On-line sentence processing of modified noun-noun compounds

by

Leonard R. P. T. Ruijters

s4393759

Supervisor: Dr Olaf Koeneman

Second reviewer: Prof Dr Ans van Kemenade

Thesis Submitted in Fulfillment

of the Requirements for the Degree of

Bachelor of Arts

Faculty of Arts, English Language & Culture

Radboud University Nijmegen

15 June 2017

Abstract

In the field of sentence processing studies, the ambiguity that occurs in modified noun-noun compound phrases, like *the porcelain egg container*, has been generally ignored. A number of models have been developed that help account for ambiguity in language (Frazier, 1987; Pritchett, 1992; MacDonald et al, 1994; Frazier & Clifton, 1996; Gibson, 1998; Ferreira & Lowder, 2016). However, none of these models seem to have been designed with accounting for [ANN] compound ambiguity in mind. In the present study, differences in processing of disambiguated [ANN] structures were compared by means of self-paced reading task. For the task, sentence-pairs were designed with [A[NN]] structures in one alternative and [[AN]N] in the other, where the semantic properties of the adjective were the disambiguating element. The experiment found no difference between the reading times for the two structures. The present paper further argued that this is evidence in favour of Ferreira and Lowder (2016)'s model, and MacDonald et al. (1994)'s model.

Keywords: Sentence processing, ambiguity, compounding, adjectives, self-paced reading Good-Enough processing, Construal, Syntactic Prediction Locality, reading times.

Table of content

1. Introduction

Sentence processing is a regularly studied topic in the field of psycholinguistics. Ambiguity in particular has been analysed thoroughly when it comes to sentence processing. Garden-path sentences, such as *The horse raced past the barn fell* (Frazier, 1987) have received a great amount of attention. The same can be said for sentences with ambiguously assigned modifying relative clauses, for instance *Someone shot the servant of the actress who was on the balcony* (Cuetos & Mitchel, 1988), or for sentences with ambiguously attached prepositional phrases, like *The man watched the woman with the binoculars* (Ferreira et al, 2002). Most models of sentence processing that have been developed are able to account for the existence of these kinds of ambiguity in one way or another, as well as account for behavioural data that was found during experiments with participants reading such constructions.

However, one type of ambiguity that has not yet been discussed in the literature is that of modified noun-noun compounds. A phrase such as *the porcelain egg container* can be interpreted in two ways, namely *a porcelain container for eggs* or *a container for porcelain eggs*. While finding an off-line preference for interpreting such a phrase should not be too difficult, looking at on-line effects would be more interesting for the study of sentence processing. On-line effects would be especially relevant to analyse when reviewing the body of models. These models focus strongly on the on-line processes that occur during language comprehension. Seeing how modified noun-noun compounds fit within these models could provide extra evidence in favour of or against some of these models. The fact that they have been overlooked so far might even prove to be disastrous for some of these models, as they may not be able to account for findings resulting from studies looking into the compound ambiguity.

In this study, the effects will be analysed of processing the structurally ambiguous modified noun-noun compound phrase with a disambiguating adjective. By disambiguating a phrase in this manner, it might become apparent which way of interpreting the modified noun-noun compound is preferred for processing by the reader (one where the adjective modifies the first noun, i.e. an [[AN]N] structure, or one where the adjective modifies the second noun, i.e. an [A[NN]] structure), if there even is a preference. Disambiguating the structure should be able to find a preference, as it forces the reader to interpret the sentence in a way that either agrees with this preference, or disagrees with it. If the forced structure disagrees with the interpretation that processing prefers, then this should cause processing

difficulties, and this should be reflected by increased reading times. Similarly, a forced interpretation in line with preferences should be accompanied by shorter reading times. If processing has no preference for one structure over the other, then reading times should reflect this as well, as reading times should not differ for the two types of structures in that case. Therefore, a self-paced reading task will be utilised so an on-line preference for one structure over the other can be measured, and what structure might be preferred. The findings of this experiment will then be compared to the predictions that some of these models make on the processing of modified noun-noun compounds, to see if these models still hold up.

The two main questions this study will focus on are: 1) Is there an on-line preference for how to interpret modified noun-noun compounds, and if so, which structure is preferred, and 2) how do the reading times for the different disambiguated modified noun-noun compounds fit within the existing literature on sentence processing and the models that have been proposed thus far?

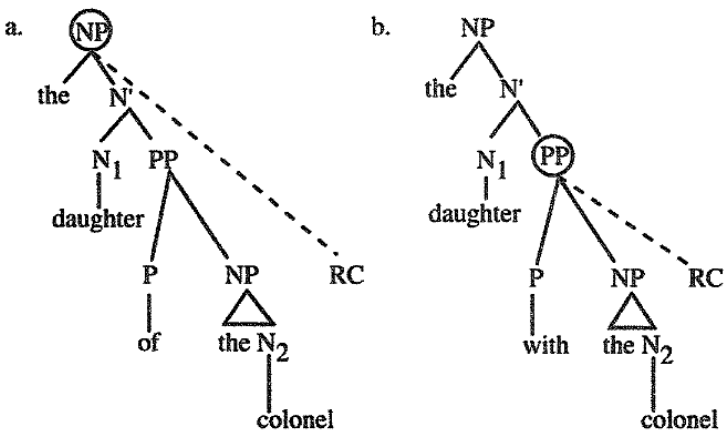
2. Background

Many models of sentence processing have been developed to describe how sentences are processed, based on the extensive research that has been conducted on the topic on ambiguity. Examples of these models are Frazier (1987)'s Garden-Path model, Frazier & Clifton (1996)'s Construal, Pritchett (1992)'s Generalised Theta Attachment model, MacDonald et al. (1994)'s Constraint-based theory, Gibson (1998)'s Syntactic Prediction Locality Theory, and the Good-Enough model (as described in Ferreira & Lowder, 2016). However, in spite of the vast body of studies on ambiguity, one kind of ambiguity has so far been mostly ignored by researchers in the field of sentence processing. This is the modified noun-noun compound ambiguity. Word category ambiguity has been researched in context of noun-noun compound (e.g. MacDonald; 1993, 1994), where phrases were used that were ambiguous between a simple verb phrase and a noun-noun compounds (e.g. *the desert trains could resupply the camp* vs. *the desert trains soldiers to be tough*), but this was not in combination with adjectives and it did not look at the underlying ambiguous structure. What is most remarkable about modified noun-noun compounds is that, in contrast to most forms of ambiguity previously studied, the ambiguous element of the modified noun-noun compound, the adjective, is introduced before the structure that facilitates its ambiguity, the double noun. To illustrate, in *The man shot the servant of the actress who was on the balcony*, the relative clause *who was on the balcony* is introduced after the two nouns that it can

modify (thereby creating ambiguity). In the case of *The porcelain egg container was on the table*, *porcelain* is introduced before the two nouns it can modify. The ambiguous element is introduced before the ambiguity is. Since this structure is in essence the inversion of the structures that are usually analysed in the field of sentence processing, seeing how the different interpretations for [ANN] phrases are processed could have very serious implications for the models that have been developed thus far. An off-line in-class experiment by A. Foltz (2016) suggested that there is an off-line preference for the [A[NN]] structure, but more relevant for the models of sentence processing is to see what the on-line preference is, if there is one.

