sentence in the passive condition.

The fact that passivization complexity disappeared once interference in retrieval was balanced across voice suggests that a difficulty in passive comprehension did not emerge at the encoding stage in previous experiments, but rather at the retrieval stage. Given that it must be assumed that participants encoded the target sentences in this experiment as in previous ones, if encoding the representation of a passive sentence introduces an additional difficulty with respect to encoding the representation of an active one, the complexity in comprehension of passive sentences with respect to active ones should have been observed in this experiment as well. Instead, the absence of a complexity effect suggests that encoding the representation of a passive sentence is not complex.

A (marginal) interaction between syntax and predicate type was also observed in the accuracy data, due to a significant effect of syntax in statives but not eventives. This is compatible with the hypothesis of an increased difficulty in passivization of stative, but not eventive, predicates. While eventives consistently deliver verbal passives across languages, including English, passivized states introduce an ambiguity between the preferred adjectival interpretation and the less frequent verbal one, which can only be resolved at the by-phrase. Moreover, it is argued in the linguistic literature that states can only be passivized if their meaning can be coerced into an event subsequent to a state (Gehrke & Grillo, 2009). Hence, the ambiguity of interpretation of stative passives and the required coercion of their meaning suggest an increased difficulty in parsing with respect to eventives. Evidence for this was never observed in the previous accuracy data (Exp. 3-6, described in this chapter and in Chapter 2) at least not to significance. It is important to consider why the interaction arises (significantly) in this experiment only. Perhaps it is a spurious effect, given that it only emerged in this study and it is only marginal. However, a trend in the same direction, i.e., increased difficulty in statives passives compared to eventive passives, clearly emerged in both Exp. 5 and 6. The additional analysis that collapsed data across Exp. 5-7 further supports this interpretation. This analysis found a significant interaction effect between syntax and predicate type, as the size of the accuracy difference across passivization was

double in statives with respect to eventives, indicating that the effect is not spurious but perhaps previously underpowered. The ability to detect an interaction would have been further exasperated by the confound of a syntactic mismatch between target sentence and comprehension question in passives only. Given that the syntactic form of the comprehension question was only controlled in the last experiment, this confound could have contributed to the (in)ability to observe an interaction in all the previous experiments. These results suggest that the interaction effect in Exp. 7 is not a spurious result, but rather indicates that passivization complexity varies with predicate type, although the interaction can be (partially) masked by characteristics of the task (i.e., syntax of the comprehension question) that affect memory retrieval.

In the acceptability task, passive sentences were found to be as acceptable as active sentences, independent of the predicate type, replicating the results of Exp. 5, but not Exp. 6. This is contrary to what we hypothesized based on the theoretical literature and previous reports in the psycholinguistic literature (Ambidge et al., 2016; Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Grodzinsky, 1995; Maratsos et al., 1985; O'Brien et al., 2006; Snyder & Hyams, 2015; Volpato et al., 2015). More importantly, these results are in direct contrast with the data collected by Ambridge and colleagues (2016), who found both that passives were less acceptable than actives overall and that passivization interacted with predicate type in the expected direction, i.e., passives of experiencer-theme predicates were less acceptable than passives of agent-patient predicates. Potential explanations for these results being discrepant with Ambridge et al. have been considered in the discussion of Exp. 5.

While the accuracy data showed evidence of an interaction between passivization and predicate type, the acceptability data did not. This difference across methodologies could be due to a differential sensitivity of acceptability to coercion and revision effects, as well as possible characteristics of the experimental design that might have prevented us from observing such an interaction effect. These reasons will be explored in the General Discussion section.

Nonetheless, the acceptability results are incompatible with theories that predict that passive sentence processing is more difficult than active sentence processing. On the other hand, they are compatible with models that suggest that passive sentences are not harder to process than active sentences.

3.5 General discussion

Overall, across three joint acceptability and comprehension question experiments (Exp. 5-7) passives were found to be as acceptable as actives sentences, indicating that they are not more complex to process than actives. Moreover, passives were more errorful than actives when the syntax of the comprehension questions was always in the active form, thus introducing interference in memory retrieval for passive sentences only. When the syntax of the comprehension questions was balanced for voice, passives were understood as accurately as actives. Furthermore, under such conditions, a marginal interaction between passivization and predicate type emerged in the expected direction, i.e., passives of statives were more difficult than their active counterparts, while no difference emerged across voice in eventives. This result was further supported by an analysis that collapsed the accuracy data from all three acceptability experiments and demonstrated an interaction at significance. Collectively, the data further confirm that passives are not more complex than actives to parse and interpret, rather the task characteristics created interference at retrieval for passive sentences only, introducing this increased difficulty. The following paragraphs discuss each of these main findings in more detail before considering further questions that they raise about passive sentence processing.

3.5.1 Passives are not more complex than actives.

The acceptability data reported in Exp. 5-7 consistently found passive sentences to be as acceptable, if not numerically more acceptable, than active sentences. These results are compatible with models that suggest that passive sentences are not harder to process than active sentences thanks to the surface morphological cues that increase expectation for a certain upcoming grammatical category, i.e., expectation-based and surprisal-based models (Hale, 2001; Levy, 2007). In the self-paced reading data (Exp. 1-4 presented in Chapter 2), the presence of these cues might have reduced surprisal and increased expectation for words in the passive structure compared to the active one, thus translating in a facilitation effect for passives across several regions.

On the contrary, the facilitation effect found in the self-paced reading data was not (reliably) replicated in the acceptability data. This difference in results could be

ascribed to the fact that self-paced reading reflects moment-by-moment processing. On the contrary, acceptability judgments are formed on the grammaticality of the sentence in its entirety. It may be the case that the increased expectations for an upcoming word category do not factor into acceptability judgments of the entire sentence or are balanced out by processing additional functional words on which the expectations rest.

Nonetheless, the acceptability data exclude a possible late passivization effect (not detected in previous self-paced reading studies due to limitations in the methodology). Given that acceptability data are based on the entire sentence parse and interpretation, they are sensitive to any complexity effect, independent of whether it occurs during parsing or after the entire sentence interpretation has been built. The absence of such an effect in the acceptability data across experiments indicates that passives are not more complex to parse or interpret than actives.

The data from the seven experiments presented thus far require a parsing model accounting for passives not being more difficult than actives (see Chapter 5 for a more detailed description of such an account).

3.5.2 Passives can be more prone than actives to failure in memory retrieval.

Although passive sentences were found to be as acceptable as actives, they were also consistently found to be more errorful than actives in comprehension, both in the experiments reported in the current and previous chapter (Exp. 2-6), as well as those in the published literature (Ferreira, 2003). Since the acceptability judgment task does not rely on memory of the sentence interpretation, as verification questions do, the combination of the acceptability and accuracy results suggests that memory demands of post-interpretative processes are greater for the passive condition. Lower accuracy on passives could be explained in two different ways that rely on post-interpretive processes of encoding or retrieving the target sentence. Firstly, it has been proposed that participants create an “active-like” representation of the target sentence (e.g., agent-action-patient; Anderson, 1974) during post-interpretative tasks, such as comprehension questions. This strategy would be adopted to avoid memory decay. However, generating such a representation for passive sentences would require an

additional transformation with respect to active sentences, thus possibly making the former more prone to errors.

Alternatively, the active comprehension question could generate interference during retrieval of the target passive sentence between the syntax of the comprehension question, i.e., active, and the structure of the target sentence stored in memory, i.e., passive. In a similar way, the active syntax of the target sentence could have primed participants' processing of the active comprehension question, thus facilitating the task.

In distinguishing between the two explanations, Exp. 7 balanced the syntax of the comprehension questions thus equating retrieval difficulty across the syntax conditions, while any difficulty in memory encoding for the passive should still present. This experiment found that the main effect of syntax on comprehension accuracy disappeared. This suggests that encoding a target passive sentence is not more complex than encoding a target active one. On the contrary, it was interference at retrieval that caused lower accuracy on passives than actives across experiments (Exp. 2-6). Once the cause of this interference (the mismatching syntax of the comprehension questions) was controlled for, the passivization complexity effect on accuracy disappeared.

3.5.3 Passivization interacts with predicate semantics.

All three experiments manipulated the predicate type to investigate a possible interaction between predicate semantics and passivization. Passivization was expected to interact with predicate semantics in the following way: passives of states should be more difficult to process than actives of states, while less or no difference should be observed in eventives. This was motivated by previous theoretical and psycholinguistic studies which identified multiple sources of difficulty with stative passives including a temporary ambiguity between verbal vs. adjectival interpretation, and coercion into an eventive interpretation (Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Snyder & Hyams, 2015).

3.5.3.1 The interaction in the accuracy data.

No evidence in support of this interaction was found in Exp. 5 and 6, while an effect appeared in Exp. 7 in the expected direction. Critically, while Exp. 5 and 6 only contained questions in the active form, thus introducing priming when an active sentence was followed by an active question but possibly interference when a passive sentence was followed by an active question, Exp. 7 balanced the syntax of the questions with respect to the syntax of the targeted sentences. As already reported, this difference in design eliminated the complexity effect insofar observed in passive sentences comprehension. At the same time, the hypothesized interaction between predicate type and passivization emerged.

Two possible explanations can be offered for this effect. The first option would be to attribute this effect to an interplay between factors that independently affect passive sentence comprehension, i.e., the syntax of the question and the interaction between predicate type and semantics. In this case, it would be possible that the confound introduced by the syntax of the question had insofar masked the effect of an interaction between predicate type and passivization. Given that the syntactic form of the comprehension question was only manipulated in the last experiment, this confound would have prevented us from observing an interaction in all the previous experiments.

Alternatively, we could interpret this result as a spurious effect. After all, predicate type was manipulated together with syntax across five experiments (three are reported in this chapter, while the other two in Chapter 2) and was only observed in one. However, the trend for this interaction was numerically present across Exp. 5-7 and the analysis that collapsed the data from Exp. 5- 7 found a significant interaction in the expected direction, i.e., larger effect size in statives with respect to eventives, thus suggesting that the first option might be more likely.

These results suggest that the temporary ambiguity between verbal vs. adjectival interpretation, and the coercion into an eventive interpretation, both at play when a passivized state should have a verbal interpretation, affect sentence comprehension offline.

3.5.3.2 No interaction in the acceptability data.

As reported above, no evidence of an interaction between passivization and predicate type was found in the acceptability data. More importantly, these results are in direct contrast with the data collected by Ambridge and colleagues (2016), who found a significant interaction between passivization and predicate type in the expected direction, i.e., passives of experiencer-theme predicates (i.e., states) were less acceptable than passives of agent-patient predicates (i.e., events).

Given the numerous differences between our design and the design of Ambridge and colleagues' study, it is only possible to speculate on the reasons for this discrepancy. As discussed in several points across this chapter, the study run by Ambridge and colleagues presents several issues. Their study did not contain any filler items, which means that the study did not contain any ungrammatical sentences (beyond non-passivizable passives), thus artificially inflating differences between sentences when subjects are required to use the whole judgment scale in accordance with the instructions, but no anchor for the lowest point (i.e., an ungrammatical item) is provided, thus maximising scale bias across participants (Schütze & Sprouse, 2013). The absence of ungrammatical items might make participants more sensitive to differences in grammatical structures and allow an interaction between passivization and predicate type to be observed. The ungrammatical fillers in our design might have, thus, prevented us from observing this effect.

A closer look at Ambridge et al.'s interaction effect (i.e., Exp. 2) shows that when unpassivizable passives are included they appear to solely account for the interaction with predicate semantics and there is little if any difference between agent-patient and experiencer-theme predicates. When non-passivizable passives are excluded and a subset of each of the three predicates (agent-patient, experiencer-theme, theme-experiencer) included, there is nonetheless an interaction, but criteria for predicate exclusion are not clear and do not seem accounted for based on predicate class or semantic affectedness of the predicate on the internal argument (see Chapter 1 for more details on semantic affectedness). Hence, the nature of the interaction effect reported in their study is less transparent than suggested.

Finally, it is important to underline that our instructions specifically asked participants to ignore plausibility when rating the sentences. However, the difficulty in

passivizing stative verbs is actually related to the verb meaning, beside their grammatical structure. Our instructions might have thereby made participants less sensitive to an interaction effect.

Overall, the numerous differences between our design and the design of Ambridge and colleagues' study do not allow to conclusively determine the reason for the discrepancy between our results. Future studies are required to assess which factors allow for an interaction between predicate type and voice to emerge in acceptability judgment ratings.

3.5.4 Moving beyond the current study.

While the accuracy data showed evidence of an interaction between passivization and predicate type, the acceptability and self-paced reading data did not. I have already considered in the previous section possible reasons for the lack of such an effect in acceptability judgments. Here, I consider a further explanation, as well as one for the lack of an interaction in self-paced reading.

In addition to design features (e.g., presence of ungrammatical fillers, instructions) possibly masking an interaction effect between passivization and predicate type, it might be possible to hypothesize that acceptability judgments are not a sensitive enough task to detect such an interaction. Previous studies on the sensitivity of acceptability judgments to processing effects (e.g., Bogal-Allbritten, 2014; Sprouse, 2008) revealed that not all effects of temporary representations, e.g., reanalysis and coercion, can be detected in acceptability judgments. Indeed, the theoretical literature on an interaction between passivization and predicate properties (Gehrke & Grillo, 2007, 2009; Snyder & Hyams, 2015) suggests that passivizing statives involves temporary ambiguity and revision from adjectival to verbal interpretation. If indeed acceptability judgments are not sensitive to reanalysis or coercion effects, then we may not be able to observe an interaction between passivization and predicate type in acceptability data.

A similar reasoning can be applied to self-paced reading. Several psycholinguistic studies observed that the two processes responsible for the additional processing complexity in passives of states, i.e., revision and coercion, are correlated with processing measures that are understood to tap into late stages of processing (e.g., re-

reading, regression; Boland & Blodgett, 2002; Frazier & Rayner, 1982; McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; Traxler, Pickering, & McElree, 2002). Crucially, self-paced reading only provides early, but not late, measures of processing, which might explain the absence of an interaction in the previous self-paced reading studies reported in Chapter 2.

Alternatively, as already mentioned in a previous section, the interaction effect found in the accuracy data in Exp. 7 might be due to a spurious effect, although the analysis collapsing data across experiments revealed the same interaction effect, going in the expected direction of an increased difficulty in passivizing states, thus possibly confirming the results of Exp. 7.

The next study should hence adopt a new methodology that is sensitive to both reanalysis and coercion to discern between these two options. In particular, the next study will use eye-tracking while reading, which allows for the measurement of both early and late measures of processing and could hence detect any effect due to reanalysis and coercion online, to establish whether passive sentences processing is modulated by predicate semantics.

4 Interaction between passivization and predicate semantics in eye-tracking.

4.1 Introduction.

In the previous series of 7 experiments (see Chapter 2 and 3), I tested the interaction between passivization and predicate type. Experiments 1-4 used self-paced reading coupled with sentence comprehension questions, while Experiments 5-7 used an acceptability judgement task along with verification questions. All experiments used the same stimuli, which incorporated two manipulations: voice (passive or active) and predicate type (stative or eventive; see (45)-(48) for example items).

(45) Eventive passive: The guitarist was rejected by the attractive and talented singer in the concert hall next to the Irish pub.

(46) Eventive active: The guitarist rejected the attractive and talented singer in the concert hall next to the Irish pub.

(47) Stative passive: The guitarist was cherished by the attractive and talented singer for keeping the band focused through the whole tour.

(48) Stative active: The guitarist cherished the attractive and talented singer for keeping the band focused through the whole tour.

Across the 4 self-paced reading experiments, passives were found to be read faster than actives, irrespective of the verb, at several regions: verb, determiner, first adjective, conjunction and second adjective. The faster reading times were consistently reported across experiments for the verb and determiner regions, while results varied across experiments with respect to the other regions. These results were consistent with previous findings reported in the literature (Carrithers, 1989; Traxler et al., 2014). The faster reading times in passives with respect to actives are compatible with expectation-based (e.g., Levy, 2007) or surprisal-based models (e.g., Hale, 2001; see Chapter 1 for a description of these models and Chapter 2 for a more detailed analysis of the self-paced reading results). Interestingly for my purposes, in the previous studies the effect of passivization on reading times did not differ across predicate types, contrary to my expectations.

On the other hand, offline evidence in favour of an interaction between passivization and predicate type was found in the comprehension accuracy data. In 5 out of the 7 verification tasks, run in our previous studies, passives were found to be more errorful than actives, independent of the predicate type (Exp. 1 found passives to be numerically more accurate than actives, while in Exp. 7 we did not find a main effect of syntax). Data across the last 3 experiments (Exp. 5-7[23]), in which a larger difference was consistently observed in accuracy across voice in statives with respect to eventives, were pooled to increase the power of the statistical analysis. This revealed a significant interaction between passivization and predicate type. Although the effect of voice was significant in both predicate groups, the interaction derived from the fact that the difference across passivization in the stative condition (stative active: 91.42%; stative passive: 84.97%) was more than double the difference in the eventive one (eventive active: 91.24%; eventive passive: 88.33%). While the increased number of errors across passivization with stative predicates was expected based on the findings of previous acquisition, aphasia and self-paced reading experiments presented above, the difference in eventives was not. In the previous chapter, I interpreted this effect to be due to increased demands imposed during the comprehension task when answering active questions about passive sentences and not due to difficulty with parsing and/or interpreting eventive passives.

An interesting question thus emerges: why does the interaction between passivization and predicate type, suggested by the theoretical literature and reported in acquisition and aphasia studies, only emerge in offline accuracy and not online or acceptability data? As described in the previous section, processing passivized states could be more difficult than passivized eventives due to revision of the preferential adjectival interpretation and/or coercion of the original meaning of the stative verb into a consequent state of an event (Belletti & Rizzi, 1988; Gehrke & Grillo 2007, 2009; Grillo, 2008; Snyder & Hyams, 2015).

Previous studies on the sensitivity of acceptability judgments to processing effects (e.g., Bogal-Allbritten, 2014; Sprouse, 2008) revealed that not all effects of temporary representations, e.g., reanalysis and coercion, can be detected in

[23] Data from previous experiments (Exp. 1-4) were not included, given that different methodologies (self-paced reading vs. acceptability) and procedures (lab based vs. web-based) were used across experiments.

acceptability judgments. Hence, it might not be possible to observe an interaction between passivization and predicate type in acceptability data.

Moreover, effects caused by both revision and coercion are usually observed in late psycholinguistic measures, which cannot be collected in self-paced reading (e.g., regression, re-reading; Boland & Blodgett, 2002; Frazier & Rayner, 1982; McElree et al., 2001; Traxler et al., 2002). In the next section I will suggest that the lack of an interaction between voice and predicate type in our previous online measures stems from a limitation of self-paced reading measurements. Moreover, this limitation is not present in eye-tracking while reading, making it a target method for studying the interaction between voice and predicate type. Nonetheless, eye-tracking while reading has its own set of limitations relevant to the main effect of voice, as will be discussed below.

4.1.1 A trade off: advantages and limitations of eye-tracking while reading compared to self-paced reading.

Self-paced reading measurements present a limitation that specifically affects the ability to detect an interaction between passivization and predicate type: self-paced reading measures only the time spent reading a region for the first time (i.e., early processing measures). However, studies investigating revision and coercion effects, in contexts other than passivization, report effects in late processing measures (not first reading times).

The early processing measure collected by self-paced reading is captured by *first-pass* reading times in eye-tracking while reading (see below for a detailed definition of first-pass). In addition to first-pass, eye-tracking while reading provides measures of re-reading, regression path and total reading durations (see below for definitions), among others. Collectively, these measures offer a more informative timeline of reading behaviour, including any late parsing or interpretation difficulties, with respect to other methodologies (e.g., self-paced reading).

- First pass: sum of all fixation durations in a region before it is exited leftward or rightward;
- Regression path: sum of all fixation durations in a region and all preceding regions before it is exited to the right;

- Re-reading time: sum of all second-pass fixation durations in a region after it is exited to either right or left;
- Total time: sum of all fixation durations in a region.

First pass commonly reflects initial reading stages, while regression path, re-reading time and total time are associated with later or second pass stages.

The ability to distinguish between early and late processes provides an advantage of the eye-tracking methodology with respect to self-paced reading. This advantage has previously been established by cross-methodological studies comparing self-paced reading and eye-tracking results. In ambiguity resolution studies it was found that the late re-analysis effects captured by eye-tracking measures could not be observed in self-paced reading (Ferreira & Henderson, 1990; Jackson, Dussias, Hristova, 2012). Given these considerations, eye-tracking while reading represents the best methodology to study the interaction between predicate type and passivization, given that the effects of coercion and revision tend to be observed in late measures and not early ones (e.g., McElree et al., 2001; Traxler et al., 2002).

Moreover, the ability to distinguish between early and late processes in eye-tracking while reading would also offer the possibility to further assess the *Good Enough* model predictions with respect to passives processing. In fact, according to this model (Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016), revision/integration of the heuristic output with the algorithmic one might occur in late measures of processing (e.g., re-reading, regression). However, a late effect of passivization cannot be detected in self-paced reading as its measures only reflect early stages of processing. Adopting the eye-tracking while reading methodology would thus provide a further opportunity to test the *Good Enough* model predictions.

Finally, it is important to point out that eye-tracking also has limitations, particularly in the case of the voice contrast. In eye-tracking while reading, a well-known, investigated phenomenon is word skipping. Short words, that have a high frequency and that can be easily predicted from context, are skipped 70% to 80% of the time they are encountered (Schotter, Angele & Rayner, 2012). Skipping words allows readers to be more efficient, without losing information required to build an overall interpretation of the sentence. However, the time required to fully access and integrate these skipped words must be then redistributed across adjacent regions. In particular, orthographic

processing of skipped words is thought to take place in the region preceding the skipped word, while syntactic and semantic processing likely occur in the region following it. Hence, syntactic and semantic processing of skipped words should likely cause longer fixation durations on and regressions out of the region following the skipped word (Angele & Rayner, 2013).

Crucially, passivization is signalled by short and highly frequent words, such as the auxiliary "was" and the preposition "by", which are not present in the active version of the sentence. If these words are skipped, then their processing and integration will be redistributed across adjacent regions, possibly causing longer fixations on and regressions out of the right-ward region to access and integrate the skipped word(s) (e.g., if "by" is skipped in (49), but it is not present in (50), we would expect to find longer reading times on "the boy" in the passive, i.e., in (49) than the active, i.e., (50), version of the sentence) for reasons independent of passivization difficulty. Hence, interpreting any main effect of voice in early eye-tracking measures is complicated by the word-skipping confound.

(49) The girl was pushed by the boy.

(50) The girl pushed the boy.

For this reason, when discussing our results, I will only focus on possible interaction effects between passivization and predicate type, which would signal an effect independent of this confound, and on passivization effects in late processing measures. Main effects of passivization in early measures will be reported for posterity but not given any focus in the interpretation of the results given that their source is complicated by the 'word skipping' confound.

4.1.2 Current study: aim and predictions.

The present experiment aims at investigating a possible interaction between predicate semantics and passivization. Given previous psycholinguistic studies on related processing phenomena (Boland & Blodgett, 2002; Frazier & Rayner, 1982; McElree et al., 2001; Traxler et al., 2002), I would expect this interaction to surface in late measures of processing. The present experiment will thus employ eye-tracking while reading, which captures both early and late measures of processing, and should

thus allow to observe the interaction effect described in the theoretical, acquisition and aphasia literature (e.g., Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Grodzinsky, 1985; Maratsos et al., 1985; O'Brien et al., 2006; Volpato et al., 2015). The ability to distinguish early and late measure of processing should additionally allow to detect possible late effects of passivization, predicted by the *Good Enough* model.

This experiment would thus have two possible outcomes with respect to my aim. Firstly, it might be possible to find evidence for an interaction between passivization and predicate type in the expected direction, i.e., longer fixation durations and more regressions in the stative passives condition with respect to their active counterpart, but no difference in eventive. This difference should only emerge in late processing measures, i.e., re-reading, total time, regression path, and not early measures, i.e., first pass. This is due to the fact that the hypothesized causes of this interaction, revision and coercion, have been found to be reflected by late, rather than early, measures of processing (Boland & Blodgett, 2002; Frazier & Rayner, 1982; McElree et al., 2001; Traxler et al., 2002). These results would be compatible with previous studies in the theoretical, acquisition and aphasia literature that reported data in support of an interaction. Moreover, they would be compatible with our previous self-paced reading and comprehension accuracy data on passivization (presented in Chapter 2 and 3), given that the former does not offer a possible comparison with late measure effects, while the latter have already provided supporting evidence in favour of an interaction.

Secondly, no evidence of an interaction could emerge in this experiment even in late measures of processing. This would be compatible with our self-paced reading data, in which passives were read faster than actives independent of predicate type. Given the previous theoretical, acquisition and aphasia studies, which observed and reported data in favour of an interaction using offline measures, a lack of evidence towards an interaction in this experiment would indicate that this effect only emerges in *very* late stages of processing, i.e., post-interpretative processes, which cannot be captured by online measures.