One of the best-known models in sentence processing is the Garden-Path model (Frazier, 1987), later turned to Construal (Frazier & Clifton, 1996). The theory relies on the idea that a parser analyses a phrase based on what analysis is available to the reader first (Frazier & Clifton, 1997). The principles proposed by the model (Minimal Attachment and Late Closure) are therefore not so much constraints by which the parser must abide, as they are a pattern arising from deciding on the first available analysis of a phrase. Construal differentiates between primary and non-primary phrases. Primary phrases are defined as the subject and main predicate of a finite clause, or as complements and constituents obligatory to such phrases (Frazier & Clifton, 1997). The cause of ambiguity here is ascribed to what Frazier and Clifton (1997) describe as underspecification. Underspecification can occur if the particular element is a non-primary phrase that is not initially interpreted as a primary phrase. According to Frazier and Clifton (1997), such underspecified phrases can associate to the last theta-assigner or its projection, as illustrated using the relative clause in (1) (taken from Frazier & Clifton, 1997).

(1)

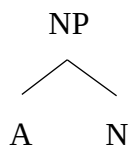


According to Frazier and Clifton (1997), the relative clause in (1) can associate to the last theta-assigner or a projection of it. In (1a), this means the relative clause can associate to the NP of *daughter*, but it can also associate with the NP of *colonel*, creating ambiguity. Frazier and Clifton (1997) argue that the case of (1b) is not ambiguous, because they suggest *with* is a theta-assigner, and as such, the prepositional phrase cannot associate with the node above *with*, according to the Construal model. Frazier and Clifton (1997) speculate that, since an alternative formulation to the structure in (1a) is accessible (*the colonel's daughter*), the interpretation is in favour of associating the relative clause with the NP_2, but an association with NP_1 is not impossible, which causes ambiguity. Frazier and Clifton (1997) do not parametrise the phenomenon of this measured off-line preference, as this association preference differs across languages (Cuetos & Mitchell, 1988). In other words, following the Construal model, ambiguity comes from non-essential elements of a sentence being able to associate with more than one item, where there is no theta-assigner to limit the association.

For this study, it is important to speculate how Construal would predict that [ANN] compounds are most likely interpreted on-line, before the off-line judgment is made, so that this prediction can be compared with the results of the experiment. Construal suggests that a parser decides on the first available analysis as soon as it becomes available, also known as Minimal Attachment. In case this first analysis turns out to be a misanalysis, it is revised at a later stage. This revision is reflected in reading times.

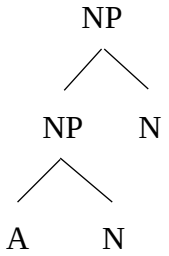
Due to Minimal Attachment, it is safe to assume Construal predicts a preference for the [[AN]N] analysis over the [A[NN]] analysis. The adjective and first noun of a modified noun-noun phrase are presented first, and assuming the principles of Construal are universal when it comes to resolving ambiguity, they are analysed as being together, giving rise to the structure in (2).

(2)



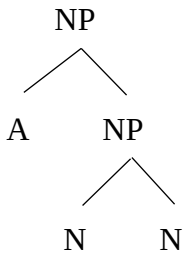
When the second noun is introduced, Construal suggests the parser prefers to keep this structure intact if possible, and only change as little as possible. This minimal change gives rise to the structure illustrated in (3), as here the structure of (2) is left intact.

(3)



In order to interpret the adjective as modifying the second noun, a reanalysis of (2) is required, as can be seen in (4). As is shown, this analysis requires a reconstruction, contrary to the analysis in (3), since the original intepretation has to be abandoned. As such, following Construal, the structure in (3) is preferred over the structure in (4).

(4)



It could be argued that the Late Closure principle applies to these structures, as they contain an equal amount of structural nodes (although (4) does not comply to the idea that nodes are integrated as straightforward as possible). Late Closure applies when two alternatives for association are equally likely according to Minimal Attachment. Should Late Closure apply here, rather than Minimal Attachment, then (4) would turn out to be the more preferred structure of modified noun-noun compounds. Whether (3) or (4) is the preferred structure is somewhat uncertain, but a strong preference for one over the other would work in favour of Construal. Other models, such as the Good-Enough model (as discussed in Ferreira & Lowder, 2016) could explain the absence of a preference more robustly than the Construal model could, but such models have a harder time explaining the existence of a structural on-line preference. In other words, the Construal model predicts that reading times for [A[NN]] and [[AN]N] differ, and findings in the present study that comply with this prediction are favourable to Construal.

Another model is provided by Pritchett (1992). Pritchett (1992)'s model of Generalised Theta Attachment suggests that ambiguity results from uncertainty in theta-roles, and that especially the Garden-Path effect is a consequence of miss-assigning theta-roles. It is assumed that theta-roles are assigned at every stage of the parse, and that leaving a theta-role empty or unassigned is avoided as much as possible. Take for example the sentence *The horse raced past the barn fell*. When the verb *raced* is encountered, its theta-roles are activated (relevant for the illustration are AGENT and PATIENT). Assigning the role of AGENT to *the horse* allows for the possibility of PATIENT to be assigned later on. Assigning PATIENT to *the horse* however, would likely mean the AGENT role would go unassigned. Therefore, interpreting *the horse* as AGENT of *raced* is the more preferable option according to Generalised Theta Attachment. It is only later, when *fell*'s theta-roles go unassigned that it becomes apparent that *the horse* is PATIENT of *raced*. The ambiguity is resolved by reassigning theta-roles within the sentence.

Since ambiguity resolution relies on theta-attachment in this model, it does not make any predictions on the preference of the on-line interpretation of noun-noun compounds, because the adjectives have no theta-roles to assign to the nouns they modify. That being said, it could be argued that a non-preference either way for the [ANN] structure is evidence in favour of the General Theta Attachment model, as theta attachment is not involved in the processing of these phrases[1], and so no preference arises through this method[2]. Other models actively predict a non-preference (e.g. Ferreira & Lowder, 2016), rather than keeping from predicting a preference. A non-preference would work in favour of such models more strongly.