Additionally, the presence or absence of a passivization effect in late measure of processing (e.g., longer fixation duration in passives with respect to actives in re-reading or regression path) would respectively provide support for or evidence against the *Good Enough* model of passive sentence processing.

In keeping the design similar to the self-paced reading ones (Chapter 2), accuracy data were also collected, by inserting comprehension questions after each item (experimental or filler) without manipulating the syntax of the comprehension question. Hence, I do not predict to observe an interaction between passivization and predicate type in the accuracy data. Rather, passives should be more errorful than actives independent of predicate type, as was reported in previous studies on healthy adults (Ferreira, 2003; our own experiments reported in Chapter 2 and 3), which did not control for task characteristics (such as the syntax of the comprehension questions, as discussed in Chapter 3) and predicate properties, or did not have sufficient statistical power to detect such interaction.

4.2 Methods and Design.

4.2.1 Participants.

Fifty-eight British English native speakers were recruited to participate in the study (38 females; average age: 24). They were all aged between 18 and 50 and had no visual or hearing impairment.

Participants were recruited through the UCL Sona System and received either payment or credits for their participation. All the participants were informed of the aims and procedures of the experiment and provided informed consent, approved by UCL ethics. None of the participants had previously participated in any of the experiments related to this study (including the pre-norming tests).

4.2.2 Stimuli.

The 28 sentence sets were kept identical to those used in Exp. 3 to 7. Given that eye-tracking is an online methodology, we re-inserted sentence-final prepositional phrases (PP; in contrast with Exp. 6 and 7) to avoid end-of-sentence effects (Just et al., 1982). However, we shortened the sentence completions that were previously employed in Exp. 1-5 for two reasons: (1) no difference in online processing of passive vs. active sentences was detected after the head of the by-phrase in self-paced reading making long sentence-final PPs unnecessary; (2) for-clauses in our stative items contained a syntactic dependency between the subject or object (subject in the passive

version of the sentence; object in the active version of the sentence) of the matrix clause and the implicit subject of the for-clause that is irrelevant in self-paced reading but not so in eye-tracking. In self-paced reading, any linguistic material presented after the region of interest could not affect reading times on the region of interest, since following regions are masked. On the contrary, in eye-tracking while reading sentences are presented in their entirety, thus allowing participants to re-read earlier regions. In this context, any additional complexity beyond the pure syntactic difference between active and passive sentences could interact with processing times and confound the data. Hence, we substituted the original sentence completions (locative PPs in eventive items and for-clauses in stative items) with simple temporal PPs. The temporal PPs were always 3 words long and were introduced by the prepositions "on", "in", "during", "at" in the eventive items, and "throughout", "since", "after", "before" in the stative ones. The 8 different prepositions were balanced across the 28 sets (e.g., 7 items contained "on", 7 "in", 7 "during", 7 "at", see Table 19; see Appendix 7.14 for the complete list of experimental items).

Eventive	Passive	The guitarist was rejected by the attractive and talented singer on Tuesday morning.
Eventive	Active	The guitarist rejected the attractive and talented singer on Tuesday morning.
Stative	Passive	The guitarist was admired by the attractive and talented singer throughout the tour.
Stative	Active	The guitarist admired the attractive and talented singer throughout the tour.

Table 19: Example of experimental stimuli in Exp. 8.

Given that our experimental items were shortened, fillers were accordingly shortened (see Appendix 7.15 for the complete list of fillers).

4.2.3 Procedure.

The items were presented in 11pt Monaco font. All sentences were presented on a single line. Comprehension questions followed each sentence, but they only targeted theta-roles after experimental items (e.g., "Did the musician reject the guitarist?"; see

Appendix 7.8 for the complete list of comprehension questions). Given that the present experiment was designed to further investigate online processing, and hence comparability to previous self-paced reading experiments was the most important, we kept comprehension questions in the active form, as in Exp. 1-6 (but in contrast to Exp. 7). To avoid creating a bias in attention towards specific parts of the sentence (in this case, the NP-VP-NP part), fillers were followed by complex and simple (i.e., targeting other sentential aspects, such as time or location) questions, as in the previous experiments.

The experiment was implemented using EyeTrack, an application developed by the UMass EyeLab to create and run eye-tracking experiments for the SR Research EyeLink trackers (http://blogs.umass.edu/eyelab/?_ga=2.174607394.1867291940.1502543167-1441954052.1502543167). Eye-movements were recorded using an EyeLink 1000 eye tracker (Desktop Mount) with a 2000Hz infrared camera (SR Research Ltd., Mississauga, Ontario, Canada), interfaced with two PC computers. Sentences were displayed on a PC monitor, positioned at a 70cm distance from the participants' eyes. The tracker was placed at a 55cm distance from the participants' eyes.

After providing informed consent, the Miles test was performed to determine the subject's dominant eye, which would be tracked during the experiment (Roth, Lora & Heilman, 2002). This test required subjects to stand approximately 150cm away from a sticker of circles on the wall in front of them. They were instructed to extend their arms and bring both hands together to form a triangle in front of their eyes. They were asked to centre the circles within the triangle keeping both eyes open. They were then asked to close one eye at a time and describe whether the circles moved from the centre of the triangle, and whether they moved more with one eye closed or not. The dominant eye was taken to be the one that would make the circles move less when open. Although viewing was binocular during testing, only the dominant eye, as determined by the Miles test, was tracked. Due to the characteristics of the tracker used, participants could not wear contact lenses during testing.

After the eye-dominance test, participants were sat on a chair, adjustable for height, with their forehead on a headrest and chin on a chinrest and asked to adjust the chair for their comfort. Calibration and validation were then carried out to insure precise

tracking of eye-movements. Participants then read the experiment instructions, which asked them to read accurately and at their normal speed. There was then a practice session of 6 trials to familiarize participants with the experimental procedure. A black box would appear on the left side of the screen (positioned at the left edge of where the sentence would appear) and participants had to fixate the box until the tracker detected their fixation. Detection of eye-gaze would then trigger stimulus presentation. After reading the sentence, participants would press the centre button in a button-box which was comfortably placed in their hands. Comprehension questions would follow each item and responses were recorded using the same button-box. After answering the question, a new black box would appear on the screen to trigger the presentation of a new trial. Re-calibration was carried out every time the tracker accuracy became unreliable. This could be identified by the inability to trigger a trial's stimulus presentation. Similarly, re-calibration occurred after any break requested by the participant.

In total, 84 experimental stimuli (28 experimental items and 56 fillers) were presented to each participant. The task lasted approximately 45 minutes and was administered in a soundproofed room.

4.2.4 Data analysis.

Ten participants were excluded from the final analysis due to the following reasons: 1 participant scored low on filler accuracy (75%); 8 participants had more than 50% of trials per condition eliminated due to excessive blinking/track loss/long saccade in a critical region; 1 participant was excluded due to a clear lack of reading (short and few fixations that span long parts of text). Data from 48 participants were analysed.

The outcome measures were: accuracy and eye-movements (first pass; regression path; re-reading time and total time).

4.2.4.1 Comprehension Accuracy data:

Accuracy data were analysed using a mixed effects logistic regression including syntax and predicate type along with their interaction as fixed effects, and both subjects and items as random effects (including both intercepts and slopes). The contrasts were passive vs. active [-0.5, 0.5] and eventive vs. stative [-0.5, 0.5]. During the discussion of

the results, possible modifications of the random effects structure due to convergence failure will be indicated.

4.2.4.2 Eye-tracking data:

Eye-movements were initially pre-processed to exclude trials based on: (1) presence of a blink/track loss in the critical region, (2) more than 15 blinks overall within the whole trial (including the stage before the sentence was triggered), or (3) presence of a long saccade. Overall, 1.19% of trials and 0.95% of fixations were excluded after pre-processing. Participant fixation count (with a lower threshold <200) across all trials was also calculated to serve as a possible exclusion criterion.

Eye-movements were then analysed in RStudio (https://www.rstudio.com/). Each eye-movement measure (first pass; regression path; re-reading time; total time) was analysed within each of the critical sub-regions (region 1, containing only the verb; region 2, containing both the determiner and the first adjective; region 3, containing both the coordinator and the second adjective; region 4, containing the head of the by-/direct object phrase). Data were analysed with a linear mixed effects model including syntax and predicate type along with their interaction as fixed effects, and both subjects and items as random effects (including both intercepts and slopes). The contrasts used were passive vs. active [-0.5, 0.5] and eventive vs. stative [-0.5, 0.5]. P-values were determined through treating the t-value as a z-statistic (Barr et al., 2013).

To increase comparability with our previous self-paced reading results (presented in Chapter 2), we separately analysed accurate trials vs. accurate and inaccurate trials. The results of both analyses, whenever not converging, will be reported in the Results section for posterity. However, only the analysis of eye-movements across all trials will be discussed in the Discussion section, given that analysing accurate trials only would decrease the power of our statistical analysis, as we would eliminate up to 20% of trials across conditions.

4.3 Results.

4.3.1 Accuracy data:

In the verification task, there was a significant main effect of syntax, due to the fact that accuracy on passives was lower than accuracy on actives (β=-1.05, z=-5.06, p<.001[24]; mean passive accuracy: 84.5%; mean active accuracy: 93.5%; see Table 20). There was no effect of predicate type or interaction between syntax and predicate type. However, the difference in accuracy across passivization in statives was double that in eventives (average difference eventive active – eventive passive = 5.95; average difference stative active – stative passive = 11.3). Despite this difference, the data overall indicate that passives were more errorful than actives, independent of predicate type.

	Passive	Active
Eventive	88.99% (±4.53)	94.94% (±3.17)
Stative	80.66% (±5.71)	91.96% (±3.93)

Table 20: Accuracy to comprehension questions in Exp. 8.

4.3.2 Eye-tracking data:

As anticipated in the Introduction section, we will discuss results for the following measures: first pass, regression path, re-reading time and total time from accurate and inaccurate trials. These will be considered for each of the four critical regions analysed: (1) verb, (2) determiner and first adjective, (3) conjunction and second adjective, (4) head of the "by"/direct object phrase.

A preliminary consideration needs to be done before discussing the results in detail: across the four measures and three out of the four regions, we consistently found a significant main effect of predicate type (at the verb in first pass: β=152.18, t=3.59, p<.001, regression path: β=152.18, t=3.59, p<.001, re-reading time: β=184.44, t=2.65, p=.008, and total time: β=295.88, t=3.78, p<.001; at the second region in re-reading time: β=135.98, t=2.12, p=.03, and total time: β=215.28, t=2.68, p=.007; at the third region in re-reading time: β=132.86, t=2.45, p=.01, and total time: β=158.24, t=2.33,

[24] The model only converged with intercept and no slope in the random effect structure.

p=.02). This effect was caused by shorter fixation durations in eventives with respect to statives, independent of passivization. It is hard to interpret this effect, given that the two predicate groups differed both in terms of frequency and plausibility. In fact, a post-hoc analysis, run on data taken from SUBTLEX-UK revealed that our eventive verbs are twice as frequent as our stative ones (eventives Zipf frequency: 4.65; statives Zipf frequency: 2.48). On the other hand, our eventive items are significantly less plausible than our stative ones (*β*=.35, *t*=2.65, *p*=.008), as established in a previous plausibility norming test (see Chapter 2 for more details). Critically, recall, predicate type does not interact with passivization in plausibility – the contrast of interest.

At the verb, we only observed a trend towards an interaction between predicate type and passivization in early measures of processing, i.e., first pass (see Figure 10A). This trend was also observed in regression path (see Figure 10B) and became stronger in later measures. The interaction was in fact marginal in re-reading time (*β*=129.65, *t*=1.8, *p*=.07; see Figure 10C[25]), and then became significant in total time (*β*=188.62, *t*=3.38, *p*<.001). When broken down, the interaction in total time revealed a significant effect of syntax in both predicate types, but, crucially, in opposite directions (see Figure 10D): passives were processed faster than actives in eventives (*β*=-88.94, *t*=-2.22, *p*=.03), but slower than actives in statives (*β*=99.68, *t*=2.51, *p*=.01[26]). These data indicate that at the verb, passivizing a stative predicate is more difficult than passivizing an eventive one, with respect to their active counterparts. Moreover, reading an eventive predicate in a passive structure is facilitated relative to an active structure. No other effect was observed across the four measures.

[25] The interaction was not marginal when considering accurate trials only (*β*=102.37, *t*=1.31, *p*=.19). Although less significant, the effect went in the same direction as in the analysis of all trials, thus indicating that the loss of significance is not due to a difference in behaviour between accurate and inaccurate trials, but rather to the reasons outlined in the Data Analysis section (e.g., loss of power).

[26] The effect of syntax in both eventives and statives was only marginal when considering accurate trials only (effect in eventives: *β*=-84.6, *t*=-1.93, *p*=.05; effect in statives: *β*=81.89, *t*=1.84, *p*=.07). Although now only marginal, the effects went in the same direction as in the analysis of all trials, thus indicating that the loss of significance is not due to a difference in behaviour between accurate and inaccurate trials, but rather to the reasons outlined in the Data Analysis section (e.g., loss of power).

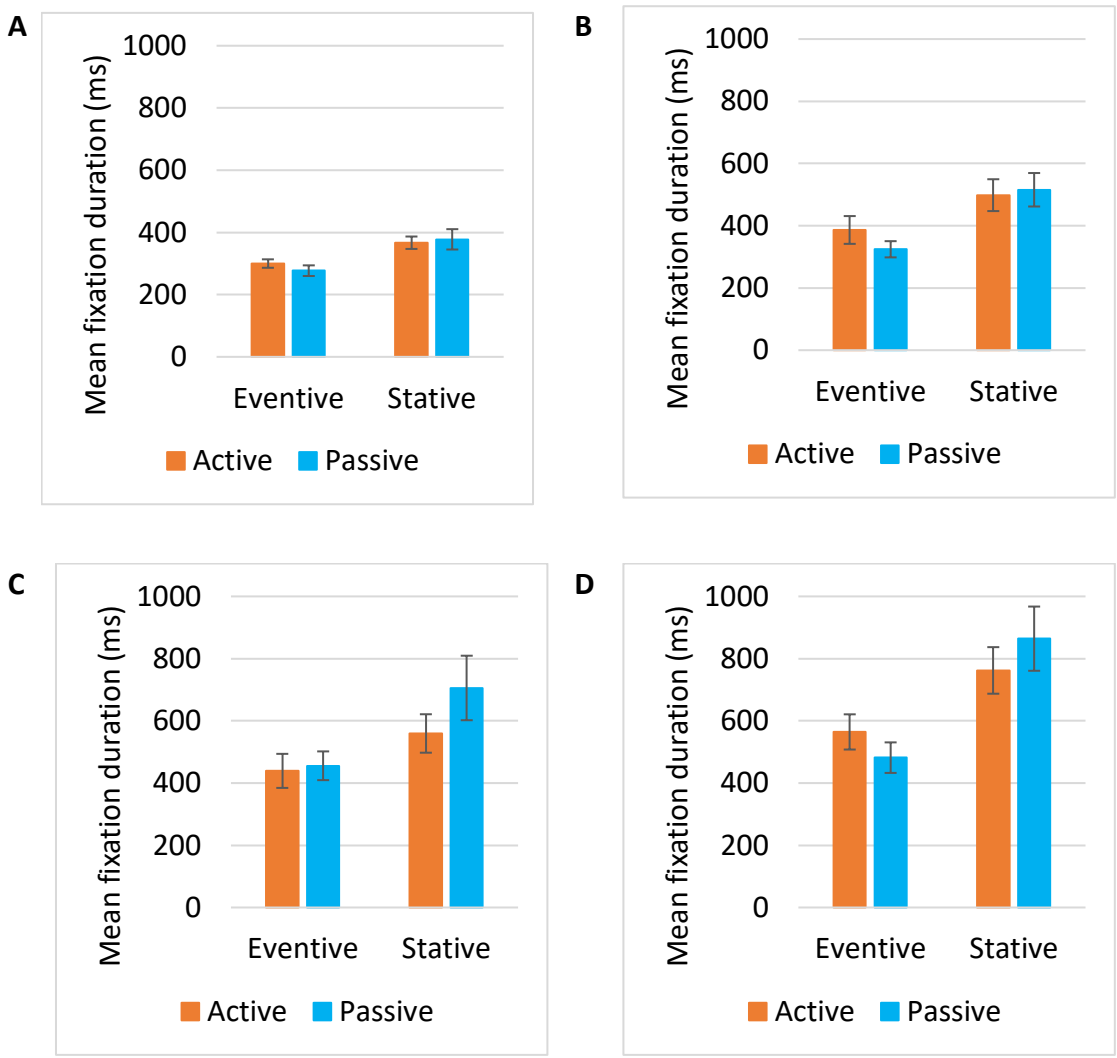

Figure 10: Eye-tracking measures at the verb in Exp. 8. ***A****: Mean fixation durations in first pass. Error bars indicate the standard error of the mean.* ***B****: Mean fixation durations in regression path. Error bars indicate the standard error of the mean.* ***C****: Mean fixation durations in re-reading time. Error bars indicate the standard error of the mean.* ***D****: Mean fixation durations in total time. Error bars indicate the standard error of the mean.*

At the second region, which comprised both determiner and first adjective, we found a significant effect of syntax (β=31, t=2.19, p=.03; see Figure 11A) in first pass, indicating that subjects fixated this region longer in the passive condition with respect to the active one, independent of predicate type. The same passivization effect was observed in total time (β=70.36, t=2.17, p=.03; see Figure 11B). However, given that no such effect was found in other late measures (regression path and re-reading time), we must conclude that the results in total time are only driven by the first pass effect. As mentioned in the Introduction section and as will be argued in more details in the

Discussion section, this passivization effect is difficult to interpret, due to the presence of a possible confound between a pure syntactic effect and effects due to word skipping. In conjunction with the self-paced reading data, which is comparable to first-pass reading times, it seems most likely that the effect of passivization in this region arises from word skipping. No other effect or interaction was observed across the four measures.

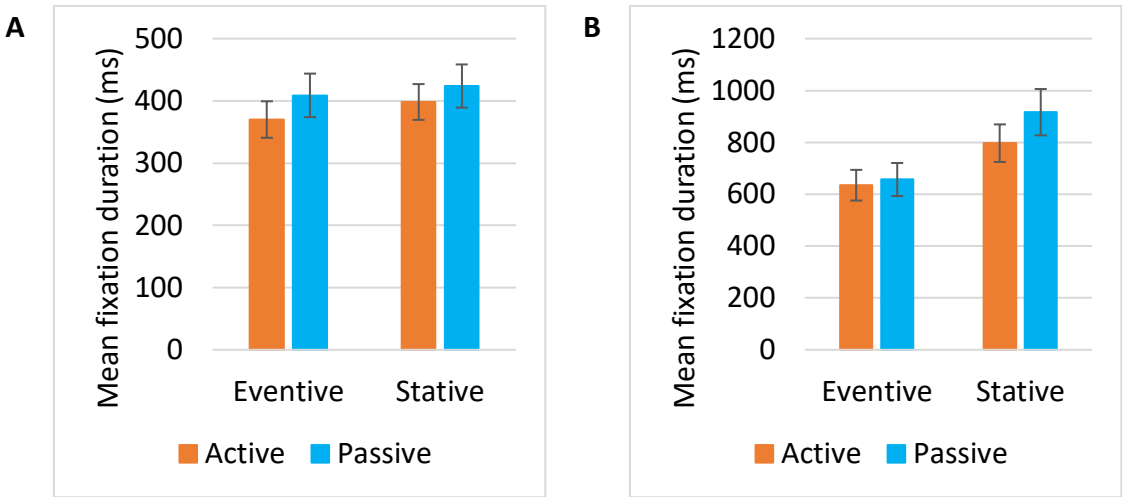

Figure 11: Eye-tracking measures at the second region (determiner and first adjective) in Exp. 8. ***A****: Mean fixation durations in first pass. Error bars indicate the standard error of the mean.* ***B****: Mean fixation durations in total time. Error bars indicate the standard error of the mean.*

The third analysed region was composed of the conjunction "and" and the second adjective. We did not observe any significant effect or interaction across the four considered measures. However, a trend towards an interaction in the expected direction, i.e., stative passives more difficult than stative actives but no difference in eventives, emerged in late measures, i.e., regression path, re-reading time and total time (see Figure 12A, B and C respectively).

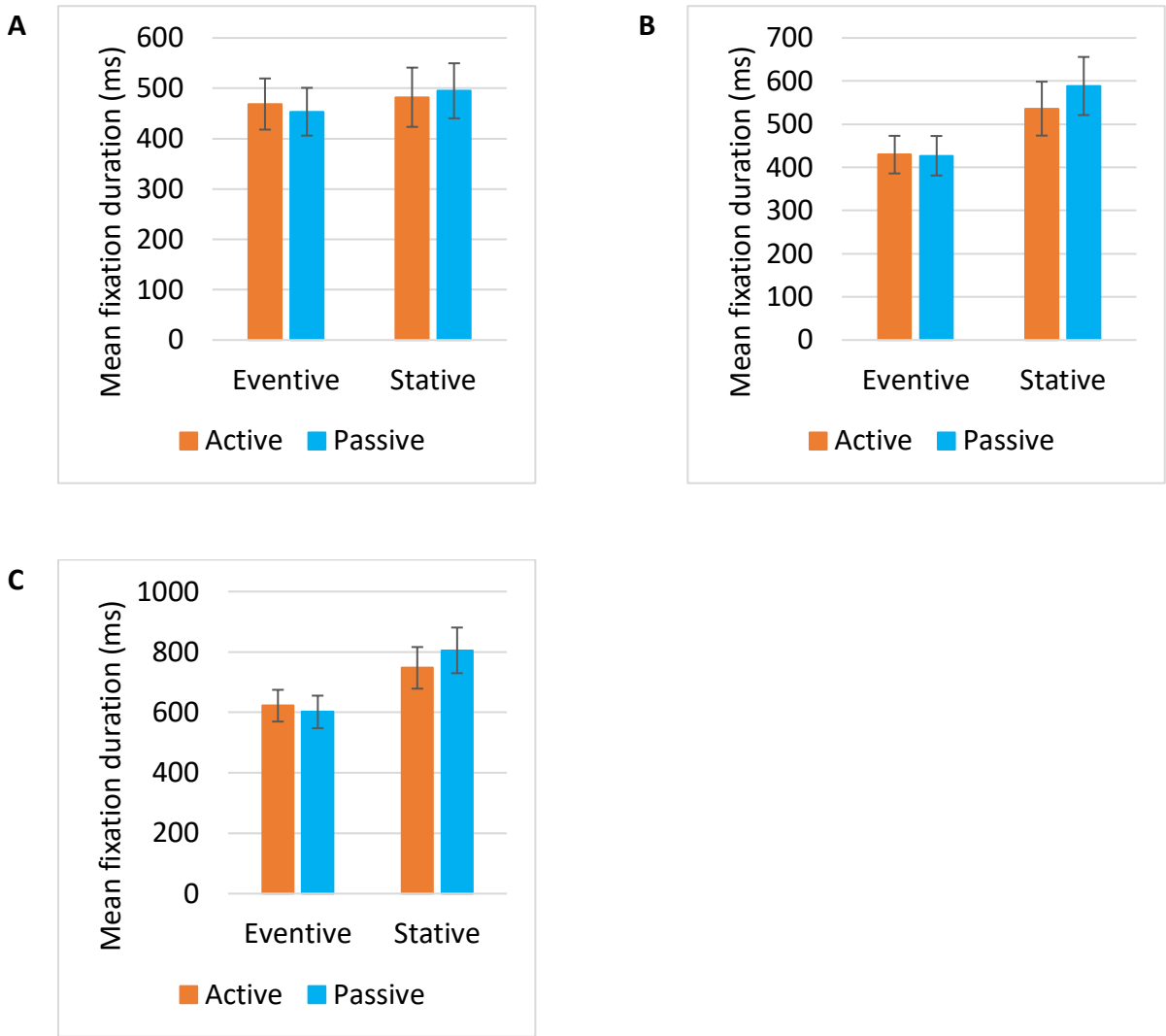

Figure 12: Eye-tracking measures at the third region (conjunction and second adjective) in Exp. 8. ***A****: Mean fixation durations in regression path. Errors bars indicate the standard error of the mean.* ***B****: Mean fixation durations in re-reading time. Errors bars indicate the standard error of the mean.* ***C****: Mean fixation durations in total time. Errors bars indicate the standard error of the mean.*

The final region only contained the head of the "by"/direct object phrase. No significant effect or interaction emerged in early measures, i.e., first pass. However, a significant interaction was observed in late measures, i.e., total time (b=98.83, t=2, p=.045; see Figure 13[27]). When broken down, this interaction revealed a significant

[27] The effect disappeared when considering accurate trials only (b=67.67, t=1.4, p=.15). Although now not significant, the effect went in the same direction as in the analysis of all trials, thus indicating that the loss of significance is not due to a difference in behaviour between accurate and inaccurate trials, but rather to the reasons outlined in the Data Analysis section (e.g., loss of power).

effect of syntax in statives (β=73.32, t=2.14, p=.03), due to the fact that passives required longer fixation durations than actives, while no difference was found in eventives. No other effect or interaction was observed across the four measures.