MacDonald et al. (1994) take a different approach, namely a constraint-based approach. They suggest that context, both within a sentence and within the world in which it is uttered, are the basis for ambiguity resolution, even at on-line speeds. Lexical properties, such as frequency of use, the lexical items a word selects for, and the part of the sentence that precedes a word all help contribute to the resolution of ambiguity in sentences. For example, in (5), *The witness* in (5a) sets the reader up for different expectations than *The evidence* in (5b), because the reader knows that it is more likely that *The evidence* is the THEME of *examined*, whereas *The witness* is also a very likely AGENT of examined.

[1] This is different from Frazier & Clifton (1997)'s approach to theta-assigning, since they argue that theta-assigning limits the distance over which modifying phrases can associate with an item thus preventing ambiguity, rather than suggesting that theta-assigning in itself causes ambiguity.

[2] It could be argued that adjectives assign theta-roles (Cinque, 1990). If that were the case, then the [[AN]N]-structure would be preferred on-line over the [A[NN]]-structure, as the theta-role is either more quickly assigned in the case of [[AN]N], or there is no reanalysis required for [[AN]N]-structures.

(5) a. The witness examined by the lawyer was lying.
b. The evidence examined by the lawyer was conclusive.

Any responses from the listener that indicate a mismatch between what they are exposed to and what they expected to hear is not a result of a preferred structure that is being violated. Instead, responses stem from an on-line prediction made by the reader, that is disproven at a later stage in the sentence. These predictions are on a case-by-case basis, so that (5a) and (5b) are subject to different on-line predictions. Consider the three interpretations when hearing the phrase *John cooked* (taken from MacDonald et al., 1994). Either *John cooked something* (however, *something* hasn't yet been reached in the sentence), *John cooked* (whereas his friends did other things), or *John cooked* (as a result of a tragic encounter with a cannibal). Depending on the context that precedes this phrase (either in the sentence or in the discourse), the way the phrase is initially interpreted changes. Does this initial interpretation not match the intended interpretation that becomes apparent later on, then an effect is seen. Depending on whether there is an equibias between two alternatives (so that two options not accounting for context are equally likely to occur according to frequency of use, for example), then context can eliminate the interference from the alternative that is less likely to occur in that context. However, a non-equibiased frequency advantage is not entirely eliminated by context. In other words, the presence of a more frequent structure in a language (for illustration's sake for instance [A[NN]]) is always measurable in contexts that would cause the reader to predict a less-frequently occurring structure (e.g. [[AN]N]).

The constraint-based theory could apply to modified noun-noun compound ambiguity as well, in one of two ways. The first way it could apply is that an adjective can select for a noun with certain properties (e.g. animacy). If the first noun is animate, then the adjective will bind to that noun, giving rise to an [AN] structure, which the second noun then binds to in a [[AN]N] manner. If the first noun is not animate, however, then the listener could anticipate another noun that is animate, and so the [A[NN]] structure might be preferred at that point.

The other way in which the constraint-based theory could apply to modified noun-noun compound ambiguity is in terms of preceding context. Take for example *porcelain egg container*. If the context suggests the adjective modifies the second noun, as in (6a), then the listener anticipates this interpretation and shows a preference for it. The same goes for (6b) with regards to the first noun.

(6) a. I only have one porcelain item and that's my porcelain egg container.
b. I have one container for chicken eggs and one porcelain egg container.

In order to measure a preference in structure, however, the non-equibiased frequency advantage comes into play. Should (6a) suffer less interference from [[AN]N] structures than (6b) from [A[NN]] structures, as could be illustrated by (6a) having shorter reading times than (6b), then that shows that (6a) is a more frequent (and as a result a more preferred) structure. Should the ratio of interference be the other way around, then of course (6b) is the more preferred structure of the two. If there is no measurable effect of interference in either structure, then that would indicate that there is no on-line preference. The only problem lies in testing this, as there are many factors that come into play according to the constraint-based theory. The frequency advantage for the preferred structure might, for example, be in favour of (6a), but the frequency advantage of *egg* following *porcelain* might be in favour of (6b). As such, there are far too many factors for which to control that would fit within the scope of this study.

Gibson (1998) builds on this idea with Syntactic Prediction Locality Theory, and suggests that, aside from frequency and probability, processing costs associated with sentence processing also allow researchers to make predictions on sentence processing. The Syntactic Prediction Locality Theory posits that the two major factors in sentence processing are memory cost and integration cost. These costs on the processor result from making predictions about the structure of a sentence, maintaining that prediction, and integrating new items into it. Simplified, when a word in a sentence is encountered, it provides information on the meaning of the sentence. Every word provides a limited amount of possible continuations, and with each additional word, the number of possible interpretations is reduced, until only a very restricted set of interpretations remains. It is this process of predicting and eliminating interpretations that the Syntactic Prediction Locality Theory focusses on.

Syntactic Prediction Locality Theory also makes predictions on ambiguous utterances. The model is based around the idea that the process that is the least taxing with regards to memory and integration costs leads to the eventual interpretation of a sentence, and this holds true for ambiguity as well. The model suggests that, overall, ambiguous utterances have a preferred interpretation where items are bound as locally as possible, as opposed to an interpretation with a more distal binding of nodes. Simply put, items that modify elements in a sentence, such as relative clauses, bind to the earliest possible node that is presented to the

reader, according to the Syntactic Prediction Locality Theory. This is illustrated in filler-gap relationships, as occur in (7) (taken from Gibson, 1998).

(7) a. The reporter who attacked the senator admitted the error.
b. The reporter who the senator attacked admitted the error.

Reading times for sentences like (7a) were shorter than for (7b), and Gibson (1998) argues that this is because the filler-gap relationship in (7a) is shorter than in (7b) (the gap positioned before *attacked* in (7a) vs. the gap positioned after *attacked* in (7b)). Locally binding an item like this frees up the faculties that facilitate sentence processing (e.g. working memory), which is why binding as early on in the sentence as possible is the least taxing process according to the Syntactic Prediction Locality Theory.

Gibson (1998) has suggested that processing a noun-noun compound is not substantially more taxing on the processor than a phrase containing a single noun and a verb, following amongst others MacDonald (1993). Because there is supposedly no effect of the noun-noun compound on the processing, Syntactic Prediction Locality Theory suggests that the preferred interpretation is the same as MacDonald et al. (1994)'s Constraint-based model. However, since the modifier in an [ANN] structure forces the listener to make a prediction, and this prediction has to be maintained (and likely has to be revised in either interpretation of the [ANN] structure as well), this could cause processing difficulties. In an [[AN]N] structure, the second noun might trigger a revision of the structure predicted so far, when a noun is introduced to a perfectly correct and complete phrase. Meanwhile, the [A[NN]] structure might trigger a revision when encountering the first noun, after a mismatch between the adjective and noun proves to be a misjudgement in the prediction. It is also possible that both of these revisions occur when encountering either the [A[NN]] structure or the [[AN]N] structure. This would make both interpretations equally taxing, as Gibson (1998) suggests. However, it should be pointed out that the [ANN] structures should not be as taxing as [AAN] structures. If they are as taxing as [AAN] structures, then that indicates that [AAN] structures also need revision for correct interpretation. If that were so, then every additional adjective in a noun phrase would increase the processing cost. That would mean that noun phrases with longer strings of adjectives are harder to interpret. However, a sentence such as (8) does not seem to suggest that.