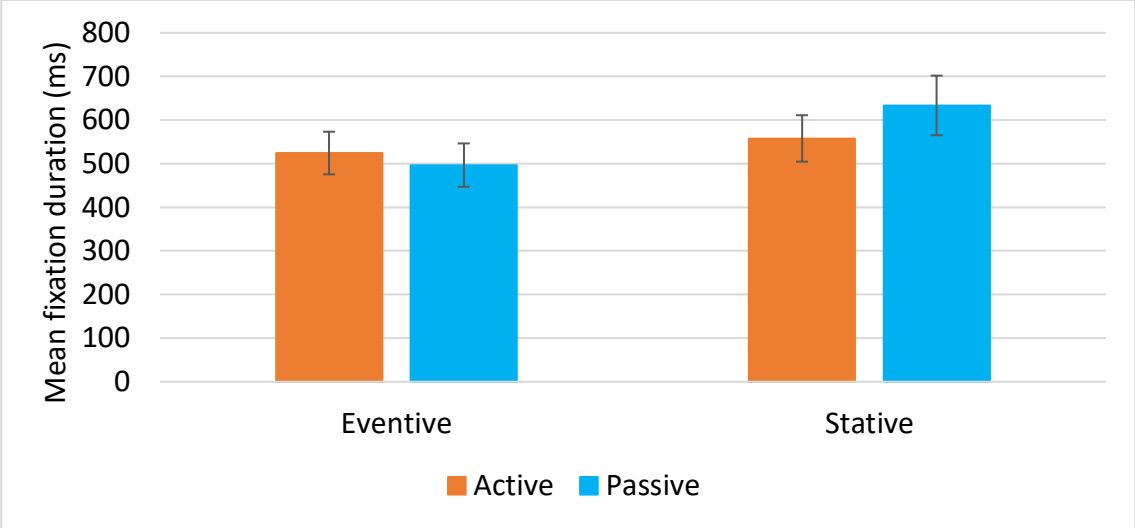

Figure 13: Mean fixation durations in total time at the head of the "by"/direct object phrase in Exp. 8. Errors bars indicate the standard error of the mean.

4.4 Discussion.

The present experiment collected eye-tracking while reading measures together with offline comprehension accuracy data to investigate whether an interaction between passivization complexity and predicate semantics is observed online when the measures target later stages of processing.

No evidence of an interaction was observed in early eye-tracking measures, while a numerical trend towards an interaction was observed at the verb in regression path and re-reading time, and at the third region (i.e., conjunction and second adjective) in regression path, re-reading time and total time. This effect became significant at the verb and in the fourth region (i.e., head of the "by"/direct object phrase) in total time. Across regions, the effect was always in the expected direction, i.e., passives of statives elicited longer fixation durations than their active counterparts, while no difference (or even a facilitation for passives with respect to actives) was found in eventives. Finally, two main effects were observed, a main effect of passivization in first pass at the second region, due to the fact that passives elicited longer fixation durations than actives, and

a main effect of predicate type across several regions and measures, due to the fact that statives elicited longer fixation durations than eventives.

In the accuracy data, passives were found to be more errorful than actives, independent of predicate type. No interaction between passivization and predicate semantics was found, although the difference in accuracy was double in statives with respect to eventives.

I will first discuss the observed interaction effects with respect to the original prediction, as these are of most interest. The other two main effects of voice and predicate type are of less interest due to the confounds mentioned previously, which I will review. Finally, I provide some consideration of the accuracy data in light of the results of the previous experiments (presented in Chapter 2 and 3).

4.4.1 Interaction between passivization and predicate semantics in online processing.

The eye-tracking data provided evidence in favour of an interaction between passivization and predicate semantics, in the expected late measures, i.e., regression path, re-reading time and total time. In particular, a significant interaction was observed in two different locations: verb and head noun of the "by"/direct object phrase (e.g., "rejected" and "singer" in (51)). In both locations, this interaction emerged in total time and showed the same numerical pattern. At the verb, we found a significant main effect of syntax in both the eventive and stative conditions. However, the effects were in opposite directions. Eventive passives had a shorter total fixation duration than eventive actives, but stative passives had a longer total fixation duration than stative actives. At the head of the "by"/direct object phrase, the effect of syntax was only significant in the stative condition, while no significant difference was found across syntax in the eventive condition.

(51) The guitarist was rejected by the attractive and talented singer on Tuesday morning.

The data are in line with the original predictions. I predicted an interaction effect between passivization and predicate type to surface in the form of stative passives being more difficult to parse than stative actives, while no difficulty should arise in eventive

passives with respect to actives, given the theoretical and psycholinguistic literature reviewed in Chapter 1 (Belletti & Rizzi, 1988; Gehrke & Grillo 2007, 2009; Grillo, 2008; Grodzinsky, 1995; Maratsos et al., 1985; McIntyre, 2013; O'Brien et al., 2006; Snyder & Hyams, 2015; Volpato et al., 2015). According to the relevant literature, both revision and coercion are considered the underlying causes for the additional difficulty in passivizing statives with respect to eventives (Gehrke & Grillo, 2007, 2009; Snyder & Hyams, 2015). Moreover, given that revision and coercion effects appear in late measures for other syntactic constructions in the literature (Boland & Blodgett, 2002; Frazier & Rayner, 1982; McElree et al., 2001; Traxler et al., 2002), I predicted this interaction to emerge in late, and not early, measures of processing.

These findings are also compatible with the absence of an interaction effect in the self-paced reading data (see Chapter 2). This is so, since self-paced reading does not provide late measures of processing.

It is important to discuss the facilitation for passives with respect to actives for the eventive predicates at the verb in total time in eye-tracking. Self-paced reading revealed a facilitation for passives compared to actives, but independent of predicate type, consistently at the verb and determiner. Thus, across these two methods there are two differences in the facilitation of passives over actives: (1) it extends until the determiner in self-paced reading and (2) applies to both predicate types in self-paced reading. The first difference can be explained by the well-known issue of spill-over effects in self-paced reading (Mitchell, 1984). It can be assumed that the main region for facilitation with passives is at the verb (see the General Discussion section of Chapter 2 for more details), and that the determiner represents a spill-over region for this effect in self-paced reading.

Explaining the second difference depends on two factors: (1) word-skipping in eye-tracking and (2) the stages of processing that the two measures capture. In eye-tracking early measures are contaminated by skipping effects, as described in the Introduction section. Hence, the absence of a facilitation effect for passives in first pass across predicate type is not surprising. In the passive, additional processing needs to occur at the verb to compensate for having skipped the preceding auxiliary across many trials. The next question should then be: why was the facilitation effect observed for eventive predicates still found in total time? This can be explained by the fact that total

time represents a mixture of early and late effects. Thus, the facilitation of the passive for eventive predicates was diminished in first pass (due to word skipping), but seems to have emerged in late measures. The nature of this late facilitation is currently not evident. In terms of the stative passives, revision and coercion are expected to affect processing in late measures, and it is thus not surprising that longer total fixation times were observed in stative passives in eye-tracking.

A final aspect to discuss in the eye-tracking results is the location at which the interaction between passivization and predicate type was observed. In particular, the interaction emerged at two specific points, i.e., the verb and head of the by-phrase, both in total time. As specified in the Introduction, total time is a composite measure, which represents the sum of all the fixation durations for a specific region. It is thus only by comparison with data collected in non-composite early and late measures that it is possible to identify the main contributors to an effect in total time. Given that no complexity was observed in early measures (i.e., first pass) at the verb, while a trend towards an interaction was present in late measures, and especially regression path, the difficulty at the verb appears to arise from late linguistic effects, as already hypothesised. In particular, the trend emerging in regression path, which is more defined than in other late measures (e.g., re-reading), suggests that participants spent more time reading the verb and other regions to the left of the verb in the stative passive condition with respect to its active counterpart. This possibly indicates revision of the initial adjectival interpretation of the verb, once more clues emerge, e.g., parafoveal processing of the “by” which would signal the verbal, rather than adjectival, nature of the passive.

Similarly to the verb region, no evidence for an interaction was observed in early measures at the head of the by-phrase, while a trend appeared in late measures, and particularly in re-reading time. Re-reading is the sum of all second-pass fixation durations for a region, after it has been exited to either right or left, which means after participants went back to re-read previous regions of the sentence or continued reading following parts of the sentence. It is hence difficult to establish the exact nature of the complexity arising at this region (a post-hoc analysis of probability of regression could not pinpoint a specific region triggering re-reading or regression). However, the difficulty seems to emerge at a point where a comprehensive interpretation of the verb phrase

(if not of the entire sentence) has been built. For this reason, it is not implausible to assume that longer fixation durations for stative passives with respect to their active counterparts in this region represent the effect of coercion of the verb meaning, coming into play once all the fundamental pieces of the sentence have been encountered and the full interpretation of the sentence can finally be construed.

Overall, the eye-tracking data are not only compatible with the predictions formulated in the Introduction section, but also with the self-paced reading and accuracy results. Taken together, the data suggest that the complexity of passivization depends on the properties of the passivized predicates. In particular, passivization of statives is more difficult than for eventives, due to revision and coercion, as evidenced by the eye-tracking data and accuracy data.

4.4.2 Main effect of passivization in online measure.

As mentioned at the beginning of this section, two main effects were found in our measures: a main effect of passivization in first pass at the second region (i.e., determiner and first adjective), due to the fact that passives required longer fixation durations; and a main effect of predicate type across the four measures and three out of the four analysed regions, caused by longer fixation durations in the stative compared to the eventive condition.

It is very difficult to interpret a main effect of passivization that occurs in a region following either the auxiliary ("was") or the "by" in the passive condition, due to the common finding that short words are skipped – 70-80% of the time – in the eye-tracking while reading methodology (Angele & Rayner, 2013; Schotter et al., 2012). When comparing passives and actives, the skipping phenomenon could introduce a confound, since actives do not contain these two short words. This concern is further strengthened in the region following "by" (the + first adjective), as two short, highly frequent words, "by" and "the" appear in sequence before the most likely first word to be fixated in that region (i.e., the first adjective).

However, "by" is a fundamental clue to passivization, and skipping it could either: (1) disrupt reading of the upcoming words, which otherwise cannot be correctly integrated or (2) entail lexical access and syntactic integration of "by" and "the" be performed at the point that is first fixated following them. Both scenarios could cause

longer fixation durations on and more probability of regressions out of the region following these words, i.e., our first adjective. Hence, longer fixation durations due to skipping cannot be disentangled from longer fixation durations due to syntactic complexity, which would be observed in the same direction (longer fixation durations on passives with respect to actives).

It is thus not possible to meaningfully interpret the observed main effect of passivization in first pass at this region. However, in consideration of the fact that it was found in first-pass, which is a measure also collected in self-paced reading, but it was nonetheless not observed in the previous four self-paced reading experiments (Experiment 1-4, presented in Chapter 2), it is highly likely that this effect is caused by skipping effects, which are only common to eye-tracking.

The absence of a passivization effect in late measures of processing (e.g., re-reading, regression path) provides evidence against the *Good Enough* model. This model would predict passives to be more difficult than actives due to revision/integration of the heuristic and algorithmic outputs. More specifically, this model might expect such a revision/integration effect to occur in late measures of processing, but the data presented here refute this hypothesis.

4.4.3 Main effect of predicate in online measure.

Similarly, it is very difficult to unequivocally interpret the main effect of predicate type observed across the four measures in the three initial regions. The main effect was due to the eventive condition giving rise to shorter fixation durations than our stative one, independent of syntactic structure.

As anticipated in the Results section, our two predicate groups differed both in terms of plausibility and frequency. Eventive predicates were more frequent than stative ones, but eventive sentences were less plausible than stative ones.

However, given that plausibility scores were collected for the entire sentence meaning, it seems unlikely that the difference in plausibility would have impacted processing at specific regions. Overall, frequency appears to be a better candidate to explain the predicate effect observed in the present data. Indeed, more frequent words usually elicit shorter fixation durations with respect to less frequent ones (Schotter et al., 2012).

It is important to note that, although a main effect of predicate type was observed in the plausibility and frequency analysis, there was no interaction between syntax and predicate type in these same measures. Hence, any findings regarding interaction effects between passivization and predicate type do not suffer from this confound.

4.4.4 Passivization in offline processing.

In the accuracy data, we found passives to be more difficult than actives independent of predicate type. This is compatible with the results collected in the previous experiments (see Chapter 2 and 3) and in the literature (Ferreira, 2003), which consistently found passives to be more errorful than actives, independent of predicate type.

Despite the fact that these data would seem compatible with models that predict passives to be inherently more difficult than actives (e.g., the *Good Enough* model, the syntactic complexity approach and the frequentist approach), I claim that passives are not inherently more complex than actives, but rather they are affected by certain properties of the sentence or the task in which they are inserted, as argued in Chapter 3.

Although passives were more errorful than actives across predicate type, the difference across voice in accuracy in the stative condition was double than in the eventive one. This indicates a trend towards a different processing of passivization in eventives with respect to statives, which clearly emerged when data were pooled across Exp. 5-7, presented in the previous chapter.

The trend observed in this experiment, together with the data collected in Exp. 5 to 7, are best accounted for by a model that poses that passives are not necessarily more complex to parse or interpret, but are, on the contrary, selectively susceptible to difficulty given specific properties of the comprehension task or of the predicate. I have identified three relevant factors in this chapter and in Chapter 3: (1) the nature of the task, i.e., whether the task targets post-interpretative processes, in particular, encoding and retrieving the interpretation of the target sentence, such as in the verification task; (2) comprehension questions only in an active voice that target theta-role assignment; (3) the semantic features of the predicate, i.e., whether it has eventive or stative properties. The verification task used in the present experiment was characterized by all

three factors, and it is hence not surprising that passive sentences were found to be more errorful than actives, independent of predicate type, although a larger difference was observed in statives with respect to eventives.

Critically however, the results of the experiments presented in Chapter 3 showed that once these factors are taken into account, i.e., when the offline task does not target post-interpretative processes (e.g., in the acceptability judgment task), when the syntax of the questions is balanced and predicate semantics is manipulated as a variable together with syntax, the offline passivization difficulty disappears and an interaction between passivization and predicate semantics emerges.

Any account of passives processing should hence take into considerations these factors and their role in passive sentences comprehension. I will present such an account in the next chapter.

4.4.5 Moving beyond these results.

Across several experiments (Exp. 1-4, see Chapter 2), passives were found to be processed faster than actives in early stages of processing, as evidenced by the self-paced reading data, indicating that passives are not more difficult than actives. This is further supported by the acceptability data (see Chapter 3), which found passives to be as acceptable as actives.

Passivization and predicate properties were found to interact online in late measures of processing in the eye-tracking data presented in this chapter. In particular, revision and/or coercion required by processing passivized statives create a slow-down with respect to their active counterparts, while no difference is found across voice in eventive predicates.

The same effects, revision and/or coercion, introduce an additional difficulty in offline tasks targeting post-interpretative processes, such as comprehension questions, as evidenced by a higher number of errors in stative passives with respect to their active counterparts. In the same offline task, i.e., comprehension questions, passive of eventives were also found to be more errorful than their active counterparts (Exp. 2-6, Exp. 8 presented in this chapter and in Chapter 2 and 3). However, the difference across statives was often much larger (in some experiments even double) than the difference in eventives.

To reconcile the online findings, i.e., facilitation in processing passives with respect to actives, and the offline findings, i.e., more errors in comprehension of passives than actives, I hypothesized that passives are not more difficult to parse or interpret than actives, but they are more susceptible to effects arising from post-interpretative processes, i.e., interference in retrieval of the interpretation of the target sentence. In Chapter 3, I tested this hypothesis and found that, once the factors contributing to this susceptibility were controlled for (see Chapter 3 for more details), the offline passivization difficulty disappeared and an interaction between passivization and predicate semantics emerged, thus nicely reconciling the online and offline findings both in terms of the comprehension of passive sentences in general and of the predicted interaction between passivization and predicate type in particular.

The mainstream models of passives processing presented in Chapter 1, i.e., the *Good Enough* model, the syntactic complexity approach and the frequentist approach, are hence not adequate to account for the data collected in this study. A new account of passives processing needs to be delineated, and I will propose my own version in the next chapter.

The next chapter will also consider the limitations of the experiments run in this study, as well as possible avenues of future research, and, in general, implications of the findings presented in this work for future studies on passive sentence comprehension.

5 General Discussion and Conclusions.

The three main theories of passive sentence processing in the literature include the *Good Enough* model, the syntactic complexity approach and the frequentist approach. The experiments presented throughout my dissertation failed to collectively support any of these theories. Here, I will recap the results of these experiments and present an alternative account of passive sentence processing that can adequately explain both the results obtained in these experiments, as well as those reported in the literature. I will then discuss possible limitations of my studies, as well as future avenues of research that might take these results as a starting point. Finally, I will consider broader implications of this study that impact psycholinguistic studies in general, particularly in terms of cross-methodological approaches.

5.1 Summary of the eight experiments and their results.

In the first four experiments (Exp. 1-4, presented in Chapter 2), data were collected through self-paced reading and comprehension questions, the two methodologies that were used independently by previous studies targeting passive sentence processing. Previous studies had found a discrepancy between online and offline results: passives were read faster online, indicating a parsing facilitation, but they were more errorful offline, indicating difficulty in comprehension (Carrithers, 1989; Ferreira, 2003; Traxler et al., 2014). The aim of these experiments was to simultaneously investigate passive sentence processing online and offline. Moreover, these experiments aimed at investigating a possible interaction between passivization and predicate semantics, observed in the theoretical and psycholinguistic literature: while eventives consistently deliver verbal passives, passives of statives can only be interpreted as verbal passives once revision and coercion occur (Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Grodzinsky, 1985; Maratsos et al., 1985; O'Brien et al., 2006; Snyder & Hyams, 2015; Volpato et al., 2015). From a processing point of view, revision and/or coercion could translate into an additional difficulty. Exp. 4 additionally collected measures of vWM to investigate whether variability in vWM across subjects might contribute to the increased number of errors on passives with respect to actives, suggesting that passives are not

more difficult than actives across the board, but rather encoding or retrieving the representation of passive sentences, as required by the verification task, might be.

Across the four self-paced reading experiments, passives were found to be read numerically faster than actives in early stages of processing, irrespective of the verb, at several regions: verb, determiner, first adjective, conjunction and second adjective. The faster reading times were significant across experiments at the verb and determiner regions, but varied across experiments with respect to the other regions. In the offline task (i.e., comprehension questions) passives were generally more errorful than actives, independent of predicate type. Moreover, a significant correlation was found between vWM span and the difference in accuracy across voice, while vWM did not correlate with online measures of processing. These findings indicate that online and offline performance stem from partially independent processes, thus implying that separate explanations of the online facilitation effect and of the offline complexity effect should be investigated. Moreover, the correlation between vWM and the difference in accuracy across voice suggests that the offline complexity effect in passives with respect to actives could partially be due to increased demands on memory components of post-interpretative processes. Specifically, accuracy failure could be explained in two different ways, either affecting encoding or retrieval of the parse in post-interpretative processing. Firstly, it has been proposed that participants create an "active-like" representation of the target sentence (e.g., agent-action-patient; Anderson, 1974) during post-interpretative tasks, such as comprehension questions. This strategy would be adopted to avoid memory decay. However, generating such a representation for passive sentences would require an additional transformation with respect to active sentences, thus possibly making the former more prone to errors during encoding.

Additionally, the complexity effect observed in passive comprehension could have had contributions from interference at retrieval due to a confound in the design of the verification task. The comprehension questions were always presented in the active form, independent of the syntax of the sentence they targeted. This might have introduced a priming effect for active sentences, whereby processing of the comprehension question would be facilitated in the active over the passive condition. Conversely, when answering a question about a passive sentence, the active syntax of

the comprehension question might have introduced interference when retrieving the representation of the target passive sentence.

The second set of acceptability judgement experiments (Exp. 5-7, presented in Chapter 3) tested the hypothesis that the offline accuracy effect was due to a failure in memory storage and/or retrieval. The acceptability judgement task was chosen as it is an offline task that does not require participants to store and/or retrieve the representation of the target sentence. The manipulations were identical to the first group of experiments: voice (active vs. passive) was manipulated within subjects, while predicate type (eventive vs. stative) between subjects. In the acceptability judgement task passives were not more difficult than actives, supporting the claims that additional demands on memory in the passive condition contributed to the complexity effect in accuracy. Furthermore, Exp. 7 introduced an additional modification to the design of the verification task: the syntax of the comprehension question was balanced across voice. This should reduce priming of an active structure and hence any facilitation in the processing of active questions. Additionally, it should reduce interference in memory between the active syntax of the question and the passive syntax of the target sentence. Once the syntax of the question was balanced across voice in the verification task, passives were not more errorful than actives across the board. The fact that passivization complexity disappeared once interference in retrieval was balanced across voice suggests that a difficulty in passive comprehension did not emerge at the encoding stage in previous experiments, but rather at the retrieval stage. Furthermore, an interaction between passivization and predicate semantics emerged in the expected direction in Exp. 7: passives of statives were more difficult to process than their active counterparts, while no difference emerged across voice in eventives.

While the accuracy data showed evidence of an interaction between passivization and predicate type, the acceptability and self-paced reading data did not. I hypothesized that this difference across methodologies could be due to the fact that acceptability and self-paced reading might not be sensitive to the effects of revision and coercion.

Abandoning a well-formed, grammatical representation (e.g., an adjectival interpretation of a stative passive) for another well-formed grammatical representation (e.g., a verbal interpretation of a stative passive) does not lower the acceptability of the

final representation (Sprouse, 2008). Hence, reanalysis does not affect acceptability. Similarly, coercion does not result in lower acceptability judgments (Bogal-Allbritten, 2014).

With regards to self-paced reading, previous studies have shown that revision and coercion were found to correlate with processing measures that are understood to tap into late stages of processing (Boland & Blodgett, 2002; Frazier & Rayner, 1982; McElree et al., 2001; Traxler et al., 2002). However, self-paced reading only allows to collect an early measure of processing, and it is thus not sensitive to late effects, such as reanalysis and coercion.

Crucially, revision and coercion are the processes responsible for the additional complexity in passivizing states (Gehrke & Grillo, 2007, 2009; Snyder & Hyams, 2015). It was thus necessary to use a different methodology, eye-tracking while reading, which allows for the measurement of both early and late measures of processing and could hence detect any effect due to reanalysis and coercion, to establish whether the complexity in passive sentences processing is modulated by predicate semantics.

The last experiment (Exp. 8, presented in Chapter 4) thus employed eye-tracking while reading with the same manipulations as previous experiments, i.e., voice (active vs. passive) was manipulated within subjects, while predicate type (eventive vs. stative) between subjects. Passivization and predicate properties were found to interact online in late measures of processing in the expected direction. In particular, revision and/or coercion intervened at the verb and at the head of the by-phrase in passivized statives creating a slow-down with respect to their active counterparts, while no difference was found across voice in eventive predicates, thus nicely reconciling the online and offline findings with respect to the predicted interaction between passivization and predicate type.

5.2 Existing models do not account for the data.

Any model of passive sentence processing should hence account for the following findings: (1) passive sentences show processing facilitation in early stages; (2) passives of statives require coercion and revision in order to be interpreted as verbal, and their effects appear in late stages of processing at the verb and at the head of the by-phrase; (3) the complexity introduced by revision and coercion in passivized states also affects

accuracy to comprehension questions; (4) passives of eventives do not induce more errors in comprehension unless the task is built in such a way to place additional demands on memory in the passive condition (e.g., when the syntax of the comprehension question is always in the active form). Hence, overall, these data indicate that passives are not more complex to process than actives. Rather, they might even facilitate processing when the passivized verb is eventive. However, specific characteristics of the task or of the passivized predicate might introduce an additional complexity in processing or retrieving the parse of a passive sentence with respect to an active one.

The mainstream models of passive sentence processing in the psycholinguistic literature argue that passives are more complex to process than actives, due to their non-canonical argument order (the *Good Enough* model, Christianson et al., 2001; Ferreira, 2003; Ferreira & Christianson, 2016; Ferreira & Patson, 2007; Karimi & Ferreira, 2016), their low frequency in language use (Johns & Jones, 2015) or an additional syntactic operation that derives passive sentences from active ones (Chomsky, 1981). These models cannot account for the present data, given that they all predict passives to be more difficult than actives, while passives were found to be as easy (if not easier) to process than actives in this study. Moreover, these models do not predict processing complexity to vary with predicate semantics.