(8) The man pet the friendly playful long-haired blue-eyed drooling dog lovingly.

In short, Gibson (1998)'s Syntactic Prediction Locality Theory predicts that [A[NN]] structures and [[AN]N] structures elicit similar reading times, whereas [AAN] structures should yield shorter reading times, in order to prevent implications that sentences like (8) are difficult sentences to process. Here it also differs from the predictions made by MacDonald et al. (1994)'s model, which makes no prediction on processing difference between [ANN] and [AAN] items.

Finally, a more recent approach to sentence processing is offered by Ferreira and Lowder (2016). Their model reconciles Information Structure, Prediction, and Good-Enough processing. According to Ferreira and Lowder (2016), sentences are structured following a given-new structure, where the start of a sentence usually contains given discourse information, and new discourse information is provided towards the end. Different processing methods are then employed to process the two different types of information.

Following Ferreira and Lowder (2016), new discourse information is processed with the help of prediction. Prediction, they argue, is not the end-goal of language, as had been previously suggested, but instead facilitates comprehension, in that the listener is preparing for what element might come next. This preparation then makes it easier for the listener to integrate new information into existing discourse.

More relevant for this study is how given information is processed, as Ferreira and Lowder (2016) suggest this follows the Good-Enough model. Given information is information previously mentioned in the discourse. This previously discussed information has to be linked to a referent in the sentence, so that discourse cohesion is made possible. However, since given information generally holds even if the referent is not interpreted every time it is encountered, Ferreira and Lowder (2016) suggest that processing power is limited for given information. Ferreira and Lowder (2016) speculate that this is in order to save more capacity for the prediction process of new information. This restricted processing power entails that given information is often processed superficially, so that any misinterpretations are generally glossed over by the listener.

This glossing over of misinterpretations in the model is especially relevant for studying ambiguity. The Good-Enough model suggests that ambiguous elements in a sentence are not necessarily reinterpreted to fit the intended meaning, but are only interpreted to the extent where it is possible to relate the ambiguous information to the relevant antecedent in the discourse (Ferreira & Lowder, 2016). In other words, ambiguity does not pose a problem for the interpreter of a sentence on a processing level, as it is only processed

on a surface level. Because of this superficial processing, there should not necessarily be a preference for an interpretation of the [ANN] phrase, at least not an on-line preference. Presumably, there also should not be a difference in processing difficulty between [ANN] structures and [AAN] structures, as these structures are of similar complexity, although the presence of different grammatical items makes it hard to strongly argue this is the case.

3. Method

To test whether English speakers have a preferred structure for the interpretation of modified noun-noun compounds, a self-paced reading task was designed. Self-paced reading tasks have been used in the past to test whether participants show a behavioural processing difficulty for the interpretation of ambiguous utterances (e.g. MacDonald; 1993, 1994). If participants have trouble processing a sentence or phrase, the reading times associated with those segments should be longer compared to a control. In other words, a sentence with a longer reading time than others requires more effort. The item with the shorter reading time is less taxing on the processing faculty of the reader, and as such has their preference in interpretation. This study is built on that same principle.

In order to find out which [ANN] interpretation is harder to process, the items for a self-paced reading task were designed to force either an [A[NN]] interpretation, or an [[AN]N] interpretation. Since this is a pilot study, the design has been kept relatively simple, and the items involve semantic disambiguation. That is to say, the adjectives used in the target items have a clear semantic relation to either the first noun in the first condition, or the second noun in the second condition, but never both. This way, an interpretation is forced. If the reading time for the first condition is longer than the other condition, it implies the forced interpretation in the first condition requires more of the reader’s processing power. This suggests it is in conflict with the structure predicted by the processer. These predictions can vary depending on which model is assumed. Longer reading times for one condition compared to the other condition indicate a preference for the item with the shorter reading time. Of course, some models do not predict a difference in processing difficulty, and an absence of difference in the reading times could be seen as evidence in favour of those models.

Twenty target items were designed for this experiment. These items were copied and modelled over two conditions, forming twenty sentence-pairs: twenty sentences containing [A[NN]] compounds paired with twenty sentences containing [[AN]N] compounds (See

Appendix A). Ten items of each condition were then combined, in a manner that each group included ten sentences of each type (= 20), and counter-balanced across two groups. In order to allow comparisons between the two groups, ten control items were included to each group, involving sentences with a non-compounded ([AAN]) construction. These sentences were otherwise similarly structured to the target items and identical over both groups. By including these items, it was possible to compare the groups, as well as analyse between-subject reading pace as an influencing factor, and they reduced the number of items required to compare from one hundred to thirty. What is more, since the underlying structure is less complicated and less prone to ambiguity, a model like Syntactic Prediction Locality Theory predicts that these items are easier to process. This helps in answering the second research question on which model best predicts the results. In order to ensure the validity of the data, 72 filler items were added to both groups to prevent participants from inferring the purpose of the study (See Appendix B). In addition, every fifth item was followed by a content-question about one of the previous filler items, in order to ensure participants interpreted the sentences rather than read the sentences superficially. The questions focussed on filler items, to prevent the attention that was drawn by these questions from influencing the reading times of the target items. Finally, a question was added to the end, inquiring if the participants could guess what the purpose of the experiment was.

To ensure that differences in reading times could only be attributed to the different interpretations, a number of random factors were recorded for the statistical analysis. A questionnaire was included at the start to ask the participants for their age, gender, native dialect and any second languages. Furthermore, the ratio by which frequency of use differed for the adjectives in target item pairs was added as a separate variable. The frequency of use was found using the Subtlex-UK database (The Centre for Reading Research, 2013). Items were adjusted to minimize differences in frequency of use as much as possible. The frequency of use of the adjectives for the [A[NN]] items was then divided by the frequency of use of the adjectives for the [[AN]N] items to find the relevant ratio. This way, it was possible to analyse the effect of frequency of use during the statistical analysis. Unfortunately, it was not possible to rule out this factor during the design-stage of the experiment using Subtlex as a source, because there was no baseline by which the actual frequencies could be compared. Therefore, it was added as a variable in the analysis. Items were not controlled for word-pair frequency; see Discussion.

In order to test the validity of the target items, three native speakers were asked to judge both the interpretation of the target items off-line, as well as the overall acceptability of

the target and control items on a seven-point scale. Due to the subjective nature of the judgement task, the given values varied a lot. Therefore, items have been singled out if two or more judges agreed on either a low acceptability score (lower than 4) or an undesirable interpretation (the opposite of the intended interpretation, an ambiguous interpretation, or an interpretation where the adjective modified neither nouns). Based on these findings, target items 1, 8, and 19, as well as filler 9 were modified. Target items 9 and 20 were excluded, as modifying these into appropriate items proved to be too difficult, leaving eighteen target items included in the study. The modified items were then judged once again by a native speaker, and deemed acceptable for this study.

10 native speakers of English (7 female; 3 Male) were found using the Max Planck Institute for Psycholinguistics participant database. Participants ranged from 19 to 54 years old (M =31; SD =11.12). One participant was left-handed. While usual conduct dictates that left-handed individuals are excluded from experiments like this one, due to the small number of participants, this individual was left in the sample to increase the power of the sentence-type effect. Six participant spoke British English, three spoke American English, and one spoke a Creole dialect. Three participants also spoke Dutch as a foreign language, two Spanish (one of which also spoke Dutch), one French, and one German. Participants were invited to the lab and sat down in front of a computer. They were instructed they could progress the sentence using the NUM4 key, and answer questions using NUM4 and NUM5. The NUMpad was used to elicit use of the right hand. A response delay was added before every question to prevent participants from accidently clicking through a question. Participants were presented three example sentences and an example question, before the experiment started.

Participants were informed beforehand that the use of the data gathered in this experiment complied with the Code of Ethics of the Max Planck Institute. This study was approved by the Ethics Committee of Social Sciences (ECSW).

4. Results

To test whether there was an effect of frequency on the reading times, a two-way ANOVA was performed, using the independent variables sentence-type ([A[NN]] or [[AN]N]) and frequency ratio, and the dependent variable reading times. The ANOVA showed that there was no main effect of frequency ($F(17,144) = 1.125$, $p = 0.336$, $\eta^2 = 0.117$). What this indicates is that there was no difference in reading times between sentence-pairs that had a high frequency ratio ($f_{[A[NN]]}/f_{[[AN]N]}$) and those that had a low ratio. In other words, sentence-pairs with a more frequent adjective in the [A[NN]] variant were not read more quickly or slowly than sentence-pairs with a more frequent adjective in the [[AN]N] variant.

There also was no interaction effect between frequency-ratio and sentence-type ($F(17,144) = 0.223$, $p = 1$, $\eta^2 = 0.026$). This shows that within a sentence-pair, a frequency-ratio in favour of one of the members of a pair did not influence reading times compared to the other. Figure 1 illustrates the reading times in a graph, where the blue line shows the reading times of the [A[NN]] sentences over the frequency-ratio of the words in the sentences, and the red line the [[AN]N] sentences. As can be seen by the error bars, there is no significant difference, which means the variance of frequency between adjectives of the sentence-pairs did not influence the reading times.

For the main analysis, the variables Native Dialect and Second Languages were transformed into the variables British and Bilingual, so a group of non-British participants and a and a group of bilingual participants were formed. This was done in order to increase the predictive power of the analysis of a small sample.

During the main analysis, to test whether there was a difference in reading times between the two types of sentences, another two-way ANOVA was performed. The analysis included the fixed factors sentence-type and sentence-pair. The ANOVA showed that there was no main-effect of type on the reading times of the sentences ($F(2,225) = 0.159$, $p = 0.853$, $\eta^2 = 0.001$). In other words, whether sentences contained the [A[NN]] compound or the [[AN]N] compound did not influence the reading times of the sentence. This can also be seen in figure 2. There was no interaction-effect between sentence-pair and type either ($F(17,255) = 0.228$, $p = 1.00$, $\eta^2 = 0.017$). The lack of an interaction-effect indicates that the difference in reading times within sentence-pairs does not differ between sentence-pairs. Figure 3 helps to illustrate the absence of an interaction effect. There is a measured main-effect of sentence ($F(26,225) = 2.281$, $p = 0.001$, $\eta^2 = 0.209$). This effect shows that the average

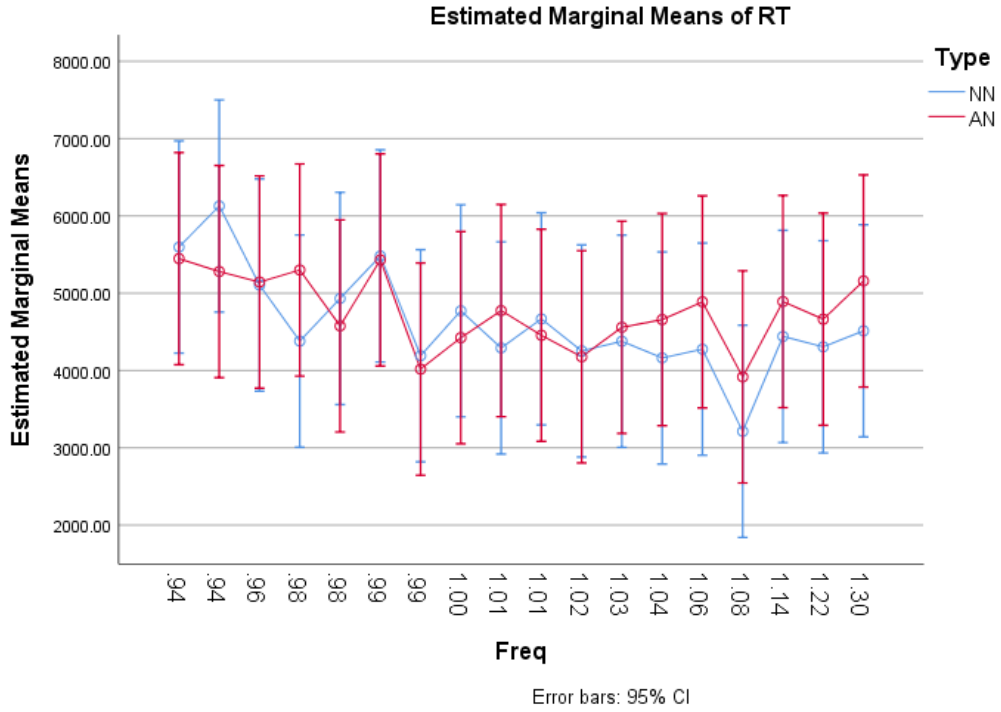


Figure 1: The reading times per sentence over the frequency-ratio of the target word within that sentence, split by [A[NN]] and [[AN]N]. As is shown, there is no significant difference between high and low ratio sentences, and no difference between [A[NN]] and [[AN]N] for each frequency-ratio.