However, this is not to say that these models cannot account for other linguistic phenomena. For example, the *Good Enough* model stipulates the existence of other types of heuristics from agent-first, which might explain different sets of data reported in the literature. In particular, the existence of a semantic plausibility heuristic, which combines the words meaning in the most plausible way, independent of the actual syntactic structure, would account for structural priming data in which listening to active implausible sentences primes the production of passive sentences (Christianson et al., 2010). Moreover, the existence of a semantic plausibility heuristic would also explain why a recent ERP experiment on semantic anomalous sentences (e.g., “The fox that shot the poacher…”) found no evidence of an N400 effect, which would be expected if sentences were only processed via algorithmic parsing that would detect a semantic anomaly (van Herten, Chwilla & Kolk 2006). Conversely, the parser might have processed the sentence along two routes, the algorithmic one and the heuristic one, with the latter

providing a plausible, despite incorrect, interpretation of the semantic anomalous target sentence (e.g., "The poacher shot the fox."). The conflict in integrating these two representations would be evidenced by a P600 effect, that was observed in the same study (van Herten et al., 2006).

Similarly, although these data are not compatible with a syntactic model of passives based on movement complexity, I do not argue against movement itself. Rather, if passives are indeed derived from an underlying active structure via movement, then this process is not costly in itself. Other structural configurations that coincide with movement might present additional difficulties for the parser or for memory mechanisms necessary to perform a specific task, thus conflating the two effects in previous studies. Studies on subject vs. object-relatives, which are unanimously thought to involve movement, reveal that complexity arises due to similarity-based interference, i.e., when the dependency has to cross a "similar" constituent to the antecedent (Fedorenko, Gibson & Rohde, 2006; Gordon et al., 2001; Van Dyke & McElree, 2006).

Given the mainstream models' inadequacy in accounting for the data collected in this study, it will be necessary to propose a new account of passive sentence processing, which is presented in the next section.

5.3 A combination account of passive sentence processing: Expectation-based models and post-interpretative task effects.

As mentioned above, the mainstream models of passive sentence processing do not adequately capture the data presented throughout this dissertation (see Chapter 2, 3 and 4), as they indicate that passives are not inherently more complex than actives, but rather their complexity can be dependent on the task or predicate type.

My approach argues that passives are not more difficult to parse and interpret than actives, as evidenced by the self-paced reading and acceptability data collected in this study, which found passives to be read faster than actives (see Chapter 2) and rated as acceptable as (or more acceptable than) actives (see Chapter 3), respectively. I have argued that passives are read faster due to their dedicated morphology, which creates expectation for upcoming syntactic categories (Hale, 2001; Levy, 2007; see section 1.4 for a detailed description of the expectation-based and the surprisal-based models and of the empirical evidence supporting them). After encountering an auxiliary, a verb is

more likely to occur than other syntactic categories (see (52)). Conversely, the subject NP in the active version of the sentence (see (53)) could be followed by several types of syntactic categories (e.g., PP, subordinate clause, AdvP), and hence the verb must compete with these alternative continuations, thus increasing surprisal and lowering expectation with respect to the same verb in the passive version of the sentence. Similarly, after encountering the past participle and "by" in (52), the only possible continuation of the sentence is a DP. At this point, the expectation for a specific grammatical structure to follow (DP) is so high (and, conversely, surprisal so low once it is actually encountered) that the processing of upcoming words is facilitated relative to the same region in the corresponding active sentence (i.e., at "pushed" in (53)). In the active, the verb can be followed by several sentence continuations (e.g. PP, AdvP), which are not possible after a "by". The upcoming DP in the active sentence is thus processed with a lower level of expectation and higher level of surprisal with respect to the corresponding DP in the passive sentence.

(52) The girl is pushed by the boy.

(53) The girl pushed the boy.

As discussed in Chapter 3, while a facilitation effect for passives with respect to actives was consistently found in the self-paced reading data, this facilitation effect was not (reliably) replicated in the acceptability data. This difference in results could be ascribed to the fact that self-paced reading reflects moment-by-moment processing. On the contrary, acceptability judgments are formed on the grammaticality of the sentence in its entirety. It may be the case that the increased expectations for an upcoming word category do not factor into acceptability judgments of the entire sentence or are balanced out by processing additional functional words on which the expectations rest.

While the noncanonical ordering is not difficult to parse and interpret, characteristics of the task, such as the syntax of the comprehension question, or the predicate semantics of the passivized verb can generate difficulty.

Chapter 3 and 4 present results demonstrating that the complexity of passives comprehension varies with predicate semantics. In particular, passives of statives were more errorful than their active counterparts, while a smaller difference (or no difference at all once confounding factors in the experimental design were controlled for) was

observed in eventives (see Chapter 3). Moreover, passives of statives elicited longer fixation durations than their active counterparts, while passives of eventives were read as fast, or even faster, than their active counterparts in eye-tracking (see Chapter 4). The observed data support the idea that certain predicates, i.e., statives, are more difficult to passivize than other predicates, i.e., events.

Finally, the characteristics of the tasks can selectively affect passive sentences during post-interpretative processes. Evidence in support of this view comes from the accuracy data collected across six experiments (Exp. 3-6 and Exp. 8), which found passives to be more errorful than actives when comprehension questions had an active syntax and targeted theta-role questions. As discussed in detail in Chapter 3, these two characteristics created memory retrieval interference in passive sentences only. During tasks that target post-interpretative processes (e.g., retrieving the stored interpretation of the target sentence), such as comprehension questions, participants store the representation of the sentence to answer the upcoming comprehension question. When the memory of the sentence needs to be retrieved to answer the question, a possible interference might arise between the active syntax of the question, and the actual passive syntax and thematic roles assignment of the original passive sentence.

This hypothesis can also apply to the results reported in the acquisition and aphasia literature pointing to a general difficulty with passive sentences. The methodology implemented in acquisition studies, the picture-sentence matching task, might introduce the same interference in matching the syntax of two stimuli considered simultaneously or in sequence. This would be due to the fact that children have to encode the actions represented by the picture in order to match them to the sentence they are simultaneously listening to. If indeed it is the case that sentences are encoded in an active-like form (Anderson, 1974), this would create interference between the active syntax of the sentence encoding the action of the picture and the passive syntax of the heard sentence. In other words, poor performance on passive sentences in acquisition studies might not reflect a delayed acquisition of passivization, but could rather be caused by interference introduced by the task. This would be confirmed by the fact that young children produce and understand passives (short and long) as well as actives, when tested using a different methodology (e.g., syntactic priming). The same reasoning could be applied to studies on aphasic patients that implemented the

sentence-picture matching task and found an overall difficulty in passives in these subjects. Alternatively, it is possible that aphasic patients rely more on heuristics such as word-order in processing sentences. In this case, they would present selective difficulties in comprehension when the sentence structure does not abide by the most frequent word order, as in the case of passive sentences.

Overall, our model argues that passives are easier to process than actives due to their multiple morphological cues. However, predicate semantics, in particular, whether a predicate requires revision or coercion in order to be interpreted as a verbal passive, can affect late stages of processing, e.g., re-reading. Moreover, the late effects of revision and coercion can increase errors in comprehension of passivized statives. Finally, post-interpretative processes can interact with passivization. In particular, if the task targets post-interpretative processes, such as memory retrieval, and the design introduces interference at memory retrieval, for example by presenting only active comprehension questions after passive sentences, then the comprehension of passive sentences is affected by a possible interference between the memory of the parse, with its passive syntax and theta-role assignment and the active syntax and theta-roles of the comprehension question.

5.4 Limitations in the present experiments.

As described in Chapter 3, our acceptability studies did not find an interaction effect between passivization and predicate type. This was contrary both to my expectations, which were based on previous theoretical and psycholinguistic findings (Belletti & Rizzi, 1988; Gehrke & Grillo, 2009; Grodzinsky, 1985; Maratsos et al., 1985; O'Brien et al., 2006; Snyder & Hyams, 2015; Volpato et al., 2015), and to the results reported by Ambridge and colleagues (2016), which found an interaction between passivization and predicate type. Although I argued that the absence of an interaction effect in the acceptability data is due to the fact that acceptability judgments are not sensitive to reanalysis and coercion effects, I will consider alternative methodological explanations here, which possibly represent the first limitation of these studies.

Firstly, I argued that several potential issues in the design of the study by Ambridge and colleagues (2016) might have hindered their conclusions. Their study did not contain any filler items, which means that the study did not contain any

ungrammatical sentences (beyond non-passivizable passives), thus artificially inflating differences between sentences when subjects are required to use the whole judgment scale in accordance with the instructions, but no anchor for the lowest point (i.e., an ungrammatical item) is provided, thus maximising scale bias across participants (Schütze & Sprouse, 2013). However, the absence of ungrammatical items might make participants more sensitive to differences in grammatical structures and allow an interaction between passivization and predicate type to be observed. Conversely, our ungrammatical fillers might have prevented us from observing this effect.

Secondly, it is important to underline that our instructions specifically asked participants to ignore plausibility when rating the sentences. However, the difficulty in passivizing stative verbs is actually related to the verb meaning, beside their grammatical structure. Our instructions might have thereby made participants less sensitive to an interaction effect.

Overall, the numerous differences between our design and the design of Ambridge and colleagues' study do not allow to conclusively determine the reason for the discrepancy between our results. Future studies are required to assess which of the above factors explain the conditions under which an interaction between predicate type and voice emerges in acceptability judgment ratings and hence its basis.

The second limitation of my conclusions is represented by the fact that my account of passive sentence processing is based in part on very robust findings, e.g., the self-paced reading data, which consistently found passives to be faster than actives, indicating that they are not more complex to process than actives. On the other hand, the interaction between predicate type and syntax, which constitutes another important building block of my account, was only found in two experiments, Exp. 7 and 8, and no further studies were conducted to possibly replicate this finding. It would thus be ideal to replicate these findings to strengthen the support to my model. These replications could use the same materials of Exp. 7 and 8, while manipulating predicate type within rather than between subjects. This would provide a cleaner test for an interaction between passivization and predicate type, as inter-subject variability would be reduced.

5.5 Future studies.

The most natural extension of this study would be a replication of Exp. 7, i.e., a new acceptability study coupled with comprehension questions. In this new study, the syntax of the comprehension questions could actually be manipulated, rather than simply balanced. This would allow to statistically analyse the role of interference at memory retrieval in sentence comprehension. A further contrast should hence be added: match (the syntax of the target sentence coincides with the syntax of the question, e.g., active sentence – active question, passive sentence – passive question) vs. mismatch (the syntax of the target sentence does not coincide with the syntax of the question, e.g., active sentence – passive question, passive sentence – active question). If indeed a mismatch between the syntax of the target sentence and the syntax of the comprehension question creates interference at retrieval, then an effect of match vs. mismatch should be observed, independent of the syntax of the target sentence, i.e., answering a passive question targeting an active sentence should be as difficult as answering an active question targeting a passive sentence, and both would be more difficult than answering an active question targeting an active sentence and answering a passive question targeting a passive sentence. This finding would also allow to replicate the results of Exp. 7, i.e., no main effect of syntax, indicating that passives are not more errorful than actives by default, and a significant interaction between passivization and predicate type indicating that passives of statives are more difficult to process than their active counterparts, while no difference emerges across voice in eventives.

Secondly, it might be interesting to replicate Ambridge and colleagues' design, i.e., running an acceptability judgment study with no fillers or ungrammatical items, but using the predicates tested in the present study. Should an interaction between passivization and predicate type emerge even using our predicates, then it should be concluded that Ambridge et al.'s design characteristics, i.e., absence of fillers or ungrammatical items, might have enlarged differences across predicate types that would otherwise become insignificant when the scale is actually balanced between ungrammatical and grammatical end-points. If, on the contrary, no interaction emerged as per our acceptability studies, then this would indicate that Ambridge and colleagues might have used a subset of predicates that generates an effect on acceptability. Such

an outcome would require further investigation into the specific properties of those predicates.

In additionally refining our understanding of the interaction between passivization and predicate type, a possible goal of future studies could be to disentangle the effects of reanalysis and coercion in passivizing states. For example, it would be interesting to modify the stimuli used in the current study by removing the by-phrases. This should eliminate the effect of reanalysis in states as the absence of a by-phrase would not trigger reanalysis. However, a possible complexity effect of coercion should still be visible in the comparison between states and eventives as the former require coercion in order to passivize while the latter do not. Hence, any complexity effect observed in such a design could be attributed to coercion independent of reanalysis.

Finally, this study could be further enriched beyond the natural methodological extensions described above. For example, it would be interesting to target possible cross-linguistic differences, e.g., by investigating a language like Chinese in which the active and passive structure are signalled by an equal number of morphological cues, to verify whether indeed the facilitation observed in passive processing is caused by a high level of expectation/low level of surprisal for the upcoming syntactic category due to a richer morphology in passives with respect to actives.

Moreover, it would be worth refining our understanding of how predicate structure interacts with passivization, beyond the broad eventive vs. stative distinction explored here, by testing other types of predicates, e.g. object-experiencers. Our experiments tested subject-experiencers (e.g., “admire”), which promote an unaffected argument into subject position when they are passivized (“The girl” in (54) is not necessarily affected by – or even aware of – the boy’s admiration). Contrary to subject-experiencers, object-experiencers promote an affected argument to subject position when they are passivized (“The girl” in (55) is the participant actually experiencing fear). This would allow to test the role of semantic affectedness in modelling the interaction between passivization and predicate type.

(54) The girl is admired by the boy.

(55) The girl is frightened by the boy.

Finally, future studies could investigate how processing in healthy adults might differ from processing in other populations (e.g., L2 speakers, aphasic patients) with respect to factors that affect passives processing as explored in this study. For example, L2 speakers are considered to be more dependent on superficial, "good-enough" processing (Lim & Christianson, 2013). Their performance on the tasks used in our experiments might hence differ from L1 speakers, e.g., it might be possible to find evidence for a revision effect in passives in late measures of processing, or passives to be more errorful than actives independent of predicate type, even when the syntax of the questions is balanced. Furthermore, this more detailed account of passive processing could help explain the variability found in the aphasia literature with respect to a selective difficulty of aphasic patients in processing passive sentences (e.g., Cho & Thompson, 2010 vs. Zimmerer, Dąbrowska, Romanowski, Blank & Varley, 2014). For example, it might be interesting to investigate how characteristics of the task, such as the ones identified in our experiments, or of the passivized predicates might modulate the difficulty found in aphasics' comprehension of passives.

5.6 Broader implications for the psycholinguistic field.

The results collected in this study showed that passives are not more complex than actives and identified factors that can selectively affect passive sentence processing, i.e., predicate semantics, as well as an interesting interaction between task characteristics and passivization in post-interpretative processes.

These results indicate a problem for studies that have used passives as a metric of syntactic complexity across various populations. As mentioned in Chapter 1, the passive-active contrast has often been employed as a metric of syntactic complexity to assess syntactic processing abilities across various populations. This contrast has in fact been used to identify areas of syntactic processing in the brain of healthy adults (e.g., Mack et al., 2013), to establish patterns of impairment in syntactic comprehension of aphasic patients (e.g., Caplan et al., 2007; Thothathiri et al., 2012) and to investigate the development of semantic and syntactic abilities in children acquiring their first language (e.g., Maratsos et al., 1985). However, the data collected in this study now indicate that these studies are based on a false notion, i.e., the idea that passives are across-the-board more complex than actives. Instead, these results show that the complexity in

comprehension of passive sentences is due to their interaction with predicate semantics, i.e., whether the passivized predicate is eventive or stative, and to its interaction with post-interpretative processes. In particular, tasks that involve memory retrieval and whose design introduce interference at retrieval for passive sentences specifically (e.g., a verification task with active comprehension questions) might artificially introduce a complexity in passives comprehension with respect to active sentences, unrelated to the difference in syntactic structure. Hence, studies that adopted the active-passive contrast as a metric of syntactic complexity might need to revise their design (and possibly conclusions) in light of these findings, by looking at whether they tested stative predicates or whether their task might have introduced interference in passives selectively (e.g., a verification task with active questions only). Moreover, future studies targeting passive sentences comprehension should henceforth be aware of possible task-related effects when using specific methodologies and should also consider predicate properties when designing their stimuli.

More generally, my study has shown the importance of using different methodologies to address the same question. This allows to approach a problem from different angles and, as a result, to delineate a more comprehensive picture of the structure of interest. This approach also protects against the risk of false positive or negative results, as consistent outcomes across methodologies are less likely to be artefacts of a specific design. Finally, the results obtained from different methodologies also allow to better understand strengths and limitations of each design, particularly with respect to its sensitivity and power in detecting the effects under investigation.

In particular, with regards to online methodologies, my study has shown that self-paced reading is sensitive to early processing effects, such as differences in expectation/surprisal levels. However, it is not sensitive to effects arising in late stages of parsing (e.g., re-reading or regression), such as revision or coercion. Conversely, these effects are captured by eye-tracking while reading, specifically by measures like re-reading, regression path and total time. Studies employing this methodology should however be alert to possible word-skipping confounds, especially if the difference across conditions is comprised of short, highly frequent words, as in the case of the active-passive contrast (the cues that facilitate processing of passive sentences are the auxiliary and “by”). Importantly, these confounds only affect early measures of processing,

implying that predictions with respect to effects occurring in late stages of processing are not impacted.

With respect to offline methodologies, my study has shown that acceptability judgments are not sensitive to differences in expectation/surprisal levels, as they are formed on the grammaticality of the sentence in its entirety. Moreover, when investigating a possible interaction between syntax and meaning (as in the case of passivization and predicate semantics), it might be tricky to instruct participants to separate plausibility effects from acceptability ones. If indeed the instructions ask participants to ignore any plausibility effect, the task might become less sensitive to effects caused by differences in semantic properties. On the other hand, comprehension questions appear to be sensitive to overall complexity effects arising from an interaction between syntax and semantics, such as the one investigated in this study between passivization and predicate type. However, these effects can only arise once task characteristics are equated across conditions. In particular, comprehension questions characteristics (e.g., their syntax) should not facilitate/affect one condition over the other(s).

Future psycholinguistic study should hence, whenever possible, adopt triangulation, i.e., use multiple methodologies to target the same question, in order to gain a more comprehensive understanding of the phenomenon under investigation and its possible interaction with task characteristics, as well as more insight into the strength and limitations of the methodologies used.

5.7 Conclusions.

The present study focused on passive sentence processing in healthy British English native speakers. Although the literature generally agreed in considering passives more difficult to process than actives, the results collected in this study showed that passives are not more complex than actives. On the contrary, their processing is facilitated by their dedicated morphology.

However, the experiments reported in this study also identified factors that *can* selectively affect passive sentence processing, i.e., predicate semantics, as well as interesting task characteristics that affect post-interpretative memory mechanisms.

These findings are problematic for existing theories of passives comprehension and advocate for a combined account of passive sentence processing, based on expectation-/surprisal-based models as well as on models that include task-based effects in accounting for sentence comprehension phenomena. Moreover, these results are problematic for studies that have used passives as a metric of syntactic complexity across various populations, whose methodologies (and possibly conclusions) should hence be revised in light of the data presented in this study. Finally, these findings demonstrate the importance of adopting triangulation, i.e., the use of different methodologies to tackle the same experimental question, in order to avoid the risk of false positive or negative results, as well as to gain further knowledge about psycholinguistic methodologies and their sensitivity to linguistic phenomena.

Future work will need to replicate some of these results, as well as expand this study by refining our understanding of how predicate structure interacts with passivization and by extending the investigation cross-linguistically, as well as across populations (e.g., L2 speakers).

6 References

Ambridge, B., Bidgood, A., Pine, J.M., Rowland, C.F., & Freudenthal, D. (2016). Is Passive Syntax Semantically Constrained? Evidence From Adult Grammaticality Judgment and Comprehension Studies. *Cognitive Science*, *40*, 1435–1459.

Anderson, J.R (1974). Verbatim and Propositional Representation of Sentences in Immediate and Long-Term Memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 149-162.

Angele, B. & Rayner, K. (2013). Processing the in the Parafovea: Are Articles Skipped Automatically? *Journal of Experimental Psychology-Learning Memory and Cognition*, *39*, 649-662.

Avrutin, S. (2006). Weak syntax. In Y. Grodzinsky, & K. Amunts (Eds.), *Broca's region* (49–62). New York, NY: Oxford University Press.

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*, 255-278.

Belletti, A. & Rizzi, L. (1988). Psych Verbs and Theta Theory. *Natural Language and Linguistic Theory*, *6*, 291-352.

Belsley, D. (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. New York, NY: Wiley.

Bever, T.G. (1970) The cognitive basis for linguistic structures. In J.R. Hayes (Ed.), *Cognitive and the Development of Language* (279-362). New York, NY: Wiley.

Bogal-Allbritten, E. (2014). Slightly coerced: Processing evidence for adjectival coercion by minimizers. *Proceedings of the 48th Meeting of the Chicago Linguistic Society (CLS 48)*. Chicago Linguistic Society, Chicago.

Boland, J.E., & Blodgett, A. (2002). Eye Movements as a Measure of Syntactic and Semantic Incongruity in Unambiguous Sentences. Unpublished manuscript, University of Michigan, Ann Arbor, MI.

Borer, H., & Wexler, K. (1987). The Maturation of Syntax. In T. Roeper & E. Williams (Eds.), *Parameter setting* (123-172). Dordrecht: Reidel.

Boyle, W., Lindell, A. K., & Kidd, E. (2013). Investigating the Role of Verbal Working Memory in Young Children's Sentence Comprehension. *Language Learning*, *63*, 211–242.

Bresnan, J. (1982). The passive in lexical theory. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations* (3–86). Cambridge, Massachusetts: The MIT Press.

Caplan, D., Dede, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging*, *26*, 439-450.

Caplan, D., Michaud, J., Hufford, R., & Makris, N. (2016). Deficit-lesion correlations in syntactic comprehension in aphasia. *Brain and Language*, *152*, 14-27.

Caplan, D., & Waters, G. (2003). On-line syntactic processing in aphasia: studies with auditory moving window presentation. *Brain and Language*, *84*, 222–249.

Caplan, D., & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77–126.

Caplan, D., Waters, G., DeDe, G., Michaud, J., & Reddy, A. (2007). A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language*, *101*, 103–150.

Caramazza, A., & Zurif, E.B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language*, *3*, 572–582.

Carrithers, C. (1989). Syntactic complexity does not necessarily make sentences harder to understand. *Journal of Psycholinguistic Research*, *18*, 75-88.

Cho, S., & Thomspon, C.K. (2010). What Goes Wrong during Passive Sentence Production in Agrammatic Aphasia: An Eyetracking Study. *Aphasiology*, *12*, 1576–1592.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Christianson, K., Luke, S.G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 538-544.

Collins, C. (2005). A Smuggling Approach to the Passive in English. *Syntax, 8*, 81–120,

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450-466.

DeDe, G., Caplan, D., Kemtes, K., & Waters, G. (2004). The Relationship between Age, Verbal Working Memory, and Language Comprehension. *Psychology and Aging*, *19*, 601-616.

Demuth, K., Moloi, F. & Machobane, M. (2010). 3-year-olds' comprehension, production, and generalization of Sesotho passives. *Cognition*, *115*, 238-251.

Den Ouden, D. (2006). *Neural networks for syntactic processing: lateralization of narrow and broad syntax*. Paper presented at 7th Science of Aphasia Conference, Neurocognition of Language: Innovative Aspects, Sardinia, Italy.

Dickey, M., & Thompson, C.K. (2009). Automatic processing of wh- and NP-movement in agrammatic aphasia: Evidence from eyetracking. *Journal of Neurolinguistics*, *22*, 563–583.

Evans, W.S., Caplan, D., Ostrowski, A., Michaud, J., Guarino, A.J., & Waters, G. (2014). Working memory and the revision of syntactic and discourse ambiguities. *Canadian Journal of Experimental Psychology*, *69*, 136-155.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, *54*, 541-553.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.

Ferreira, F. & Christianson, K. (2016). Is Now-or-Never Language Processing Good Enough? *Behavioral and Brain Sciences*, *39*, 1-72.

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 555-568.

Ferreira, F. & Patson, ND. (2007). The 'Good Enough' approach to language comprehension. *Language and Linguistics Compass*, *1*, 71-83.

Finley, A.J., & Penningroth, S.L. (2015). Online versus In-lab: Pros and Cons of an Online Prospective Memory Experiment. In A.M. Columbus (Ed.), *Advances in Psychology Research* (135-162). Hauppauge, NY: NOVA.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178-210.

Garnham, A., & Oakhill, J. (1987). Interpreting elliptical verb phrases. *Quarterly Journal of Experimental Psychology, 39*, 611–625.

Gehrke, B., & Grillo, N. (2007). Aspects on passives. In S. Blaho, L. Vicente & E. Schoorlemmer (Eds.) *Proceedings of Console XIV* (121-141). Leiden: SOLE.

Gehrke, B., & Grillo, N. (2009). How to become passive. In K. Grohmann (Ed.), *Exploration of Phase Theory: features, arguments, and interpretation at the Interfaces* (231-268). Berlin: Mouton de Gruyter.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*, 1-76.

Gordon, P., & Chafetz, J. (1990). Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition*, *36*, 227-254.

Gordon, P.C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology*, *27*, 1-13.