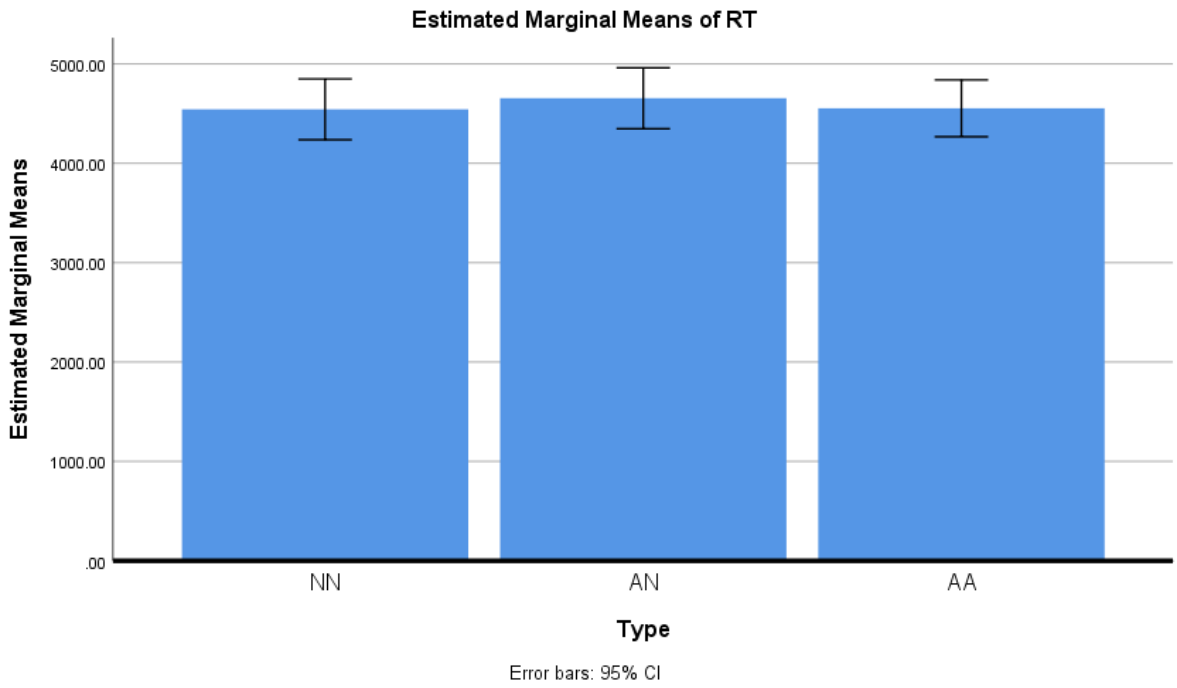


Figure 2: The figure shows that there is no significant difference in reading times between the [A[NN]] type sentences, the [[AN]N] type sentences, and the control sentences.

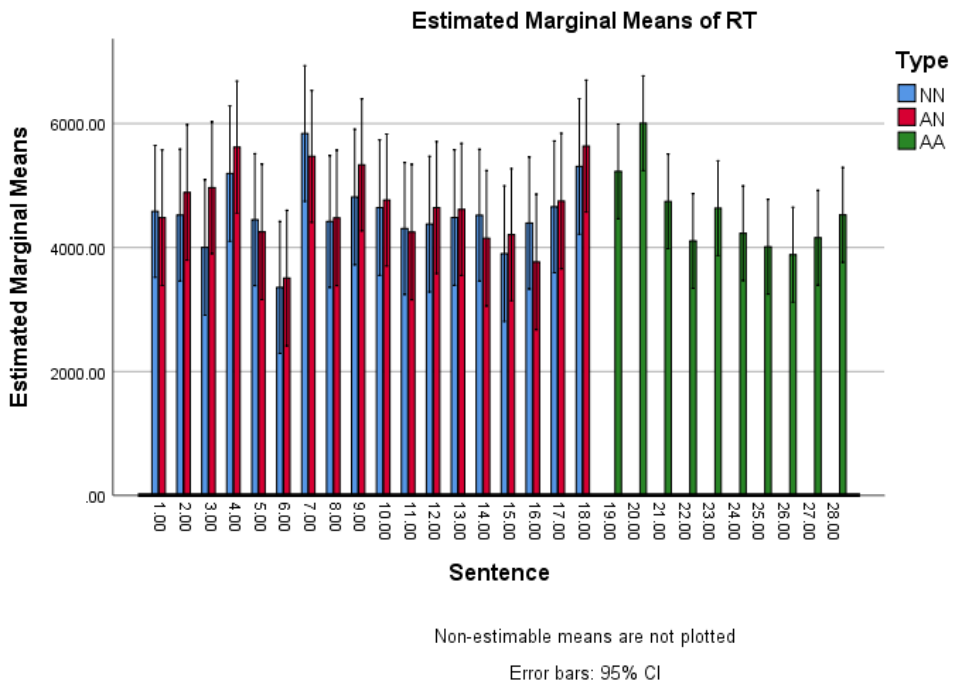


Figure 3: This figure illustrates the interaction between sentence-pair and sentence-type. The difference between the two bars for every sentence does differ significantly between different pairs (i.e. the blue bar is not significantly shorter than the red in one pair compared to the blue and red bar in another pair, or vice versa).

reading times of one or more individual sentences differed from one or more other sentences. This effect tells us that certain sentences were easier to read than others, regardless of the modified noun-noun compounds.

In addition to the fixed factors sentence-type and sentence-pair, the random factors gender, British, and bilingualism were included in the ANOVA. The ANOVA showed that there was no main-effect of gender on the reading times($F(1,2014)=3.586$, $p=0.198$, $\eta2=0.640$), but that there was a main-effect for Britishness ($F(1,2057)=96.180$, $p=0.009$, $\eta2=0.979$), as well as a main-effect for bilingualism ($F(1,2028)=125.867$, $p=0.007$, $\eta2=0.984$) on the reading times. Figure 4 demonstrates that non-Brits had a significantly shorter reading time overall than Brits had, and figure 5 shows that bilingual speakers had a significantly shorter reading time overall than monolingual speakers had. Figure 6 demonstrates that there is no difference in reading time between men and women.