Grillo, N. (2008). *Generalized minimality: Syntactic underspecification in Broca's aphasia* (Doctoral dissertation).

Grodner, D.J., & Gibson, E.A.F. (2005). Consequences of the Serial Nature of Linguistic Input for Sentential Complexity. *Cognitive Science*, *29*, 261–291.

Grodzinsky, Y. (1990). *Theoretical perspectives on language deficits*. Cambridge, MA: MIT Press.

Grodzinsky, Y. (1995). Trace Deletion, Θ-roles, and Cognitive Strategies. *Brain and Language*, *51*, 469-497.

Grodzinsky, Y. (2000). The neurology of syntax: language use without Broca's area. *Behavioral and Brain Sciences, 23*, 1–71.

Haegeman, L. (1991). *Introduction to Government and Binding theory*. Oxford & Cambridge, MA: Basil Blackwell.

Hakes, B., Evans, J., & Brannon, L. (1976). Understanding sentences with relative clauses. *Memory and Cognition*, *4*, 283–296.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In L. Levin (Ed.), *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics* (159–166). Pittsburgh, PA.

Ito, A., Martin, A.E., & Nieuwland, M.S. (2016). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, *32*, 1-12.

Jackson, C.N., Dussias, P.E., Hristova, A. (2012). Using eye-tracking to study the on-line processing of case-marking information among intermediate L2 learners of German. *International review of applied linguistics in language teaching*, *50,* 101-133.

Jaeggi, S.M., Buschkuehl, M., Jonides, J., & Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6829-6833.

Jaeggi, S.M., Buschkuehl, M., Perrig, W.J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*, 394-412.

Johns, B.T., & Jones, M.N. (2015). Generating Structure from Experience: A Retrieval-Based Model of Language Processing. *Canadian Journal of Experimental Psychology*, *69*, 1196-1961.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Just, M.A., Carpenter, P.A., & Woolley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238.

Kamide, Y., Scheepers, C., & Altmann, G.T.M. (2003). Integration of Syntactic and Semantic Information in Predictive Processing: Cross-Linguistic Evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37-55.

Kane, M.J. (2005). Full frontal fluidity? Looking in on the neuroimaging of reasoning and intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (141-163). Thousand Oaks, CA: Sage.

Kane, M.J., & Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*, 637-671

Karimi, K., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology, 69*, 1013-1040.

Keenan, E., & Dryer, M. (2007). Passive in the world's languages. In T. Shopen (Ed.), *Syntactic Typology and linguistic description, Volume 1: Clause structure, 2nd ed.* (325-361). Cambridge: Cambridge University Press.

Kim, J.H., & Christianson, K. (2013). Sentence Complexity and Working Memory Effects in Ambiguity Resolution. *Journal of Psycholinguistic Research*, *42*, 393-411.

Kiparsky, P (2013). Towards a null theory of the passive. *Lingua, 125*, 7-33.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*, 627-645.

Levy, R. (2007). Expectation-based syntactic comprehension. *Cognition*, *106*, 1-53.

Lim, J.H., & Christianson, K. (2013). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition*, *16*, 518-537.

MacDonald, M.C. (2013) How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*: 226.

Mack, J.E., Meltzer-Asscher, A., Barbieri, E., & Thompson, C.K. (2013). Neural correlates of processing passive sentences. *Brain Sciences*, 3, 1198-1214.

Manetti, C. (2013). On the production of passives in Italian: evidence from an elicited production task and a syntactic priming study with preschool children. In S. Baiz, N. Goldman & R. Hawkes (Eds.), *BUCLD 37 Online Proceedings Supplement* (1-16). Boston, MA: Boston University.

Maratsos. M.P. and Abramovitch, R. (1975) How children understand full, truncated and anomalous passives. *Journal of Verbal Learning and Verbal Behavior*, *14*, 145-157.

Maratsos, M.P., Fox, E.C., Becher, J., & Chalkley, M.A. (1985). Semantic restrictions on children's passives. *Cognition*, *19*, 167–191.

McElree, B., Traxler, M., Pickering, M., Seely, R., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, *78*, 17–25.

McIntyre, A. (2006). *The integration of directional PPs: Thoughts on the way to getting towards knowing if I'm (be)coming or going*. Paper presented at the conference 'Syntax and semantics of spatial P' in Utrecht.

McIntyre, A. (2013). Adjectival passives and adjectival participles in English. In A. Alexiadou & F. Schäfer (Eds.), *Non-canonical passives* (21–42). Amsterdam: John Benjamins.

Messenger, K. (2009). *Syntactic priming and children's production and representation of the passive* (Doctoral dissertation).

Messenger, K., Branigan, H.P., McLean, J.F., & Sorace, A. (2012). Are children's early passives semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language, 66*, 568-587.

Mitchell, D.C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D.E. Kieras & M.A. Just (Eds.), *New Methods in Reading Comprehension Research* (69-89). Hillsdale, N.J: Erlbaum.

O'Brien, K., Grolla, E., & Lillo-Martin, D. (2006). Long passives are understood by young children. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (441-451). Somerville, TX: Cascadilla Press.

Osterhout, L., & Swinney, D. (1993). On the temporal course of gap-filling during comprehension of verbal passives. *Journal of Psycholinguistic Research*, *22*, 273-286.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pinker, S., Lebeaux, D. S., & Frost, L.A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition*, *26*, 195–267.

Reali, F., & Christiansen, M.H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*, 1–23.

Roberts, R., & Gibson, E. (2002). Individual differences in working memory. *Journal of Psycholinguistic Research, 31*, 573–598.

Roth, H. L., Lora, A. N., & Heilman, K. M. (2002). Effects of monocular viewing and eye dominance on spatial attention. *Brain*, *125*, 2023–2035.

Saffran, E.M., Schwartz, M.F., & Linebarger, M.C. (1998). Semantic Influences on Thematic Role Assignment: Evidence from Normals and Aphasics. *Brain and language, 62*, 255–297.

Santi, A., & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage*, *51*, 1285-1293.

Schotter, E.R., Angele, B. & Rayner, K. (2012). Parafoveal processing in reading. *Attention Perception & Psychophysics*, *74*, 5-35.

Schütze, C.T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Schütze, C.T. & Sprouse, J. (2013). Judgment data. In R.J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (27–50). New York: Cambridge University Press.

Shanks, D.R., Vadillo, M.A., Riedel, B., Clymo, A., Govind, S., Hickin, N., Tamman, A.J., & Puhlmann, L.M. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General*, *144*, 142-158.

Snyder, W. & Hyams, N. (2015). Minimality effects in children's passives. In E. Di Domenico, C. Hamann & S. Matteini (Eds.) *Structures, Strategies and Beyond* (343-368). Amsterdam: John Benjamins.

Sprouse, J. (2008). The Differential Sensitivity of Acceptability Judgments to Processing Effects. *Linguistic Inquiry*, *39*, 686-694.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working-memory capacity and syntactic island effects. *Language*, *88*, 401-407.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*, 71-86.

Street, J.A., & Dąbrowska, E. (2010). Lexically-specific knowledge and individual differences in adult native speakers' processing of the English passive. *Applied Psycholinguistics*, *35*, 97-118.

Swets, B., Desmet, T., Hambrick, D.Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology*, *136*, 64–81.

Tenny, C. (1987). *Grammaticalizing aspect and affectedness* (Doctoral dissertation).

Thothathiri, M., Kim, A., Trueswell, J.C., & Thompson-Schill, S.L. (2012). Parametric effects of syntactic-semantic conflict in Broca's area during sentence processing. *Brain and Language*, *120*, 259-264.

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.

Traxler, M.J., Corina, D.P., Morford, J.P., Hafer, S., & Hoversten, L.J. (2014). Deaf readers' response to syntactic complexity: Evidence from self-paced reading. *Memory & Cognition*, *42*, 97–111.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, *47*, 530–547.

Van Dyke, J.A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55,* 157–166.

van Herten, M., Chwilla, D.J., & Kolk, H.H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*, 1181-1197.

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*, 1176-1190.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, *66*, 143-160.

Volpato, F., Verin, L., & Cardinaletti, A. (2015). The comprehension and production of verbal passives by Italian preschool-age children. *Applied Psycholinguistics*, *37*, 901-931.

Zimmerer, V.C., Dąbrowska, E., Romanowski, C.A., Blank, C., & Varley, R.A. (2014). Preservation of passive constructions in a patient with primary progressive aphasia. *Cortex*, *50*, 7-18.

7 Appendices

7.1 Experimental items in Exp. 1.

1	Passive	The guitarist was rejected by the attractive and talented singer in the concert hall next to the Irish pub.
	Progressive	The guitarist was rejecting the attractive and talented singer in the concert hall next to the Irish pub.
	Active	The guitarist rejected the attractive and talented singer in the concert hall next to the Irish pub.
2	Passive	The caterer was paid by the creative and competent chef in the newly refurbished kitchen opposite the big cold-store.
	Progressive	The caterer was paying the creative and competent chef in the newly refurbished kitchen opposite the big cold-store.
	Active	The caterer paid the creative and competent chef in the newly refurbished kitchen opposite the big cold-store.
3	Passive	The apprentice was injured by the naïve and clumsy electrician in the darkened room adjacent to the power line.
	Progressive	The apprentice was injuring the naïve and clumsy electrician in the darkened room adjacent to the power line.
	Active	The apprentice injured the naïve and clumsy electrician in the darkened room adjacent to the power line.
4	Passive	The detective was harassed by the loud and hostile witness in the front lobby next to the glass lift.
	Progressive	The detective was harassing the loud and hostile witness in the front lobby next to the glass lift.
	Active	The detective harassed the loud and hostile witness in the front lobby next to the glass lift.
5	Passive	The producer was sued by the untalented and pretentious artist in the legal offices next to the cinema studios.
	Progressive	The producer was suing the untalented and pretentious artist in the legal offices next to the cinema studios.

	Active	The producer sued the untalented and pretentious artist in the legal offices next to the cinema studios.
6	Passive	The chef was stabbed by the aggressive and insubordinate waiter in the restaurant kitchen next to the hotel foyer.
	Progressive	The chef was stabbing the aggressive and insubordinate waiter in the restaurant kitchen next to the hotel foyer.
	Active	The chef stabbed the aggressive and insubordinate waiter in the restaurant kitchen next to the hotel foyer.
7	Passive	The general was abandoned by the deceptive and cowardly soldier in the vast forest adjacent to the military campsite.
	Progressive	The general was abandoning the deceptive and cowardly soldier in the vast forest adjacent to the military campsite.
	Active	The general abandoned the deceptive and cowardly soldier in the vast forest adjacent to the military campsite.
8	Passive	The politician was injured by the jealous and resentful waitress in the large bathroom next to the crowded press-room.
	Progressive	The politician was injuring the jealous and resentful waitress in the large bathroom next to the crowded press-room.
	Active	The politician injured the jealous and resentful waitress in the large bathroom next to the crowded press-room.
9	Passive	The lawyer was corrupted by the nervous and dishonest judge in the famous cocktail bar opposite the County courthouse.
	Progressive	The lawyer was corrupting the nervous and dishonest judge in the famous cocktail bar opposite the County courthouse.
	Active	The lawyer corrupted the nervous and dishonest judge in the famous cocktail bar opposite the County courthouse.
10	Passive	The student was harassed by the misogynistic and racist professor in the science laboratory next to the language department.
	Progressive	The student was harassing the misogynistic and racist professor in the science laboratory next to the language department.
	Active	The student harassed the misogynistic and racist professor in the science laboratory next to the language department.

11	Passive	The salesman was hired by the experienced and competent associate in the colourful hotel lounge opposite the guest library.
	Progressive	The salesman was hiring the experienced and competent associate in the colourful hotel lounge opposite the guest library.
	Active	The salesman hired the experienced and competent associate in the colourful hotel lounge opposite the guest library.
12	Passive	The director was rejected by the charming and charismatic actress in the café on the corner near the theatre.
	Progressive	The director was rejecting the charming and charismatic actress in the café on the corner near the theatre.
	Active	The director rejected the charming and charismatic actress in the café on the corner near the theatre.
13	Passive	The pilot was stabbed by the bald and overweight steward in the luxury airport lounge opposite the unisex toilets.
	Progressive	The pilot was stabbing the bald and overweight steward in the luxury airport lounge opposite the unisex toilets.
	Active	The pilot stabbed the bald and overweight steward in the luxury airport lounge opposite the unisex toilets.
14	Passive	The sculptor was mugged by the strange and temperamental photographer in the art gallery next to the book shop.
	Progressive	The sculptor was mugging the strange and temperamental photographer in the art gallery next to the book shop.
	Active	The sculptor mugged the strange and temperamental photographer in the art gallery next to the book shop.
15	Passive	The protester was attacked by the aggressive and stocky policeman in the crowded street near the old town centre.
	Progressive	The protester was attacking the aggressive and stocky policeman in the crowded street near the old town centre.
	Active	The protester attacked the aggressive and stocky policeman in the crowded street near the old town centre.
16	Passive	The nun was saved by the wise and respected teacher in the rose garden outside the old Catholic church.

	Progressive	The nun was saving the wise and respected teacher in the rose garden outside the old Catholic church.
	Active	The nun saved the wise and respected teacher in the rose garden outside the old Catholic church.
17	Passive	The nurse was seduced by the handsome and distinguished doctor in the cosy staff kitchen on the seventh floor.
	Progressive	The nurse was seducing the handsome and distinguished doctor in the cosy staff kitchen on the seventh floor.
	Active	The nurse seduced the handsome and distinguished doctor in the cosy staff kitchen on the seventh floor.
18	Passive	The scientist was rescued by the brave and reckless reporter in the chemical analysis laboratory near the great fire.
	Progressive	The scientist was rescuing the brave and reckless reporter in the chemical analysis laboratory near the great fire.
	Active	The scientist rescued the brave and reckless reporter in the chemical analysis laboratory near the great fire.
19	Passive	The waitress was seduced by the fascinating and flirtatious model in the Chinese restaurant opposite the multi-brand fashion showroom.
	Progressive	The waitress was seducing the fascinating and flirtatious model in the Chinese restaurant opposite the multi-brand fashion showroom.
	Active	The waitress seduced the fascinating and flirtatious model in the Chinese restaurant opposite the multi-brand fashion showroom.
20	Passive	The babysitter was paid by the mature and trustworthy pharmacist in the office at the back of the shop.
	Progressive	The babysitter was paying the mature and trustworthy pharmacist in the office at the back of the shop.
	Active	The babysitter paid the mature and trustworthy pharmacist in the office at the back of the shop.
21	Passive	The industrialist was sued by the angry and resentful client in the legal services office opposite the company building.

	Progressive	The industrialist was suing the angry and resentful client in the legal services office opposite the company building.
	Active	The industrialist sued the angry and resentful client in the legal services office opposite the company building.
22	Passive	The mechanic was fired by the ill-mannered and annoying auto-electrician in the spacious garage next to the congested highway.
	Progressive	The mechanic was firing the ill-mannered and annoying auto-electrician in the spacious garage next to the congested highway.
	Active	The mechanic fired the ill-mannered and annoying auto-electrician in the spacious garage next to the congested highway.
23	Passive	The suspect was shocked by the astute and perceptive detective in the dark interrogation room inside the police station.
	Progressive	The suspect was shocking the astute and perceptive detective in the dark interrogation room inside the police station.
	Active	The suspect shocked the astute and perceptive detective in the dark interrogation room inside the police station.
24	Passive	The salesgirl was pushed by the nervous and hysterical client in the shopping mall next to the tube station.
	Progressive	The salesgirl was pushing the nervous and hysterical client in the shopping mall next to the tube station.
	Active	The salesgirl pushed the nervous and hysterical client in the shopping mall next to the tube station.
25	Passive	The child was slapped by the young and skinny babysitter in the colourful nursery inside the big country-style house.
	Progressive	The child was slapping the young and skinny babysitter in the colourful nursery inside the big country-style house.
	Active	The child slapped the young and skinny babysitter in the colourful nursery inside the big country-style house.
26	Passive	The soldier was kidnapped by the armed and dangerous rebel in the dark cave next to the combat zone.
	Progressive	The soldier was kidnapping the armed and dangerous rebel in the dark cave next to the combat zone.

	Active	The soldier kidnapped the armed and dangerous rebel in the dark cave next to the combat zone.
27	Passive	The student was rescued by the aged and revered tutor near the collapsed basement inside the old humanities library.
	Progressive	The student was rescuing the aged and revered tutor near the collapsed basement inside the old humanities library.
	Active	The student rescued the aged and revered tutor near the collapsed basement inside the old humanities library.
28	Passive	The dancer was slapped by the oppressive and abusive choreographer in the messy dressing room of the renowned theatre.
	Progressive	The dancer was slapping the oppressive and abusive choreographer in the messy dressing room of the renowned theatre.
	Active	The dancer slapped the oppressive and abusive choreographer in the messy dressing room of the renowned theatre.
29	Passive	The investigator was bribed by the unpleasant and bald bartender in the crowded pub next to the crime scene.
	Progressive	The investigator was bribing the unpleasant and bald bartender in the crowded pub next to the crime scene.
	Active	The investigator bribed the unpleasant and bald bartender in the crowded pub next to the crime scene.
30	Passive	The climber was abandoned by the unwise and incompetent guide in the disused shelter on the rugged mountain top.
	Progressive	The climber was abandoning the unwise and incompetent guide in the disused shelter on the rugged mountain top.
	Active	The climber abandoned the unwise and incompetent guide in the disused shelter on the rugged mountain top.

7.2 Filler items in Exp. 1, 2, 3 and 4.

1 The teacher took the car instead of the express train due to the previously announced public transport strike.

2 The researcher presented his most recent work to the commission and obtained very positive comments regarding the experimental design.

3 The brilliant student easily solved complicated logarithmic equations during the math lecture but the distracted student did not.

4 The judge asked for an explanation of the events and the accused answered in a complete and sincere way.

5 The artist created a controversial representation of contemporary society and exhibited it in the most popular square in the city.

6 The producer announced the release of the new thriller based on the best-selling novel by the famous Chinese author.

7 The dancer practised intensely for several days in the ball room and participated in the International Street Dancers Competition.

8 The police reported several details regarding the severe accident on the high street near the crossroads and the media did too.

9 The art student participated in a sculpture class and exhibited her creations in a collective show at the university.

10 The mathematician solved the difficult problem and obtained the most famous and prestigious prize in his discipline.

11 The politician announced his candidacy for the presidential election and answered several questions from the media at the conference.

12 The dangerous rebels attacked the school in the isolated area and all the students screamed in terror.

13 The beautiful princess married her young and handsome chauffeur and shocked the royal family and the press.

14 The brave and strong lifeguard rapidly swam towards the boat on fire and saved the old couple.

15 The big law firm began to fire employees in several departments due to an irreversible financial crisis.

16 The girl was kissed by her loving and handsome fiancé under a cherry tree in the scented garden.

17 The violinist was scolded by her tutor due to her lack of discipline during the intensive training for the concert.

18 The driver was scolded by his supervisor due to his consistently long delays in delivering their products.

19 The lady was admired by several young and bold soldiers along the path covered by yellow leaves in the park.

20 The princess was kissed by the handsome and valiant prince in the magical wood near the royal palace.

21 The debutante was presented to the distinguished and respectable members of the society during a thrilling event at the ballroom.

22 The swimmer was admired by her numerous fans for her excellent performance in the pool during the Olympic Games.

23 The child was taken to hospital by his apprehensive and worried mother due to his severe and long-lasting symptoms.

24 The ambassador was nominated Honorary Councillor by the president due to her important contribution to the cooperation between European nations.

25 The professor was unanimously elected as chairman of the Archaeology department by the board composed of distinguished scholars from important institutions.

26 The scientist was nominated for the Nobel Prize by the Norwegian committee for his ground-breaking studies on neuronal activity in rats.

27 The cheerleader was elected as Prom Queen by her schoolmates thanks to her stunning beauty and well-known kindness.

28 The heir to the rich and powerful family was kidnapped by criminals after school on Tuesday afternoon.

29 The rich and elegant old lady was mugged outside the jewellery store after a very expensive purchase.

30 The policeman was corrupted by the defence attorney during the investigation on the bloody and dreadful murder.

31 The policeman did not report the big fight between the rude and hostile neighbours in the quiet and tidy street.

32 The Japanese sculptor could not create a universally admired reproduction of the famous statue for the Ancient Culture Exhibition.

33 The teacher did not ask the lazy student for his attention during the lecture on morality in Ancient Egypt.

34 The politician did not practise her speech on the amendment of a law on public debate because of a cold.

35 The little girl did not play with her brother in the colourful playground next to their weedy garden.

36 The German striker could not score against the famous Italian football team of Milan for the entirety of his unsuccessful career.

37 The lady could not offer a well-paid job to the young unemployed woman at the shelter near the tube station.

38 The lawyer could not offer a good defence to his clients due to his lack of preparation for the trial.

39 The school director did not expel the student for his strange behaviour towards his classmates during the school trip.

40 The moderator did not expel the representative for his unorthodox response to a member of the opposing party.

41 The director did not fairly assign roles to the actors for the new Broadway production of the world-renowned play.

42 The butler did not assign tasks to the maids and waiters of the house in time to adequately prepare for the dinner party.

43 The senior manager could not hire new employees despite the urgent need of extra help for the upcoming event.

44 The young waitress could not push the heavy trolley with the pile of dirty plates toward the kitchen.

45 The district attorney did not bribe the juror during the controversial trial against the pharmaceutical company last year.

46 While the boy and the girl hugged the father appeared at the door of the tool shed next to the bushes.

47 While the quick and expert burglar and his accomplice fought the police appeared in front of the National bank.

48 While the neurologist of the highly ranked hospital in Chicago and the patient met the wife waited in the room.

49 While the student of the famous university in Florida and his parents saluted the chauffeur waited outside.

50 While the father and the mother cuddled the little girl played the piano in the living room during the Christmas party.

51 While the famous cellist and the singer dated the director gave an extraordinary performance at the Royal Albert Hall.

52 While the officer and the superintendent debated the new rule about safety showed the weak points in the chain of command.

53 While the rock guitarist and the agent hugged the sound technician loudly screamed in the soundproofed studio near the elevator.

54 While the recovering patient and the doctor saluted the nurse gave a leaflet to each participant in the staff meeting.

55 While the football player and the therapist dated the Dutch striker scored the final goal for the victory.

56 While the caring coach and the sweet lifeguard cuddled the little girl swam for the first time without armbands.

57 While the researcher and the director of the laboratory debated the machine showed a significant increase in the level of toxicity in the substance.

58 While the brilliant student and the history professor met the dean of the department organised the next staff meeting.

59 While the soldier and the captain fought the lieutenant organised the retreat from the line of fire behind the fortress.

60 While the foreign ambassador and the princess married the president began his speech about the reform of the National Security Agency.

7.3 Implausible items in the pre-norming plausibility task.

1 The celebrities interviewed the journalist and received very positive reviews from the readers of the magazine it was published in.

2 The illiterate painter spent two years in his chalet in the Alps and wrote out the material for his autobiography.

3 The patient discovered the past traumas the therapist had endured during his childhood at the beginning of treatment in the clinic.

4 The religious nun went to the striptease at the famous night cub in London with her enthusiastic sisters and the archbishop.

5 The villager haunted the ghost of the man who had previously lived in his home on the top of the hill.

6 The man scored twelve goals in the football match he refereed as a favour to his older brother.

7 The Roman Catholic priest married his hard working wife on a rainy day in Portofino during the spring.

8 The actor shaved the barber’s hair in preparation for his new role in the latest action film shot in Hollywood.

9 Santa Claus gave coal to all the children on the nice list and presents to those on the naughty list.

10 The audience entertained the circus performers travelling the country with their new show in the circus tent.

11 The law abiding police man was arrested by the criminal last Saturday on the town's busy high street.

12 The friar was discovered seducing the temptress by the priest in the old church with the cherry tree at its rear.

13 The dogs were herded by the sheep while the shepherd tried to move the flock away from the ferocious wolf.

14 The dragon was eaten by the prince trying to win the hand of the beautiful princess from the neighbouring kingdom.

15 The irate train conductor was removed from the train by the passenger on her way to see the West end show.

16 The paramedic was saved by the severely injured woman after the fire in her house started by her cat.

17 The scientists were experimented on by the mice in the immensely high tech laboratory in the university.

18 The children who were fed bread by the ducks lived in a quiet village in the north of England.

19 The jeweller who had stolen many diamond necklaces from the burglar usually went on holiday to his time share in Spain.

20 The cruel master who had escaped from his slave ran through the streets of ancient Rome towards the forum.

21 The demon who expelled the exorcist from the possessed room with his sliver cross knew the residents many sordid secrets.

22 The little boy who scared the clown at his birthday party received a remote controlled race car as a present.

23 The refugees who helped the philanthropist were happy to know that they were now citizens in a new country.

24 The child who rescued the lifeguard at the local beach had once spotted a dolphin in the water.

25 The waitress who the customer served always collected her tips from the tables at the end of her shift.

26 The tabby cat who the little mouse hunted lived in the house with a swing in its front garden.

27 The millionaire who the sweet old housekeeper hired to clean and tidy had great references from his previous employers

28 The dog who the cat chased was a husky who had once pulled sleighs through the snow for famous adventurers.

29 The rabbit who the little girl fed steak lived in a large yellow hutch in the garden.

30 The lord who served the chef his dinner only frequented the House of Lords to vote on big issues.

7.4 Comprehension questions in Exp. 1.

1	Active	Did the musician play the guitar?
	Passive	Did the musician play the guitar?
2	Active	Did the caterer pay the chef?
	Passive	Did the chef pay the caterer?
3	Active	Did the electrician injure the apprentice?
	Passive	Did the apprentice injure the electrician?
4	Active	Did the detective harass the witness?
	Passive	Did the witness harass the detective?
5	Active	Did the producer sue the artist?

	Passive	Did the artist sue the producer?
6	Active	Did the chef work in a café?
	Passive	Did the chef work in a café?
7	Active	Did the soldier abandon the general?
	Passive	Did the general abandon the soldier?
8	Active	Did the waitress injure the politician?
	Passive	Did the politician injure the waitress?
9	Active	Did the lawyer commit a felony?
	Passive	Did the judge commit a felony?
10	Active	Did the student mistreat the professor?
	Passive	Did the professor mistreat the student?
11	Active	Did the hotel designer like colours?
	Passive	Did the hotel designer like colours?
12	Active	Did the director reject the actress?
	Passive	Did the actress reject the director?
13	Active	Did the pilot stab the steward?
	Passive	Did the steward stab the pilot?
14	Active	Did the photographer mug the sculptor?
	Passive	Did the sculptor mug the photographer?
15	Active	Did the policeman attack the protester?
	Passive	Did the protester attack the policeman?
16	Active	Did the nuns cultivate tulips in the church garden?
	Passive	Did the nuns cultivate tulips in the church garden?
17	Active	Was the kitchen designed to be comfortable?
	Passive	Was the kitchen designed to be comfortable?
18	Active	Did the scientist rescue the reporter?
	Passive	Did the reporter rescue the scientist?
19	Active	Did the waitress work in a Chinese restaurant?
	Passive	Did the waitress work in a Chinese restaurant?
20	Active	Did the babysitter purchase products from the shop?
	Passive	Did the pharmacist hire the babysitter?
21	Active	Did the client sue the industrialist?

	Passive	Did the industrialist sue the client?
22	Active	Did the auto-electrician fire the mechanic?
	Passive	Did the mechanic fire the auto-electrician?
23	Active	Did the detective shock the suspect?
	Passive	Did the suspect shock the detective?
24	Active	Did the salesgirl work in a shopping mall?
	Passive	Did the salesgirl work in a shopping mall?
25	Active	Did the babysitter slap the child?
	Passive	Did the child slap the babysitter?
26	Active	Was the rebel hidden in a ruined shed?
	Passive	Was the soldier hidden in a ruined shed?
27	Active	Was the basement made of resistant materials?
	Passive	Was the basement made of resistant materials?
28	Active	Did the dancer approve the attitude of the choreographer?
	Passive	Did the choreographer approve the attitude of the dancer?
29	Active	Was the bartender known for his thick hair?
	Passive	Was the bartender known for his thick hair?
30	Active	Did the climber abandon the guide?
	Passive	Did the guide abandon the climber?

7.5 Experimental items in Exp. 2.

1	Passive	The guitarist was cherished by the attractive and talented singer for keeping the band focused through the whole tour.
	Active	The guitarist cherished the attractive and talented singer for keeping the band focused through the whole tour.
2	Passive	The caterer was admired by the creative and competent chef for being very professional at the French embassy's dinner.
	Active	The caterer admired the creative and competent chef for being very professional at the French embassy's dinner.
3	Passive	The apprentice was resented by the naïve and clumsy electrician for causing severe damage to the new electrical system.

	Active	The apprentice resented the naïve and clumsy electrician for causing severe damage to the new electrical system.
4	Passive	The detective was hated by the loud and hostile witness for having a violent outburst at the crime scene.
	Active	The detective hated the loud and hostile witness for having a violent outburst at the crime scene.
5	Passive	The producer was detested by the untalented and pretentious artist for showing a lack of commitment to the movie.
	Active	The producer detested the untalented and pretentious artist for showing a lack of commitment to the movie.
6	Passive	The chef was hated by the aggressive and insubordinate waiter for having an attitude of superiority in the kitchen.
	Active	The chef hated the aggressive and insubordinate waiter for having an attitude of superiority in the kitchen.
7	Passive	The spy was feared by the deceptive and cowardly soldier for behaving strangely during the most recent military engagement.
	Active	The spy feared the deceptive and cowardly soldier for behaving strangely during the most recent military engagement.
8	Passive	The politician was loved by the jealous and insecure waitress for buying a beautiful present for their recent anniversary.
	Active	The politician loved the jealous and insecure waitress for buying a beautiful present for their recent anniversary.
9	Passive	The lawyer was respected by the honest and wise judge for providing the poor suspect with excellent legal representation.
	Active	The lawyer respected the honest and wise judge for providing the poor suspect with excellent legal representation.
10	Passive	The student was cherished by the young and friendly professor for showing great enthusiasm during each of the lectures.
	Active	The student cherished the young and friendly professor for showing great enthusiasm during each of the lectures.
11	Passive	The broker was envied by the experienced and competent associate for proposing a brilliant idea to attract new customers.

	Active	The broker envied the experienced and competent associate for proposing a brilliant idea to attract new customers.
12	Passive	The director was adored by the charming and charismatic actress for receiving very good reviews of the first performance.
	Active	The director adored the charming and charismatic actress for receiving very good reviews of the first performance.
13	Passive	The pilot was resented by the bald and overweight steward for making a bad joke about the scared passenger.
	Active	The pilot resented the bald and overweight steward for making a bad joke about the scared passenger.
14	Passive	The sculptor was admired by the strange and temperamental photographer for achieving great success in the recent museum exhibition.
	Active	The sculptor admired the strange and temperamental photographer for achieving great success in the recent museum exhibition.
15	Passive	The protester was detested by the aggressive and stocky policeman for committing an extremely violent act during the demonstration.
	Active	The protester detested the aggressive and stocky policeman for committing an extremely violent act during the demonstration.
16	Passive	The nun was valued by the wise and respected teacher for working to improve the conditions of underprivileged children.
	Active	The nun valued the wise and respected teacher for working to improve the conditions of underprivileged children.
17	Passive	The nurse was valued by the handsome and distinguished doctor for working tirelessly during the eight hour surgical procedure.
	Active	The nurse valued the handsome and distinguished doctor for working tirelessly during the eight hour surgical procedure.
18	Passive	The scientist was despised by the brave and reckless reporter for demonstrating a lack of respect for animals' rights.
	Active	The scientist despised the brave and reckless reporter for demonstrating a lack of respect for animals' rights.

19	Passive	The waitress was loved by the fascinating and flirtatious model for saving the little cat trapped on the tree.
	Active	The waitress loved the fascinating and flirtatious model for saving the little cat trapped on the tree.
20	Passive	The babysitter was trusted by the mature and reliable pharmacist for having promptly intervened during the child's crying fit.
	Active	The babysitter trusted the mature and reliable pharmacist for having promptly intervened during the child's crying fit.
21	Passive	The industrialist was despised by the angry and resentful client for refusing sternly to settle the potentially damaging lawsuit.
	Active	The industrialist despised the angry and resentful client for refusing sternly to settle the potentially damaging lawsuit.
22	Passive	The mechanic was trusted by the kind and naïve customer for having helped during the fire in the neighbourhood.
	Active	The mechanic trusted the kind and naïve customer for having helped during the fire in the neighbourhood.
23	Passive	The detective was feared by the impulsive and panicky suspect for having behaved very frighteningly during the last interrogation.
	Active	The detective feared the impulsive and panicky suspect for having behaved very frighteningly during the last interrogation.
24	Passive	The salesgirl was pitied by the nervous and overworked businesswoman for reacting hysterically in the famous haute couture boutique.
	Active	The salesgirl pitied the nervous and overworked businesswoman for reacting hysterically in the famous haute couture boutique.
25	Passive	The child was adored by the young and skinny babysitter for being very creative and imaginative in their games.
	Active	The child adored the young and skinny babysitter for being very creative and imaginative in their games.
26	Passive	The soldier was loathed by the armed and dangerous rebel for imprisoning women and children in the recent raids.

	Active	The soldier loathed the armed and dangerous rebel for imprisoning women and children in the recent raids.
27	Passive	The student was respected by the aged and revered tutor for making interesting observations about the Ancient Ottoman Empire.
	Active	The student respected the aged and revered tutor for making interesting observations about the Ancient Ottoman Empire.
28	Passive	The dancer was loathed by the oppressive and abusive choreographer for performing very badly during the show on Friday.
	Active	The dancer loathed the oppressive and abusive choreographer for performing very badly during the show on Friday.
29	Passive	The investigator was pitied by the unpleasant and bald bartender for making very racist jokes at the local pub.
	Active	The investigator pitied the unpleasant and bald bartender for making very racist jokes at the local pub.
30	Passive	The guide was envied by the handsome and athletic climber for reaching the top of the mountain without resting.
	Active	The guide envied the handsome and athletic climber for reaching the top of the mountain without resting.

7.6 Comprehension questions in Exp. 2.

1	Active	Did the musician play the guitar?
	Passive	Did the musician play the guitar?
2	Active	Did the caterer admire the chef?
	Passive	Did the chef admire the caterer?
3	Active	Did the electrician resent the apprentice?
	Passive	Did the apprentice resent the electrician?
4	Active	Did the detective hate the witness?
	Passive	Did the witness hate the detective?
5	Active	Did the producer detest the artist?
	Passive	Did the artist detest the producer?
6	Active	Was the waiter humble and obedient?
	Passive	Was the waiter humble and obedient?

7	Active	Did the soldier fear the spy?
	Passive	Did the spy fear the soldier?
8	Active	Did the anniversary occur several months before?
	Passive	Did the anniversary occur several months before?
9	Active	Did the suspect receive legal representation?
	Passive	Did the suspect receive legal representation?
10	Active	Did the student cherish the professor?
	Passive	Did the professor cherish the student?
11	Active	Did the associate have a lot of experience?
	Passive	Did the associate have a lot of experience?
12	Active	Did the director adore the actress?
	Passive	Did the actress adore the director?
13	Active	Did the pilot resent the steward?
	Passive	Did the steward resent the pilot?
14	Active	Did the photographer admire the sculptor?
	Passive	Did the sculptor admire the photographer?
15	Active	Did the policeman detest the protester?
	Passive	Did the protester detest the policeman?
16	Active	Did the teacher value the nun?
	Passive	Did the nun value the teacher?
17	Active	Was the surgical procedure long?
	Passive	Was the surgical procedure long?
18	Active	Did the scientist despise the reporter?
	Passive	Did the reporter despise the scientist?
19	Active	Was the cat on a tree?
	Passive	Was the cat on a tree?
20	Active	Did the babysitter trust the pharmacist?
	Passive	Did the pharmacist trust the babysitter?
21	Active	Did the client despise the industrialist?
	Passive	Did the industrialist despise the client?
22	Active	Did the customer trust the mechanic?
	Passive	Did the customer trust the mechanic?

23	Active	Did the detective fear the suspect?
	Passive	Did the suspect fear the detective?
24	Active	Did the salesgirl work in a famous boutique?
	Passive	Did the salesgirl work in a famous boutique?
25	Active	Did the babysitter adore the child?
	Passive	Did the child adore the babysitter?
26	Active	Did the rebel imprison old men?
	Passive	Did the soldier imprison old men?
27	Active	Was the tutor young and unexperienced?
	Passive	Was the tutor young and unexperienced?
28	Active	Did the dancer appreciate the work of the choreographer?
	Passive	Did the choreographer appreciate the performance of the dancer?
29	Active	Was the bartender known for his thick hair?
	Passive	Was the bartender known for his thick hair?
30	Active	Did the climber envy the guide?
	Passive	Did the guide envy the climber?

7.7 Experimental items in Exp. 3, 4 & 5.

1	Eventive	Passive	The guitarist was rejected by the attractive and talented singer in the concert hall next to the Irish pub.
	Eventive	Active	The guitarist rejected the attractive and talented singer in the concert hall next to the Irish pub.
	Stative	Passive	The guitarist was cherished by the attractive and talented singer for keeping the band focused through the whole tour.
	Stative	Active	The guitarist cherished the attractive and talented singer for keeping the band focused through the whole tour.
2	Eventive	Passive	The caterer was paid by the creative and competent chef in the newly refurbished kitchen opposite the big cold-store.
	Eventive	Active	The caterer paid the creative and competent chef in the newly refurbished kitchen opposite the big cold-store.

	Stative	Passive	The caterer was admired by the creative and competent chef for being very professional at the French embassy's dinner.
	Stative	Active	The caterer admired the creative and competent chef for being very professional at the French embassy's dinner.
3	Eventive	Passive	The apprentice was injured by the naïve and clumsy electrician in the darkened room adjacent to the power line.
	Eventive	Active	The apprentice injured the naïve and clumsy electrician in the darkened room adjacent to the power line.
	Stative	Passive	The apprentice was resented by the naïve and clumsy electrician for causing severe damage to the new electrical system.
	Stative	Active	The apprentice resented the naïve and clumsy electrician for causing severe damage to the new electrical system.
4	Eventive	Passive	The detective was harassed by the loud and hostile witness in the front lobby next to the glass lift.
	Eventive	Active	The detective harassed the loud and hostile witness in the front lobby next to the glass lift.
	Stative	Passive	The detective was hated by the loud and hostile witness for having a violent outburst at the crime scene.
	Stative	Active	The detective hated the loud and hostile witness for having a violent outburst at the crime scene.
5	Eventive	Passive	The producer was sued by the untalented and pretentious artist in the legal offices next to the cinema studios.
	Eventive	Active	The producer sued the untalented and pretentious artist in the legal offices next to the cinema studios.
	Stative	Passive	The producer was detested by the untalented and pretentious artist for showing a lack of commitment to the movie.
	Stative	Active	The producer detested the untalented and pretentious artist for showing a lack of commitment to the movie.

6	Eventive	Passive	The chef was stabbed by the aggressive and insubordinate waiter in the restaurant kitchen next to the hotel foyer.
	Eventive	Active	The chef stabbed the aggressive and insubordinate waiter in the restaurant kitchen next to the hotel foyer.
	Stative	Passive	The chef was hated by the aggressive and insubordinate waiter for having an attitude of superiority in the kitchen.
	Stative	Active	The chef hated the aggressive and insubordinate waiter for having an attitude of superiority in the kitchen.
7	Eventive	Passive	The spy was abandoned by the deceptive and cowardly soldier in the vast forest adjacent to the military campsite.
	Eventive	Active	The spy abandoned the deceptive and cowardly soldier in the vast forest adjacent to the military campsite.
	Stative	Passive	The spy was feared by the deceptive and cowardly soldier for behaving strangely during the most recent military engagement.
	Stative	Active	The spy feared the deceptive and cowardly soldier for behaving strangely during the most recent military engagement.
8	Eventive	Passive	The politician was injured by the jealous and insecure waitress in the large bathroom next to the crowded press-room.
	Eventive	Active	The politician injured the jealous and insecure waitress in the large bathroom next to the crowded press-room.
	Stative	Passive	The politician was loved by the jealous and insecure waitress for buying a beautiful present for their recent anniversary.
	Stative	Active	The politician loved the jealous and insecure waitress for buying a beautiful present for their recent anniversary.
9	Eventive	Passive	The student was harassed by the young and inexperienced professor in the science laboratory next to the language department.

	Eventive	Active	The student harassed the young and inexperienced professor in the science laboratory next to the language department.
	Stative	Passive	The student was cherished by the young and inexperienced professor for showing great enthusiasm during each of the lectures.
	Stative	Active	The student cherished the young and inexperienced professor for showing great enthusiasm during each of the lectures.
10	Eventive	Passive	The director was rejected by the charming and charismatic actress in the café on the corner near the theatre.
	Eventive	Active	The director rejected the charming and charismatic actress in the café on the corner near the theatre.
	Stative	Passive	The director was adored by the charming and charismatic actress for receiving very good reviews of the first performance.
	Stative	Active	The director adored the charming and charismatic actress for receiving very good reviews of the first performance.
11	Eventive	Passive	The pilot was stabbed by the bald and overweight steward in the luxury airport lounge opposite the unisex toilets.
	Eventive	Active	The pilot stabbed the bald and overweight steward in the luxury airport lounge opposite the unisex toilets.
	Stative	Passive	The pilot was resented by the bald and overweight steward for making a bad joke about the scared passenger.
	Stative	Active	The pilot resented the bald and overweight steward for making a bad joke about the scared passenger.
12	Eventive	Passive	The sculptor was abandoned by the famous and ecletic photographer in the art gallery next to the book shop.
	Eventive	Active	The sculptor abandoned the famous and ecletic photographer in the art gallery next to the book shop.

	Stative	Passive	The sculptor was admired by the famous and ecletic photographer for achieving great success in the recent museum exhibition.
	Stative	Active	The sculptor admired the famous and ecletic photographer for achieving great success in the recent museum exhibition.
13	Eventive	Passive	The protester was attacked by the aggressive and stocky policeman in the crowded street near the old town centre.
	Eventive	Active	The protester attacked the aggressive and stocky policeman in the crowded street near the old town centre.
	Stative	Passive	The protester was detested by the aggressive and stocky policeman for committing an extremely violent act during the demonstration.
	Stative	Active	The protester detested the aggressive and stocky policeman for committing an extremely violent act during the demonstration.
14	Eventive	Passive	The nun was saved by the wise and respected teacher in the rose garden outside the old Catholic church.
	Eventive	Active	The nun saved the wise and respected teacher in the rose garden outside the old Catholic church.
	Stative	Passive	The nun was valued by the wise and respected teacher for working to improve the conditions of underprivileged children.
	Stative	Active	The nun valued the wise and respected teacher for working to improve the conditions of underprivileged children.
15	Eventive	Passive	The nurse was seduced by the handsome and distinguished doctor in the cosy staff kitchen on the seventh floor.
	Eventive	Active	The nurse seduced the handsome and distinguished doctor in the cosy staff kitchen on the seventh floor.

	Stative	Passive	The nurse was valued by the handsome and distinguished doctor for working tirelessly during the eight hour surgical procedure.
	Stative	Active	The nurse valued the handsome and distinguished doctor for working tirelessly during the eight hour surgical procedure.
16	Eventive	Passive	The scientist was rescued by the brave and reckless reporter in the chemical analysis laboratory near the great fire.
	Eventive	Active	The scientist rescued the brave and reckless reporter in the chemical analysis laboratory near the great fire.
	Stative	Passive	The scientist was despised by the brave and reckless reporter for demonstrating a lack of respect for animals' rights.
	Stative	Active	The scientist despised the brave and reckless reporter for demonstrating a lack of respect for animals' rights.
17	Eventive	Passive	The waitress was seduced by the fascinating and flirtatious model in the Chinese restaurant opposite the multi-brand fashion showroom.
	Eventive	Active	The waitress seduced the fascinating and flirtatious model in the Chinese restaurant opposite the multi-brand fashion showroom.
	Stative	Passive	The waitress was loved by the fascinating and flirtatious model for saving the little cat trapped on the tree.
	Stative	Active	The waitress loved the fascinating and flirtatious model for saving the little cat trapped on the tree.
18	Eventive	Passive	The babysitter was paid by the mature and reliable pharmacist in the office at the back of the shop.
	Eventive	Active	The babysitter paid the mature and reliable pharmacist in the office at the back of the shop.

	Stative	Passive	The babysitter was trusted by the mature and reliable pharmacist for having promptly intervened during the child's crying fit.
	Stative	Active	The babysitter trusted the mature and reliable pharmacist for having promptly intervened during the child's crying fit.
19	Eventive	Passive	The industrialist was sued by the angry and resentful client in the legal services office opposite the company building.
	Eventive	Active	The industrialist sued the angry and resentful client in the legal services office opposite the company building.
	Stative	Passive	The industrialist was despised by the angry and resentful client for refusing sternly to settle the potentially damaging lawsuit.
	Stative	Active	The industrialist despised the angry and resentful client for refusing sternly to settle the potentially damaging lawsuit.
20	Eventive	Passive	The mechanic was hired by the skilled and friendly auto-electrician in the spacious garage next to the congested highway.
	Eventive	Active	The mechanic hired the skilled and friendly auto-electrician in the spacious garage next to the congested highway.
	Stative	Passive	The mechanic was trusted by the skilled and friendly auto-electrician for having helped to complete repairs on the car.
	Stative	Active	The mechanic trusted the skilled and friendly auto-electrician for having helped to complete repairs on the car.
21	Eventive	Passive	The detective was pushed by the impulsive and panicky suspect in the dark interrogation room inside the police station.
	Eventive	Active	The detective pushed the impulsive and panicky suspect in the dark interrogation room inside the police station.

	Stative	Passive	The detective was feared by the impulsive and panicky suspect for having behaved very frighteningly during the last interrogation.
	Stative	Active	The detective feared the impulsive and panicky suspect for having behaved very frighteningly during the last interrogation.
22	Eventive	Passive	The salesgirl was pushed by the nervous and overworked businesswoman in the shopping mall next to the tube station.
	Eventive	Active	The salesgirl pushed the nervous and overworked businesswoman in the shopping mall next to the tube station.
	Stative	Passive	The salesgirl was pitied by the nervous and overworked businesswoman for reacting hysterically in the famous haute couture boutique.
	Stative	Active	The salesgirl pitied the nervous and overworked businesswoman for reacting hysterically in the famous haute couture boutique.
23	Eventive	Passive	The child was slapped by the young and skinny babysitter in the colourful nursery inside the big country-style house.
	Eventive	Active	The child slapped the young and skinny babysitter in the colourful nursery inside the big country-style house.
	Stative	Passive	The child was adored by the young and skinny babysitter for being very creative and imaginative in their games.
	Stative	Active	The child adored the young and skinny babysitter for being very creative and imaginative in their games.
24	Eventive	Passive	The soldier was attacked by the armed and dangerous rebel in the dark tunnel next to the combat zone.
	Eventive	Active	The soldier attacked the armed and dangerous rebel in the dark tunnel next to the combat zone.
	Stative	Passive	The soldier was loathed by the armed and dangerous rebel for imprisoning women and children in the recent raids.

	Stative	Active	The soldier loathed the armed and dangerous rebel for imprisoning women and children in the recent raids.
25	Eventive	Passive	The student was rescued by the aged and revered tutor near the collapsed basement inside the old humanities library.
	Eventive	Active	The student rescued the aged and revered tutor near the collapsed basement inside the old humanities library.
	Stative	Passive	The student was respected by the aged and revered tutor for making interesting observations about the Ancient Ottoman Empire.
	Stative	Active	The student respected the aged and revered tutor for making interesting observations about the Ancient Ottoman Empire.
26	Eventive	Passive	The dancer was slapped by the oppressive and abusive choreographer in the messy dressing room of the renowned theatre.
	Eventive	Active	The dancer slapped the oppressive and abusive choreographer in the messy dressing room of the renowned theatre.
	Stative	Passive	The dancer was loathed by the oppressive and abusive choreographer for performing very badly during the show on Friday.
	Stative	Active	The dancer loathed the oppressive and abusive choreographer for performing very badly during the show on Friday.
27	Eventive	Passive	The investigator was bribed by the unpleasant and bald bartender in the crowded pub next to the crime scene.
	Eventive	Active	The investigator bribed the unpleasant and bald bartender in the crowded pub next to the crime scene.
	Stative	Passive	The investigator was pitied by the unpleasant and bald bartender for making very racist jokes at the local pub.

	Stative	Active	The investigator pitied the unpleasant and bald bartender for making very racist jokes at the local pub.
28	Eventive	Passive	The guide was saved by the handsome and athletic climber on the perilous track on the rugged mountain top.
	Eventive	Active	The guide saved the handsome and athletic climber on the perilous track on the rugged mountain top.
	Stative	Passive	The guide was envied by the handsome and athletic climber for reaching the top of the mountain without resting.
	Stative	Active	The guide envied the handsome and athletic climber for reaching the top of the mountain without resting.

7.8 Comprehension questions in Exp. 3, 4, 6 and 8.

1	Eventive	Passive	Did the singer reject the guitarist?
	Eventive	Active	Did the guitarist reject the singer?
	Stative	Passive	Did the singer cherish the guitarist?
	Stative	Active	Did the guitarist cherish the singer?
2	Eventive	Passive	Did the chef pay the caterer?
	Eventive	Active	Did the caterer pay the chef?
	Stative	Passive	Did the chef admire the caterer?
	Stative	Active	Did the caterer admire the chef?
3	Eventive	Passive	Did the apprentice injure the electrician?
	Eventive	Active	Did the electrician injure the apprentice?
	Stative	Passive	Did the apprentice resent the electrician?
	Stative	Active	Did the electrician resent the apprentice?
4	Eventive	Passive	Did the witness harass the detective?
	Eventive	Active	Did the detective harass the witness?
	Stative	Passive	Did the witness hate the detective?
	Stative	Active	Did the detective hate the witness?
5	Eventive	Passive	Did the artist sue the producer?
	Eventive	Active	Did the producer sue the artist?
	Stative	Passive	Did the artist detest the producer?