However, while there was a main-effect of British and of bilingualism, there was no interaction-effect for both type and British ($F(2,225)=0.486$, $p=0.616$, $\eta^2=0.004$) and type and bilingualism ($F(2,225)=0.290$, $p=0.748$, $\eta^2=0.003$). There was no interaction-effect for gender

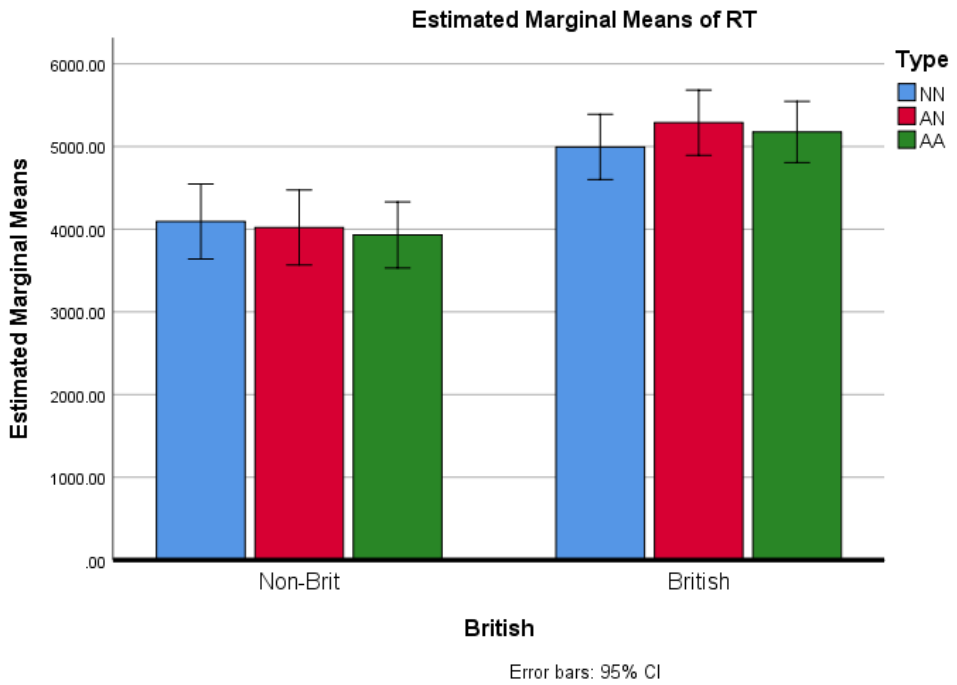


Figure 4: Figure 4 shows that there is a difference in mean reading times of Brits vs. non-Brits (namely that non-Brits read faster than Brits). However, there is no difference in reading times within either group over the three types of sentences.

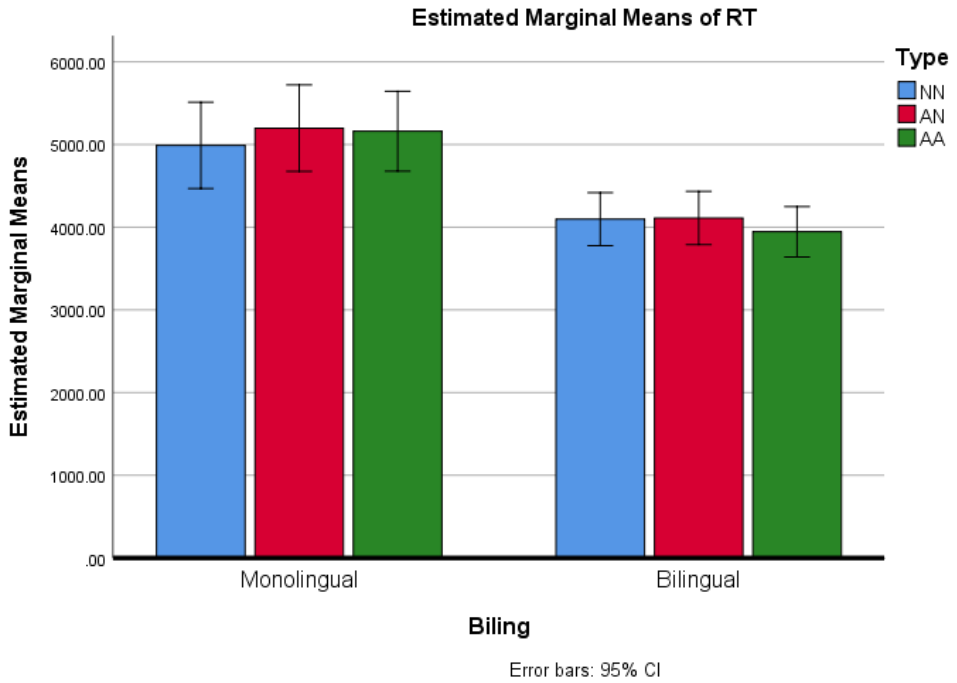


Figure 5: Here is illustrated that bilingual speakers have faster reading times than monolingual speakers. However, as was the case with Brits vs. non-Brits, there is no difference in reading times within groups when it comes to the three types of constructions.

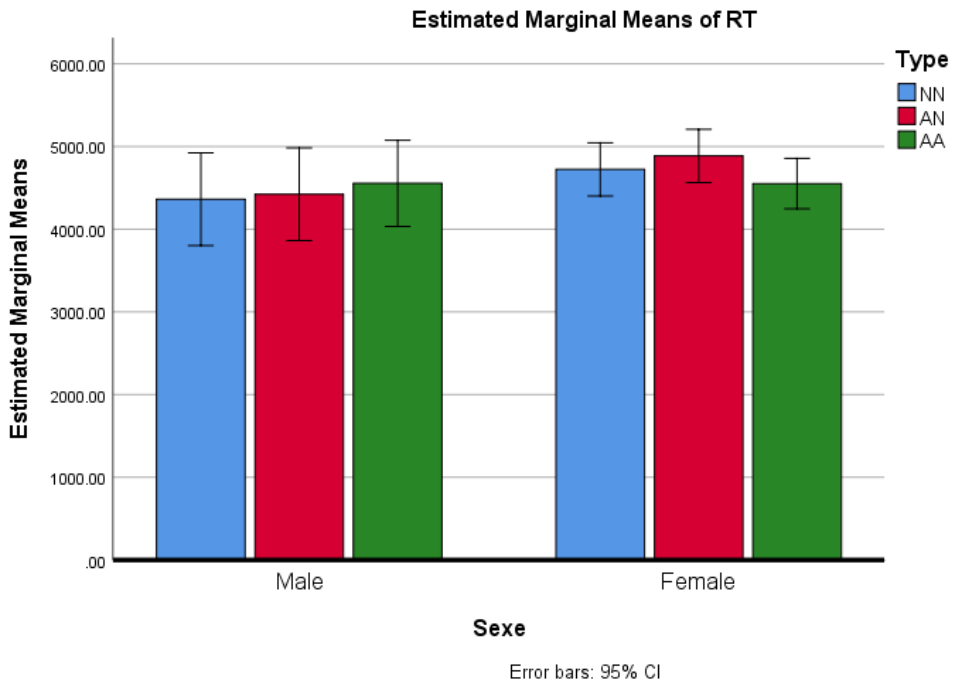


Figure 6: This bar-graph illistrates that there is neither a main-effect for gender, as the average reading times are at around the same level, nor an interaction-effect between gender and sentence-type.

and type on reading time either ($F(2,225)=0.550$, $p=0.577$, $\eta^2=0.005$). This can be seen in figures 4, 5, and 6 as well, where there is no significant difference in reading times for the three sentence-types within the groups for brits vs. non-brits, bilinguals vs. monolinguals, and men vs. women.