	Stative	Active	Did the producer detest the artist?
6	Eventive	Passive	Did the waiter stab the chef?
	Eventive	Active	Did the chef stab the waiter?
	Stative	Passive	Did the waiter hate the chef?
	Stative	Active	Did the chef hate the waiter?
7	Eventive	Passive	Did the spy abandon the soldier?
	Eventive	Active	Did the soldier abandon the spy?
	Stative	Passive	Did the spy fear the soldier?
	Stative	Active	Did the soldier fear the spy?
8	Eventive	Passive	Did the politician injure the waitress?
	Eventive	Active	Did the waitress injure the politician?
	Stative	Passive	Did the politician love the waitress?
	Stative	Active	Did the waitress love the politician?
9	Eventive	Passive	Did the professor harass the student?
	Eventive	Active	Did the student harass the professor?
	Stative	Passive	Did the professor cherish the student?
	Stative	Active	Did the student cherish the professor?
10	Eventive	Passive	Did the actress reject the director?
	Eventive	Active	Did the director reject the actress?
	Stative	Passive	Did the actress adore the director?
	Stative	Active	Did the director adore the actress?
11	Eventive	Passive	Did the steward stab the pilot?
	Eventive	Active	Did the pilot stab the steward?
	Stative	Passive	Did the steward resent the pilot?
	Stative	Active	Did the pilot resent the steward?
12	Eventive	Passive	Did the sculptor abandon the photographer?
	Eventive	Active	Did the photographer abandon the sculptor?
	Stative	Passive	Did the sculptor admire the photographer?
	Stative	Active	Did the photographer admire the sculptor?
13	Eventive	Passive	Did the protester attack the policeman?
	Eventive	Active	Did the policeman attack the protester?
	Stative	Passive	Did the protester detest the policeman?

	Stative	Active	Did the policeman detest the protester?
14	Eventive	Passive	Did the nun save the teacher?
	Eventive	Active	Did the teacher save the nun?
	Stative	Passive	Did the nun value the teacher?
	Stative	Active	Did the teacher value the nun?
15	Eventive	Passive	Did the doctor seduce the nurse?
	Eventive	Active	Did the nurse seduce the doctor?
	Stative	Passive	Did the doctor value the nurse?
	Stative	Active	Did the nurse value the doctor?
16	Eventive	Passive	Did the reporter rescue the scientist?
	Eventive	Active	Did the scientist rescue the reporter?
	Stative	Passive	Did the reporter despise the scientist?
	Stative	Active	Did the scientist despise the reporter?
17	Eventive	Passive	Did the model seduce the waitress?
	Eventive	Active	Did the waitress seduce the model?
	Stative	Passive	Did the model love the waitress?
	Stative	Active	Did the waitress love the model?
18	Eventive	Passive	Did the pharmacist pay the babysitter?
	Eventive	Active	Did the babysitter pay the pharmacist?
	Stative	Passive	Did the pharmacist trust the babysitter?
	Stative	Active	Did the babysitter trust the pharmacist?
19	Eventive	Passive	Did the industrialist sue the client?
	Eventive	Active	Did the client sue the industrialist?
	Stative	Passive	Did the industrialist despise the client?
	Stative	Active	Did the client despise the industrialist?
20	Eventive	Passive	Did the mechanic hire the auto-electrician?
	Eventive	Active	Did the auto-electrician hire the mechanic?
	Stative	Passive	Did the mechanic trust the auto-electrician?
	Stative	Active	Did the auto-electrician trust the mechanic?
21	Eventive	Passive	Did the detective push the suspect?
	Eventive	Active	Did the suspect push the detective?
	Stative	Passive	Did the detective fear the suspect?

	Stative	Active	Did the suspect fear the detective?
22	Eventive	Passive	Did the businesswoman push the salesgirl?
	Eventive	Active	Did the salesgirl push the businesswoman?
	Stative	Passive	Did the businesswoman pity the salesgirl?
	Stative	Active	Did the salesgirl pity the businesswoman?
23	Eventive	Passive	Did the child slap the babysitter?
	Eventive	Active	Did the babysitter slap the child?
	Stative	Passive	Did the child adore the babysitter?
	Stative	Active	Did the babysitter adore the child?
24	Eventive	Passive	Did the soldier attack the rebel?
	Eventive	Active	Did the rebel attack the soldier?
	Stative	Passive	Did the soldier loath the rebel?
	Stative	Active	Did the rebel loath the soldier?
25	Eventive	Passive	Did the student rescue the tutor?
	Eventive	Active	Did the tutor rescue the student?
	Stative	Passive	Did the student respect the tutor?
	Stative	Active	Did the tutor respect the student?
26	Eventive	Passive	Did the dancer slap the choreographer?
	Eventive	Active	Did the choreographer slap the dancer?
	Stative	Passive	Did the dancer loath the choreographer?
	Stative	Active	Did the choreographer loath the dancer?
27	Eventive	Passive	Did the investigator bribe the bartender?
	Eventive	Active	Did the bartender bribe the investigator?
	Stative	Passive	Did the investigator pity the bartender?
	Stative	Active	Did the bartender pity the investigator?
28	Eventive	Passive	Did the climber save the guide?
	Eventive	Active	Did the guide save the climber?
	Stative	Passive	Did the climber envy the guide?
	Stative	Active	Did the guide envy the climber?

7.9 Fillers in Exp. 5.

Legend:

- Gramm= grammatical
- Ungramm= ungrammatical
- Agr= agreement
- Prep= preposition
- Compl= complementizer
- _2= 2 errors

	Condition	Error	Sentence
1	Gramm	-	The young couple who spent the weekend abroad were recently married after a long engagement.
2	Gramm	-	The researcher who presented his recent study at the prestigious conference had finished collecting his data only two days before.
3	Gramm	-	The captain who organised the retreat of his battalion had fought very bravely in the recent unapproved attacks.
4	Gramm	-	The ambassador who attended the party on Saturday is married to a young and bright journalist.
5	Gramm	-	The company executive who introduced the new safety rules was deeply concerned about the employees' working conditions.
6	Gramm	-	The opera singer who gave an extraordinary performance at La Scala had studied at Julliard in New York.
7	Gramm	-	The student who the professor greeted was the most brilliant of his year and president of the students' association.
8	Gramm	-	The boy who the girl hugged was the son of a very influential politician in Washington.

9	Gramm	-	The judge who the attorney corrupted was about to decide on a controversial case against a pharmaceutical company.
10	Gramm	-	The employee who the manager promoted was very happy and bought champagne for his colleagues.
11	Gramm	-	The activist who the moderator disqualified from the debate was well known by the media and the press.
12	Gramm	-	The rich lady who the criminal mugged outside the jewellery store had just bought an expensive necklace.
13	Gramm	-	The boy who the mob kidnapped outside a school was the son of an oil company manager.
14	Gramm	-	The young girl who the prince married was a bright alumna of Oxford University who graduated top of her class.
15	Ungramm	Agr	The journalist interviewed the celebrity and received very positive reviews from the readers of the magazine them was published in.
16	Ungramm	Agr_2	The writer spent two year in his chalet in a Alps and wrote his autobiography.
17	Ungramm	Agr	The therapist discovered the past traumas that the patient have endured during his childhood at the beginning of treatment in the clinic.
18	Ungramm	Agr_2	The gangster will gone to the striptease at the most famous night club in London with a enthusiastic friends and the boss.
19	Ungramm	Agr	The ghost haunted the villager who were currently living in the ghost's house on the top of the hill.
20	Ungramm	Agr_2	The player might scored twelve goal in the football match against the last team in the tournament.
21	Ungramm	Agr	The door-to-door salesperson married his hard working wife on a rainy days in Portofino during the spring.

22	Ungramm	Agr	The barber shaved the actor's hair in preparation for his new role in the latests action movie shot in Hollywood.
23	Ungramm	Agr	Santa Claus gave presents to all the child on the nice list and coal to those on the naughty list.
24	Ungramm	Agr	The circus performers entertained the audience while travelling the country with her new show in the circus tent.
25	Ungramm	Agr	The criminals was arrested by the law abiding policeman last Saturday on the town's busy high street.
26	Ungramm	Agr	The temptress were discovered seducing the friar by the priest in the old church with the cherry tree at its rear.
27	Ungramm	Agr	The sheep was herded by the dogs while the shepherd tried to move the flock away from the ferocious wolf.
28	Ungramm	Agr	The princes was eaten by the dragon trying to win the hand of the beautiful princess from the neighbouring kingdom.
29	Ungramm	Prep	The severely injured woman was saved by the paramedic after the fire for her house destroyed everything.
30	Ungramm	Prep_2	The irate passenger will removed on the train by the train conductor at the end of the work shift.
31	Ungramm	Prep	The solo for the symphony orchestra was played by a very talented cellist this year among the St. Pancras church.
32	Ungramm	Prep	The nobleman was amused by the comedian that performed many jokes and tricks that he had practiced before years.
33	Ungramm	Prep_2	The mice would experimented in by the scientists in the immensely high tech laboratory in the university.
34	Ungramm	Prep	The migrant was arrested by the police after she tried at enter the European country illegally.

35	Ungramm	Prep	The ducks who were fed bread by the children were swimming in a quiet lake on the north of England.
36	Ungramm	Prep	The burglar who had stolen many diamond necklaces from the jeweller went for prison several times over the years.
37	Ungramm	Prep_2	The philantropist who helped a refugees was happy for know that they were now citizens in a new country.
38	Ungramm	Prep	The tailor who hired the handy man at paint his house in the village liked very bright colours.
39	Ungramm	Prep	The slave who had escaped in his cruel master ran through the streets of ancient Rome towards the forum.
40	Ungramm	Prep	The exorcist who expelled the demon from the possessed girl at his silver cross knew the residents' many sordid secrets.
41	Ungramm	Prep	The clown who scared the little boy at his birthday party received a meagre tip with the parents.
42	Ungramm	Prep	The leader who appointed the advisor felt that the country was heading above a tough time in the years to come.
43	Ungramm	Compl	The zombie who bit the man was once a compassionate king how had ruled his lands fairly and kindly.
44	Ungramm	Compl_2	The lifeguard why rescued a children at the local beach had once spotted a dolphin in the water.
45	Ungramm	Compl	The plumber how spread butter on his toast with a knife had several jobs to complete that day.
46	Ungramm	Compl	The chef when served the lord his dinner only used very fresh ingredients for his recipes.
47	Ungramm	Compl_2	The customer where the blonde waitress served always left a generous tips on the table at the end of his lunch.
48	Ungramm	Compl	The little mouse how the tabby cat hunted hid in the house with a swing in its front garden.

49	Ungramm	Compl	The teacher where the parking attendant gave a ticket to for parking on double yellow lines disliked his job.
50	Ungramm	Compl	The model why the artist painted a picture of had won many prizes for his beauty.
51	Ungramm	Compl	The delinquent teenagers how the policeman instructed to vacate the area had drunk a lot of beer.
52	Ungramm	Compl	The sweet old housekeeper because the millionaire hired to clean and tidy had great references from her previous employers.
53	Ungramm	Compl	The man who the elephant trampled spent many months recovering in hospital because eventually returned to his family.
54	Ungramm	Compl_2	The cat that the dog will chased was a Siamese where had run away from its home.
55	Ungramm	Compl	The rabbit how the little girl fed grass lived in a large yellow hutch in the garden.
56	Ungramm	Compl	The terrorist since the army killed had carried out several attacks to innocent people over the years.

7.10 Comprehension questions in Exp. 5.

1	Eventive	Passive	Did the singer reject the guitarist?
	Eventive	Active	Did the guitarist reject the singer?
	Stative	Passive	Did the guitarist cherish the singer?
	Stative	Active	Did the singer cherish the guitarist?
2	Eventive	Passive	Did the chef pay the caterer?
	Eventive	Active	Did the caterer pay the chef?
	Stative	Passive	Did the chef admire the caterer?
	Stative	Active	Did the caterer admire the chef?
3	Eventive	Passive	Did the apprentice injure the electrician?
	Eventive	Active	Did the electrician injure the apprentice?
	Stative	Passive	Did the apprentice resent the electrician?
	Stative	Active	Did the electrician resent the apprentice?

4	Eventive	Passive	Did the witness harass the detective?
	Eventive	Active	Did the detective harass the witness?
	Stative	Passive	Did the witness hate the detective?
	Stative	Active	Did the detective hate the witness?
5	Eventive	Passive	Did the artist sue the producer?
	Eventive	Active	Did the producer sue the artist?
	Stative	Passive	Did the artist detest the producer?
	Stative	Active	Did the producer detest the artist?
6	Eventive	Passive	Did the waiter stab the chef?
	Eventive	Active	Did the chef stab the waiter?
	Stative	Passive	Did the waiter hate the chef?
	Stative	Active	Did the chef hate the waiter?
7	Eventive	Passive	Did the spy abandon the soldier?
	Eventive	Active	Did the soldier abandon the spy?
	Stative	Passive	Did the spy fear the soldier?
	Stative	Active	Did the soldier fear the spy?
8	Eventive	Passive	Did the politician injure the waitress?
	Eventive	Active	Did the waitress injure the politician?
	Stative	Passive	Did the politician love the waitress?
	Stative	Active	Did the waitress love the politician?
9	Eventive	Passive	Did the student harass the professor?
	Eventive	Active	Did the professor harass the student?
	Stative	Passive	Did the student cherish the professor?
	Stative	Active	Did the professor cherish the student?
10	Eventive	Passive	Did the actress reject the director?
	Eventive	Active	Did the director reject the actress?
	Stative	Passive	Did the actress adore the director?
	Stative	Active	Did the director adore the actress?
11	Eventive	Passive	Did the pilot stab the steward?
	Eventive	Active	Did the steward stab the pilot?
	Stative	Passive	Did the pilot resent the steward?
	Stative	Active	Did the steward resent the pilot?

12	Eventive	Passive	Was the art gallery next to a grocery shop?
	Eventive	Active	Was the art gallery next to a grocery shop?
	Stative	Passive	Did the sculptor present his work in a museum?
	Stative	Active	Did the photographer present his work in a museum?
13	Eventive	Passive	Did the attack took place near the city centre?
	Eventive	Active	Did the attack took place near the city centre?
	Stative	Passive	Did the protester commit a violent act during a meeting?
	Stative	Active	Did the policeman commit a violent act during a meeting?
14	Eventive	Passive	Was the garden outside an Anglican church?
	Eventive	Active	Was the garden outside an Anglican church?
	Stative	Passive	Did the nun work in a school for underprivileged children?
	Stative	Active	Did the teacher work in a school for underprivileged children?
15	Eventive	Passive	Was the kitchen on the seventh floor?
	Eventive	Active	Was the kitchen on the seventh floor?
	Stative	Passive	Were the nurse and the doctor in a waiting room?
	Stative	Active	Were the nurse and the doctor in a waiting room?
16	Eventive	Passive	Was the scientist in the library?
	Eventive	Active	Was the scientist in the library?
	Stative	Passive	Was the reporter daring?
	Stative	Active	Was the reporter daring?
17	Eventive	Passive	Did the waitress work in a Chinese restaurant?
	Eventive	Active	Did the waitress work in a Chinese restaurant?
	Stative	Passive	Was the cat trapped on a tree?
	Stative	Active	Was the cat trapped on a tree?
18	Eventive	Passive	Was the office at the front of the shop?
	Eventive	Active	Was the office at the front of the shop?
	Stative	Passive	Was the pharmacist trustworthy?
	Stative	Active	Was the pharmacist trustworthy?
19	Eventive	Passive	Was the client irate and spiteful?
	Eventive	Active	Was the client irate and spiteful?
	Stative	Passive	Was the client happy and satisfied?

	Stative	Active	Was the client happy and satisfied?
20	Eventive	Passive	Was the garage next to a dead-end street?
	Eventive	Active	Was the garage next to a dead-end street?
	Stative	Passive	Did the mechanic work on a motorcycle?
	Stative	Active	Did the electrician work on a motorcycle?
21	Eventive	Passive	Was the suspect calm and relaxed?
	Eventive	Active	Was the suspect calm and relaxed?
	Stative	Passive	Was the suspect calm and relaxed?
	Stative	Active	Was the suspect calm and relaxed?
22	Eventive	Passive	Was the businesswoman nervous?
	Eventive	Active	Was the businesswoman nervous?
	Stative	Passive	Did the salesgirl work at H&M?
	Stative	Active	Did the salesgirl work at H&M?
23	Eventive	Passive	Was the babysitter slim?
	Eventive	Active	Was the babysitter slim?
	Stative	Passive	Was the child unimaginative?
	Stative	Active	Was the babysitter unimaginative?
24	Eventive	Passive	Was the rebel unarmed and harmless?
	Eventive	Active	Was the rebel unarmed and harmless?
	Stative	Passive	Had the raids been carried out recently?
	Stative	Active	Had the raids been carried out recently?
25	Eventive	Passive	Was the tutor respected?
	Eventive	Active	Was the tutor respected?
	Stative	Passive	Was the student interested in the Ancient Ottoman Empire?
	Stative	Active	Was the tutor interested in the Ancient Ottoman Empire?
26	Eventive	Passive	Was the theatre unknown?
	Eventive	Active	Was the theatre unknown?
	Stative	Passive	Was the choreographer demanding and scathing?
	Stative	Active	Was the choreographer demanding and scathing?
27	Eventive	Passive	Was the pub crowded?
	Eventive	Active	Was the pub crowded?

	Stative	Passive	Was the investigator in a Chinese restaurant?
	Stative	Active	Was the investigator in a Chinese restaurant?
28	Eventive	Passive	Was the track safe and secure?
	Eventive	Active	Was the track safe and secure?
	Stative	Passive	Was the climber fit?
	Stative	Active	Was the climber fit?

7.11 Experimental items in Exp. 6 and 7.

1	Eventive	Passive	The guitarist was rejected by the attractive and talented singer.
	Eventive	Active	The guitarist rejected the attractive and talented singer.
	Stative	Passive	The guitarist was cherished by the attractive and talented singer.
	Stative	Active	The guitarist cherished the attractive and talented singer.
2	Eventive	Passive	The caterer was paid by the creative and competent chef.
	Eventive	Active	The caterer paid the creative and competent chef.
	Stative	Passive	The caterer was admired by the creative and competent chef.
	Stative	Active	The caterer admired the creative and competent chef.
3	Eventive	Passive	The apprentice was injured by the naïve and clumsy electrician.
	Eventive	Active	The apprentice injured the naïve and clumsy electrician.
	Stative	Passive	The apprentice was resented by the naïve and clumsy electrician.
	Stative	Active	The apprentice resented the naïve and clumsy electrician.
4	Eventive	Passive	The detective was harassed by the loud and hostile witness.
	Eventive	Active	The detective harassed the loud and hostile witness.
	Stative	Passive	The detective was hated by the loud and hostile witness.
	Stative	Active	The detective hated the loud and hostile witness.
5	Eventive	Passive	The producer was sued by the untalented and pretentious artist.
	Eventive	Active	The producer sued the untalented and pretentious artist.

	Stative	Passive	The producer was detested by the untalented and pretentious artist.
	Stative	Active	The producer detested the untalented and pretentious artist.
6	Eventive	Passive	The chef was stabbed by the aggressive and insubordinate waiter.
	Eventive	Active	The chef stabbed the aggressive and insubordinate waiter.
	Stative	Passive	The chef was hated by the aggressive and insubordinate waiter.
	Stative	Active	The chef hated the aggressive and insubordinate waiter.
7	Eventive	Passive	The spy was abandoned by the deceptive and cowardly soldier.
	Eventive	Active	The spy abandoned the deceptive and cowardly soldier.
	Stative	Passive	The spy was feared by the deceptive and cowardly soldier.
	Stative	Active	The spy feared the deceptive and cowardly soldier.
8	Eventive	Passive	The politician was injured by the jealous and insecure waitress.
	Eventive	Active	The politician injured the jealous and insecure waitress.
	Stative	Passive	The politician was loved by the jealous and insecure waitress.
	Stative	Active	The politician loved the jealous and insecure waitress.
9	Eventive	Passive	The student was harassed by the young and inexperienced professor.
	Eventive	Active	The student harassed the young and inexperienced professor.
	Stative	Passive	The student was cherished by the young and inexperienced professor.
	Stative	Active	The student cherished the young and inexperienced professor.
10	Eventive	Passive	The director was rejected by the charming and charismatic actress.
	Eventive	Active	The director rejected the charming and charismatic actress.

	Stative	Passive	The director was adored by the charming and charismatic actress.
	Stative	Active	The director adored the charming and charismatic actress.
11	Eventive	Passive	The pilot was stabbed by the bald and overweight steward.
	Eventive	Active	The pilot stabbed the bald and overweight steward.
	Stative	Passive	The pilot was resented by the bald and overweight steward.
	Stative	Active	The pilot resented the bald and overweight steward.
12	Eventive	Passive	The sculptor was abandoned by the famous and eclectic photographer.
	Eventive	Active	The sculptor abandoned the famous and eclectic photographer.
	Stative	Passive	The sculptor was admired by the famous and eclectic photographer.
	Stative	Active	The sculptor admired the famous and eclectic photographer.
13	Eventive	Passive	The protester was attacked by the aggressive and stocky policeman.
	Eventive	Active	The protester attacked the aggressive and stocky policeman.
	Stative	Passive	The protester was detested by the aggressive and stocky policeman.
	Stative	Active	The protester detested the aggressive and stocky policeman.
14	Eventive	Passive	The nun was saved by the wise and respected teacher.
	Eventive	Active	The nun saved the wise and respected teacher.
	Stative	Passive	The nun was valued by the wise and respected teacher.
	Stative	Active	The nun valued the wise and respected teacher.
15	Eventive	Passive	The nurse was seduced by the handsome and distinguished doctor.
	Eventive	Active	The nurse seduced the handsome and distinguished doctor.
	Stative	Passive	The nurse was valued by the handsome and distinguished doctor.

	Stative	Active	The nurse valued the handsome and distinguished doctor.
16	Eventive	Passive	The scientist was rescued by the brave and reckless reporter.
	Eventive	Active	The scientist rescued the brave and reckless reporter.
	Stative	Passive	The scientist was despised by the brave and reckless reporter.
	Stative	Active	The scientist despised the brave and reckless reporter.
17	Eventive	Passive	The waitress was seduced by the fascinating and flirtatious model.
	Eventive	Active	The waitress seduced the fascinating and flirtatious model.
	Stative	Passive	The waitress was loved by the fascinating and flirtatious model.
	Stative	Active	The waitress loved the fascinating and flirtatious model.
18	Eventive	Passive	The babysitter was paid by the mature and reliable pharmacist.
	Eventive	Active	The babysitter paid the mature and reliable pharmacist.
	Stative	Passive	The babysitter was trusted by the mature and reliable pharmacist.
	Stative	Active	The babysitter trusted the mature and reliable pharmacist.
19	Eventive	Passive	The industrialist was sued by the angry and resentful client.
	Eventive	Active	The industrialist sued the angry and resentful client.
	Stative	Passive	The industrialist was despised by the angry and resentful client.
	Stative	Active	The industrialist despised the angry and resentful client.
20	Eventive	Passive	The mechanic was hired by the skilled and friendly auto-electrician.
	Eventive	Active	The mechanic hired the skilled and friendly auto-electrician.
	Stative	Passive	The mechanic was trusted by the skilled and friendly auto-electrician.
	Stative	Active	The mechanic trusted the skilled and friendly auto-electrician.

21	Eventive	Passive	The detective was pushed by the impulsive and panicky suspect.
	Eventive	Active	The detective pushed the impulsive and panicky suspect.
	Stative	Passive	The detective was feared by the impulsive and panicky suspect.
	Stative	Active	The detective feared the impulsive and panicky suspect.
22	Eventive	Passive	The salesgirl was pushed by the nervous and overworked businesswoman.
	Eventive	Active	The salesgirl pushed the nervous and overworked businesswoman.
	Stative	Passive	The salesgirl was pitied by the nervous and overworked businesswoman.
	Stative	Active	The salesgirl pitied the nervous and overworked businesswoman.
23	Eventive	Passive	The child was slapped by the young and skinny babysitter.
	Eventive	Active	The child slapped the young and skinny babysitter.
	Stative	Passive	The child was adored by the young and skinny babysitter.
	Stative	Active	The child adored the young and skinny babysitter.
24	Eventive	Passive	The soldier was attacked by the armed and dangerous rebel.
	Eventive	Active	The soldier attacked the armed and dangerous rebel.
	Stative	Passive	The soldier was loathed by the armed and dangerous rebel.
	Stative	Active	The soldier loathed the armed and dangerous rebel.
25	Eventive	Passive	The student was rescued by the aged and revered tutor.
	Eventive	Active	The student rescued the aged and revered tutor.
	Stative	Passive	The student was respected by the aged and revered tutor.
	Stative	Active	The student respected the aged and revered tutor.
26	Eventive	Passive	The dancer was slapped by the oppressive and abusive choreographer.
	Eventive	Active	The dancer slapped the oppressive and abusive choreographer.

	Stative	Passive	The dancer was loathed by the oppressive and abusive choreographer.
	Stative	Active	The dancer loathed the oppressive and abusive choreographer.
27	Eventive	Passive	The investigator was bribed by the unpleasant and bald bartender.
	Eventive	Active	The investigator bribed the unpleasant and bald bartender.
	Stative	Passive	The investigator was pitied by the unpleasant and bald bartender.
	Stative	Active	The investigator pitied the unpleasant and bald bartender.
28	Eventive	Passive	The guide was saved by the handsome and athletic climber.
	Eventive	Active	The guide saved the handsome and athletic climber.
	Stative	Passive	The guide was envied by the handsome and athletic climber.
	Stative	Active	The guide envied the handsome and athletic climber.

7.12 Filler items in Exp. 6 and 7.

Legend:

- Gramm= grammatical
- Ungramm= ungrammatical
- Agr= agreement
- Prep= preposition
- Compl= complementizer
- _2= 2 errors

	Condition	Error	Sentence
1	Gramm	-	The young couple were recently married after a long engagement.
2	Gramm	-	The researcher had finished collecting his data two days ago.

3	Gramm	-	The captain had fought very bravely in the recent unapproved attacks.
4	Gramm	-	The ambassador is married to a young and bright journalist.
5	Gramm	-	The company executive was deeply concerned about the employees' working conditions.
6	Gramm	-	The opera singer had studied at Julliard in New York.
7	Gramm	-	The brilliant student was the head of the Bridge association.
8	Gramm	-	The boy was the son of an influential politician in Washington.
9	Gramm	-	The judge was about to decide on a controversial case.
10	Gramm	-	After the employee was promoted he bought champagne for his colleagues.
11	Gramm	-	The activist left the debate after a heated discussion.
12	Gramm	-	The rich lady had just bought an expensive necklace.
13	Gramm	-	The boy was kidnapped outside of the prestigious school.
14	Gramm	-	The young girl graduated at the top of her class at Oxford University.
15	Ungramm	Agr	The journalist interviewed the celebrity about its recent movie.
16	Ungramm	Agr_2	The writer spent two year in his chalet in a Alps.
17	Ungramm	Agr	The therapist discovered the past traumas that the patient have endured.
18	Ungramm	Agr_2	The gangster will gone to the striptease at a very famous night clubs.
19	Ungramm	Agr	The ghost haunted the woman who were living in the house.
20	Ungramm	Agr_2	The player might scored twelve goal in the football match.

21	Ungramm	Agr	The door-to-door salesperson married his hard working wife on a rainy days.
22	Ungramm	Agr	The barber would shaved the actor's hair in preparation for his new role.
23	Ungramm	Agr	Santa Claus gave presents to all the child on the nice list.
24	Ungramm	Agr	The circus performers entertains the audience while travelling the country.
25	Ungramm	Agr	The criminals was fleeing from the police on a stolen car.
26	Ungramm	Agr	The temptress were discovered seducing the friar in the old church.
27	Ungramm	Agr	The goats that was grazing in the fields belonged to an unfriendly farmer.
28	Ungramm	Agr	The princesses who lived in an enchanted castle was good at cooking.
29	Ungramm	Prep	The severely injured woman was saved at the paramedic.
30	Ungramm	Prep_2	The irate passenger will removed until the train by the conductor.
31	Ungramm	Prep	The solo for the symphony orchestra was played across a cellist.
32	Ungramm	Prep	The nobleman was amused upon the comedian that performed many jokes.
33	Ungramm	Prep_2	The scientist who experimented until mice have written a famous article.
34	Ungramm	Prep	The migrant who tried at enter the country was fleeing from war.
35	Ungramm	Prep	The ducks were fed bread through the children on the lake's shore.
36	Ungramm	Prep	The burglar went to prison several times for the years.
37	Ungramm	Prep_2	The doctor who helped to a refugees was very generous.
38	Ungramm	Prep	The tailor hired the handy man at paint his house in the village.

39	Ungramm	Prep	The slave who had escaped in his cruel master ran through the streets.
40	Ungramm	Prep	The exorcist expelled the demon from the possessed girl until his silver cross.
41	Ungramm	Prep	The little boy was scared by the clown upon the party.
42	Ungramm	Prep	The leader felt that the country was heading between a tough time.
43	Ungramm	Compl	The zombie where bit the man was once a compassionate king.
44	Ungramm	Compl_2	The lifeguard why rescued a children had once spotted a dolphin.
45	Ungramm	Compl	The plumber how spread butter on his toast had several jobs.
46	Ungramm	Compl	The chef since served the lord only used fresh ingredients.
47	Ungramm	Compl_2	The customer where the blonde waitress served left a generous tips.
48	Ungramm	Compl	The little mouse how the tabby cat hunted hid in the house.
49	Ungramm	Compl	The actress where the teacher admired liked Shakespeare.
50	Ungramm	Compl	The model why the artist painted a picture of was tall.
51	Ungramm	Compl	The teenager how the policeman ordered to leave had freckles.
52	Ungramm	Compl	The housekeeper because the millionaire hired to clean had great references.
53	Ungramm	Compl	The man when the elephant trampled spent many months in hospital.
54	Ungramm	Compl_2	The cat where the dog will chased was a Siamese.
55	Ungramm	Compl	The rabbit how the little girl fed grass lived in a hutch.
56	Ungramm	Compl	The terrorist since the army killed had carried out several attacks.

7.13 Comprehension questions in Exp. 7.

1	Eventive	Passive	Did the singer reject the guitarist?
	Eventive	Active	Was the singer rejected by the guitarist?
	Stative	Passive	Did the singer cherish the guitarist?
	Stative	Active	Was the singer cherished by the guitarist?
2	Eventive	Passive	Was the chef paid by the caterer?
	Eventive	Active	Did the chef pay the caterer?
	Stative	Passive	Was the chef admired by the caterer?
	Stative	Active	Did the chef admire the caterer?
3	Eventive	Passive	Did the apprentice injure the electrician?
	Eventive	Active	Was the apprentice injured by the electrician?
	Stative	Passive	Did the apprentice resent the electrician?
	Stative	Active	Was the apprentice resented by the electrician?
4	Eventive	Passive	Did the witness harass the detective?
	Eventive	Active	Was the witness harassed by the detective?
	Stative	Passive	Did the witness hate the detective?
	Stative	Active	Was the witness hated by the detective?
5	Eventive	Passive	Was the producer sued by the artist?
	Eventive	Active	Did the producer sue the artist?
	Stative	Passive	Was the producer detested by the artist?
	Stative	Active	Did the producer detest the artist?
6	Eventive	Passive	Did the waiter stab the chef?
	Eventive	Active	Was the waiter stabbed by the chef?
	Stative	Passive	Did the waiter hate the chef?
	Stative	Active	Was the waiter hated by the chef?
7	Eventive	Passive	Did the spy abandon the soldier?
	Eventive	Active	Was the spy abandoned by the soldier?
	Stative	Passive	Did the spy fear the soldier?
	Stative	Active	Was the spy feared by the soldier?
8	Eventive	Passive	Was the waitress injured by the politician?
	Eventive	Active	Did the waitress injure the politician?

	Stative	Passive	Was the waitress loved by the politician?
	Stative	Active	Did the waitress love the politician?
9	Eventive	Passive	Did the student harass the professor?
	Eventive	Active	Was the student harassed by the professor?
	Stative	Passive	Did the student cherish the professor?
	Stative	Active	Was the student cherished by the professor?
10	Eventive	Passive	Was the director rejected by the actress?
	Eventive	Active	Did the director reject the actress?
	Stative	Passive	Was the director adored by the actress?
	Stative	Active	Did the director adore the actress?
11	Eventive	Passive	Was the steward stabbed by the pilot?
	Eventive	Active	Did the steward stab the pilot?
	Stative	Passive	Was the steward resented by the pilot?
	Stative	Active	Did the steward resent the pilot?
12	Eventive	Passive	Did the photographer abandon the sculptor?
	Eventive	Active	Was the photographer abandoned by the sculptor?
	Stative	Passive	Did the photographer admire the sculptor?
	Stative	Active	Was the photographer admired by the sculptor?
13	Eventive	Passive	Was the policeman attacked by the protester?
	Eventive	Active	Did the policeman attack the protester?
	Stative	Passive	Was the policeman detested by the protester?
	Stative	Active	Did the policeman detest the protester?
14	Eventive	Passive	Was the nun saved by the teacher?
	Eventive	Active	Did the nun save the teacher?
	Stative	Passive	Was the nun valued by the teacher?
	Stative	Active	Did the nun value the teacher?
15	Eventive	Passive	Was the doctor seduced by the nurse?
	Eventive	Active	Did the doctor seduce the nurse?
	Stative	Passive	Was the doctor valued by the nurse?
	Stative	Active	Did the doctor value the nurse?
16	Eventive	Passive	Did the reporter rescue the scientist?
	Eventive	Active	Was the reporter rescued by the scientist?

	Stative	Passive	Did the reporter despise the scientist?
	Stative	Active	Was the reporter despised by the scientist?
17	Eventive	Passive	Was the waitress seduced by the model?
	Eventive	Active	Did the waitress seduce the model?
	Stative	Passive	Was the waitress loved by the model?
	Stative	Active	Did the waitress love the model?
18	Eventive	Passive	Was the babysitter paid by the pharmacist?
	Eventive	Active	Did the babysitter pay the pharmacist?
	Stative	Passive	Was the babysitter trusted by the pharmacist?
	Stative	Active	Did the babysitter trust the pharmacist?
19	Eventive	Passive	Did the industrialist sue the client?
	Eventive	Active	Was the industrialist sue by the client?
	Stative	Passive	Did the industrialist despise the client?
	Stative	Active	Was the industrialist despised by the client?
20	Eventive	Passive	Did the mechanic hire the auto-electrician?
	Eventive	Active	Was the mechanic hired by the auto-electrician?
	Stative	Passive	Did the mechanic trust the auto-electrician?
	Stative	Active	Was the mechanic trusted by the auto-electrician?
21	Eventive	Passive	Did the detective push the suspect?
	Eventive	Active	Was the detective pushed by the suspect?
	Stative	Passive	Did the detective fear the suspect?
	Stative	Active	Was the detective feared by the suspect?
22	Eventive	Passive	Was the businesswoman pushed by the salesgirl?
	Eventive	Active	Did the businesswoman push the salesgirl?
	Stative	Passive	Was the businesswoman pitied by the salesgirl?
	Stative	Active	Did the businesswoman pity the salesgirl?
23	Eventive	Passive	Did the child slap the babysitter?
	Eventive	Active	Was the child slapped by the babysitter?
	Stative	Passive	Did the child adore the babysitter?
	Stative	Active	Was the child adored by the babysitter?
24	Eventive	Passive	Was the soldier attacked by the rebel?
	Eventive	Active	Did the soldier attack the rebel?

	Stative	Passive	Was the soldier loathed by the rebel?
	Stative	Active	Did the soldier loath the rebel?
25	Eventive	Passive	Was the student rescued by the tutor?
	Eventive	Active	Did the student rescue the tutor?
	Stative	Passive	Was the student respected by the tutor?
	Stative	Active	Did the student respect the tutor?
26	Eventive	Passive	Did the choreographer slap the dancer?
	Eventive	Active	Was the choreographer slapped by the dancer?
	Stative	Passive	Did the choreographer loath the dancer?
	Stative	Active	Was the choreographer loathed by the dancer?
27	Eventive	Passive	Was the bartender bribed by the investigator?
	Eventive	Active	Did the bartender bribe the investigator?
	Stative	Passive	Was the bartender pitied by the investigator?
	Stative	Active	Did the bartender pity the investigator?
28	Eventive	Passive	Did the climber save the guide?
	Eventive	Active	Was the climber saved by the guide?
	Stative	Passive	Did the climber envy the guide?
	Stative	Active	Was the climber envied by the guide?

7.14 Experimental items in Exp. 8.

1	Eventive	Passive	The guitarist was rejected by the attractive and talented singer on Tuesday morning.
	Eventive	Active	The guitarist rejected the attractive and talented singer on Tuesday morning.
	Stative	Passive	The guitarist was cherished by the attractive and talented singer throughout the tour.
	Stative	Active	The guitarist cherished the attractive and talented singer throughout the tour.
2	Eventive	Passive	The caterer was paid by the creative and competent chef in the afternoon.
	Eventive	Active	The caterer paid the creative and competent chef in the afternoon.

	Stative	Passive	The caterer was admired by the creative and competent chef since the banquet.
	Stative	Active	The caterer admired the creative and competent chef since the banquet.
3	Eventive	Passive	The apprentice was injured by the naïve and clumsy electrician during the repair.
	Eventive	Active	The apprentice injured the naïve and clumsy electrician during the repair.
	Stative	Passive	The apprentice was resented by the naïve and clumsy electrician after the accident.
	Stative	Active	The apprentice resented the naïve and clumsy electrician after the accident.
4	Eventive	Passive	The detective was harassed by the loud and hostile witness on Thursday evening.
	Eventive	Active	The detective harassed the loud and hostile witness on Thursday evening.
	Stative	Passive	The detective was hated by the loud and hostile witness after the investigation.
	Stative	Active	The detective hated the loud and hostile witness after the investigation.
5	Eventive	Passive	The producer was sued by the untalented and pretentious artist in early September.
	Eventive	Active	The producer sued the untalented and pretentious artist in early September.
	Stative	Passive	The producer was detested by the untalented and pretentious artist throughout the filming.
	Stative	Active	The producer detested the untalented and pretentious artist throughout the filming.
6	Eventive	Passive	The chef was stabbed by the aggressive and insubordinate waiter on Friday evening.
	Eventive	Active	The chef stabbed the aggressive and insubordinate waiter on Friday evening.

	Stative	Passive	The chef was hated by the aggressive and insubordinate waiter after the fire.
	Stative	Active	The chef hated the aggressive and insubordinate waiter after the fire.
7	Eventive	Passive	The spy was abandoned by the deceptive and cowardly soldier during the night.
	Eventive	Active	The spy abandoned the deceptive and cowardly soldier during the night.
	Stative	Passive	The spy was feared by the deceptive and cowardly soldier since the attack.
	Stative	Active	The spy feared the deceptive and cowardly soldier since the attack.
8	Eventive	Passive	The politician was injured by the jealous and insecure waitress at 6 o'clock.
	Eventive	Active	The politician injured the jealous and insecure waitress at 6 o'clock.
	Stative	Passive	The politician was loved by the jealous and insecure waitress throughout the election.
	Stative	Active	The politician loved the jealous and insecure waitress throughout the election.
9	Eventive	Passive	The student was harassed by the young and inexperienced professor at the party.
	Eventive	Active	The student harassed the young and inexperienced professor at the party.
	Stative	Passive	The student was cherished by the young and inexperienced professor before the argument.
	Stative	Active	The student cherished the young and inexperienced professor before the argument.
10	Eventive	Passive	The director was rejected by the charming and charismatic actress at the premiere.
	Eventive	Active	The director rejected the charming and charismatic actress at the premiere.

	Stative	Passive	The director was adored by the charming and charismatic actress since the debut.
	Stative	Active	The director adored the charming and charismatic actress since the debut.
11	Eventive	Passive	The pilot was stabbed by the bald and overweight steward at 2 o'clock.
	Eventive	Active	The pilot stabbed the bald and overweight steward at 2 o'clock.
	Stative	Passive	The pilot was resented by the bald and overweight steward after the flight.
	Stative	Active	The pilot resented the bald and overweight steward after the flight.
12	Eventive	Passive	The sculptor was abandoned by the famous and eclectic photographer in early June.
	Eventive	Active	The sculptor abandoned the famous and eclectic photographer in early June.
	Stative	Passive	The sculptor was admired by the famous and eclectic photographer since the exhibition.
	Stative	Active	The sculptor admired the famous and eclectic photographer since the exhibition.
13	Eventive	Passive	The protester was attacked by the aggressive and stocky policeman on Wednesday afternoon.
	Eventive	Active	The protester attacked the aggressive and stocky policeman on Wednesday afternoon.
	Stative	Passive	The protester was detested by the aggressive and stocky policeman after the rally.
	Stative	Active	The protester detested the aggressive and stocky policeman after the rally.
14	Eventive	Passive	The nun was saved by the wise and respected teacher during the fire.
	Eventive	Active	The nun saved the wise and respected teacher during the fire.

	Stative	Passive	The nun was valued by the wise and respected teacher before the confrontation.
	Stative	Active	The nun valued the wise and respected teacher before the confrontation.
15	Eventive	Passive	The nurse was seduced by the handsome and distinguished doctor in early May.
	Eventive	Active	The nurse seduced the handsome and distinguished doctor in early May.
	Stative	Passive	The nurse was valued by the handsome and distinguished doctor since the surgery.
	Stative	Active	The nurse valued the handsome and distinguished doctor since the surgery.
16	Eventive	Passive	The scientist was rescued by the brave and reckless reporter at 3 o'clock.
	Eventive	Active	The scientist rescued the brave and reckless reporter at 3 o'clock.
	Stative	Passive	The scientist was despised by the brave and reckless reporter throughout the interview.
	Stative	Active	The scientist despised the brave and reckless reporter throughout the interview.
17	Eventive	Passive	The waitress was seduced by the fascinating and flirtatious model during the night.
	Eventive	Active	The waitress seduced the fascinating and flirtatious model during the night.
	Stative	Passive	The waitress was loved by the fascinating and flirtatious model before the affair.
	Stative	Active	The waitress loved the fascinating and flirtatious model before the affair.
18	Eventive	Passive	The babysitter was paid by the mature and reliable pharmacist on Saturday morning.
	Eventive	Active	The babysitter paid the mature and reliable pharmacist on Saturday morning.

	Stative	Passive	The babysitter was trusted by the mature and reliable pharmacist before the incident.
	Stative	Active	The babysitter trusted the mature and reliable pharmacist before the incident.
19	Eventive	Passive	The industrialist was sued by the angry and resentful client in early January.
	Eventive	Active	The industrialist sued the angry and resentful client in early January.
	Stative	Passive	The industrialist was despised by the angry and resentful client after the lawsuit.
	Stative	Active	The industrialist despised the angry and resentful client after the lawsuit.
20	Eventive	Passive	The mechanic was hired by the skilled and friendly auto-electrician in the morning.
	Eventive	Active	The mechanic hired the skilled and friendly auto-electrician in the morning.
	Stative	Passive	The mechanic was trusted by the skilled and friendly auto-electrician throughout the contract.
	Stative	Active	The mechanic trusted the skilled and friendly auto-electrician throughout the contract.
21	Eventive	Passive	The detective was pushed by the impulsive and panicky suspect during the arrest.
	Eventive	Active	The detective pushed the impulsive and panicky suspect during the arrest.
	Stative	Passive	The detective was feared by the impulsive and panicky suspect throughout the interrogation.
	Stative	Active	The detective feared the impulsive and panicky suspect througout the interroogation.
22	Eventive	Passive	The salesgirl was pushed by the nervous and overworked businesswoman during the fight.
	Eventive	Active	The salesgirl pushed the nervous and overworked businesswoman during the fight.

	Stative	Passive	The salesgirl was pitied by the nervous and overworked businesswoman since the breakdown.
	Stative	Active	The salesgirl pitied the nervous and overworked businesswoman since the breakdown.
23	Eventive	Passive	The child was slapped by the young and skinny babysitter in the afternoon.
	Eventive	Active	The child slapped the young and skinny babysitter in the afternoon.
	Stative	Passive	The child was adored by the young and skinny babysitter throughout the holiday.
	Stative	Active	The child adored the young and skinny babysitter throughout the holiday.
24	Eventive	Passive	The soldier was attacked by the armed and dangerous rebel during the battle.
	Eventive	Active	The soldier attacked the armed and dangerous rebel during the battle.
	Stative	Passive	The soldier was loathed by the armed and dangerous rebel after the battle.
	Stative	Active	The soldier loathed the armed and dangerous rebel after the battle.
25	Eventive	Passive	The student was rescued by the aged and revered tutor at 5 o'clock.
	Eventive	Active	The student rescued the aged and revered tutor at 5 o'clock.
	Stative	Passive	The student was respected by the aged and revered tutor before the exam.
	Stative	Active	The student respected the aged and revered tutor before the exam.
26	Eventive	Passive	The dancer was slapped by the oppressive and abusive choreographer at the rehearsal.
	Eventive	Active	The dancer slapped the oppressive and abusive choreographer at the rehearsal.

	Stative	Passive	The dancer was loathed by the oppressive and abusive choreographer before the chat.
	Stative	Active	The dancer loathed the oppressive and abusive choreographer before the chat.
27	Eventive	Passive	The investigator was bribed by the unpleasant and bald bartender on Wednesday evening.
	Eventive	Active	The investigator bribed the unpleasant and bald bartender on Wednesday evening.
	Stative	Passive	The investigator was pitied by the unpleasant and bald bartender since the confession.
	Stative	Active	The investigator pitied the unpleasant and bald bartender since the confession.
28	Eventive	Passive	The guide was saved by the handsome and athletic climber on Sunday afternoon.
	Eventive	Active	The guide saved the handsome and athletic climber on Sunday afternoon.
	Stative	Passive	The guide was envied by the handsome and athletic climber before the trip.
	Stative	Active	The guide envied the handsome and athletic climber before the trip.

7.15 Filler items in Exp. 8.

1 The young couple were recently married after a very long engagement.

2 The researcher had just finished collecting his data two days ago.

3 The captain had fought very bravely in the recent unapproved attacks.

4 The French ambassador is married to a young and bright journalist.

5 The company executive was deeply concerned about the employees' working conditions.

6 The opera singer had studied for five years at Julliard in New York.

7 The brilliant student was the head of the Bridge association of his University.

8 The boy was the son of an influential politician in Washington.

9 The old judge was about to decide on a controversial case.

10 After the employee was promoted he bought champagne for his colleagues.
11 The activist left the debate after a heated discussion with the organizers.
12 The rich lady had just bought an expensive necklace and matching earrings.
13 The boy was kidnapped outside of the prestigious school in the afternoon.
14 The young girl graduated at the top of her class at Oxford University.
15 The journalist interviewed the celebrity about her recent movie about aliens.
16 The writer spent two years in his chalet in the Alps.
17 The therapist discovered the past traumas that the patient had endured.
18 The gangster went to the striptease at a very famous night club.
19 The ghost haunted the woman who was living in the house.
20 The player scored three goals in the last football match against Aston Villa.
21 The door-to-door salesperson married his hard working wife on a rainy day.
22 The barber shaved the actor's hair in preparation for his new role.
23 Santa Claus gave presents to all the children on the nice list.
24 The circus performers entertained various audiences several times while travelling the country.
25 The criminals were fleeing from the police in a stolen car.
26 The temptress was discovered seducing the friar in the old church.
27 The goats that were grazing in the fields belonged to an unfriendly farmer.
28 The princess who lived in an enchanted castle was good at cooking.
29 The severely injured woman was saved by the paramedic in the morning.
30 The irate passenger was removed from the train by the conductor.
31 The solo for the symphony orchestra was played by a cellist.
32 The nobleman was amused by the comedian that performed many jokes.
33 The scientist who experimented on mice had written a famous article.
34 The migrant who tried to enter the country was fleeing from war.
35 The ducks were fed bread by the children on the lake's shore.
36 The burglar went to prison several times throughout the following years.
37 The doctor who gave help to the refugees was very generous.
38 The tailor hired the handy man to paint his house in the village.
39 The slave who had escaped from his cruel master ran through the streets.
40 The exorcist expelled the demon from the possessed girl with his silver cross.

41 The little boy was scared by the clown at the party.

42 The leader felt that the country was heading towards a tough time.

43 The zombie who bit the man was once a compassionate king.

44 The lifeguard who rescued the child had once seen a dolphin.

45 The plumber who spread butter on his toast had several jobs.

46 The chef who served the lord only used seasonally fresh ingredients.

47 The customer who the blonde waitress served left a generous tip.

48 The little mouse who the tabby cat hunted hid in the house.

49 The actress who the teacher admired liked thriller movies and comedies.

50 The model who the artist painted a picture of was tall.

51 The teenager who the policeman ordered to leave had freckles on his nose.

52 The housekeeper who the millionaire hired to clean had great references.

53 The man who the elephant trampled spent many months in hospital.

54 The cat who the big dog was chasing is a Siamese.

55 The rabbit who the little girl fed grass lived in a hutch.

56 The terrorist who the army killed had carried out several attacks.