5. Discussion

As discussed in the results section, there was no main-effect of sentence-type on the reading times of the sentences, nor was there an interaction-effect between sentence-type and sentence. These results showed that within sentence-pairs, there was no significant difference in reading times for any of the items. This lack of a difference could not be the result of using items with a strongly varying frequency, as frequency did not show a main-effect on the reading times, nor did frequency and sentence-type show an interaction effect. This showed that a frequency-ratio skewed in favour of one interpretation of [ANN] in a sentence-pair over the other did not influence reading times for these items, as frequency did not affect reading times between and within sentence-pairs. A participant's gender did not influence their reading times at all, while a participant's native dialect (British or non-British) and their bilingualism did affect reading times. However, there was no interaction-effect between these factors and sentence-type, so it simply meant that one group read faster than the other overall,

but not in any specific sentence-types. In other words, there was no on-line difference found between [A[NN]] sentences and [[AN]N] sentences in this study. There was also no significant difference between the target conditions ([A[NN]] and [[AN]N]) and the structurally unambiguous [AAN] sentences.

Before looking at the models discussed on the background section, it should be mentioned that it is still possible that future studies similar to this one might find a significant effect of any of these factors. As the partial η^2 for most factors indicate, the ANOVA lacks in power ($\eta^2 < 0.1$). This is a result of the small number of participants (n=10). As a result, small differences in the means of certain groups might go unnoticed due to naturally occurring error-rates, so that a factor that might actually have a significant effect on the reading times is missed. In order to properly study the effects of the different structures used in this study, a larger sample-size is required. However, due to time-constraints, it was not possible to include more participants in the study.

With that being said, on to the goal of this study of looking at how the findings of the experiment compare to the predictions made by the sentence processing models previously constructed. The results of the experiment seem to favour Ferreira and Lowder (2016)'s model and MacDonald et al. (1994)'s constraint-based model the most. The fact that there was no difference in reading times between the two target conditions, nor the target conditions and the control condition, indicates that the two target conditions do not cause any more processing difficulty than structurally similar, less complex conditions. This could mean that these sentences either do not cause any difficulty in processing, or that all of these sentences cause difficulty in processing.

A lack of difficulty in both sentence-types could be indicative of Good-Enough processing, a favourable prospect for Ferreira and Lowder (2016). The participants did not process the sentences any more than was required in order to gain a global understanding of what could be meant, rather than understand what was meant in detail. As a result, both target-structures, as well as the controls, were parsed in a similar superficial way. This is reflected by the fact that reading times were similar for all conditions. This could even go for the control items, where the double adjective might be processed superficially, or perhaps the processing of uncomplicated sentences and superficial sentence processing are similar processes in general. Conceptually, this makes sense, as it is reasonable to assume that when a sentence is processed superficially, it is processed as an uncomplicated sentence, in order to save as much processing-power for other tasks. Alternatively, it is also possible that even uncomplicated sentences are processed superficially, so that in both cases an inadequate

representation is constructed by the parser, leading to similar reading times. The only difference is that for uncomplicated structures, only one possible interpretation is present, whereas in the target items here, two interpretations are possible (albeit not equally plausible). Structurally they can be the same, and so it is possible they are processed the same, under Good-Enough processing.

A similar approach to analyse the data works for MacDonald et al. (1994)'s Constraint-based model. However, instead of attributing the fact that there is no difference in reading times between the two sentence-types to Good-Enough processing, it could be attributed to the fact that there is no frequency advantage for one structure over the other. According to MacDonald et al. (1994)'s Constraint-based model, if there is no frequency advantage for one structure over the other, then reading times for both structures should be similar. This is because if there is no frequency advantage, then both structures are predicted equally strong, and so regardless of which structure is required for the interpretation, both structures are available to the reader. That there is no effect of frequency-ratio between pairs on the reading times also supports this view, because then at least that factor can be excluded as an influence on the processing cost, as the Constrained-based model suggests. The fact that there is no difference between [ANN] and [AAN] items can be explained by this model as well. The target and control items are of the same length. If the first adjective has an equibias for the [ANN] structures and the [AAN] structures, then the model suggests that processing for both types of sentences should be equally difficult, and so reading times should be similar. This was found in this study, providing favourable data for the Constraint-based model.

Of course, the fact that reading times did not differ between conditions could also indicate that instead of no difficulty in any of the conditions, readers experienced difficulty in all of the conditions. Such an interpretation of the results would be more favourable for Gibson (1998)'s Syntactic Prediction Locality Theory. If predictions for the structure of the modified noun-noun compounds are made, then these predictions most likely arise at the adjective, as this is the first element of the phrase where a prediction can be made (other than there being a noun somewhere following the determiner). If the first noun agrees with the predictions made as a result of the adjective, then the second noun could surprise the reader, forcing them to reinterpret the phrase as a compound. If the first noun does not agree with the predictions made as a result of the adjective, then that is the point where the reader has to make new predictions (namely, that another noun likely follows that does agree with the adjective). Both of these cases cause an increase in required processing power, and as a result

an increase in reading times, leaving no difference between the sentence-types. An issue for the Syntactic Prediction Locality Theory, however, might be that there is also no difference between reading times for the target conditions and the controls. What this indicates is that there is no difference in difficulty to interpret the controls compared to the target items. The problem is that if both [A[NN]] and [[AN]N] interpretations cause difficulties for the reader, then so does the [AAN] structure according to the results of the experiment. If the [AAN] structure causes difficulties, it means that this structure requires a reanalysis of the prediction made. This reanalysis could only be triggered by encountering an adjective after the first adjective, where a noun would be expected. If this were the case, then that would mean that incrementally longer sentences with each one more adjective than the last in the noun-phrase, would cause increasingly more difficulty in processing. Since a sentence like (8), repeated in (9), can be read without issues, this seems to be very unlikely, although research is needed to confirm this.

(9) The man pet the friendly playful long-haired blue-eyed drooling dog lovingly.

It is possible to explain away the similar reading times for target and control items due to the fact that the control items are not minimally paired with the target items, as the target items are amongst pairs. As a result, the structure in the controls could be less taxing on the processor by default, but the specific items used could, by chance, require more processing power than the target items do. In order to verify this, a follow-up study is needed where the [AAN] items are also minimally paired with the [ANN] items. However, as it stands, the data do not support the Syntactic Prediction Locality Theory as strongly as they do for the Good-Enough model or the Constraint-based model.

The next model to examine is the Construal model by Frazier and Clifton (1996, 1997). It seems that the data from the experiment conducted in this study do not support Frazier and Clifton (1996, 1997)’s Construal model. What they suggest for their model is that fully ambiguous sentences, as well as sentences with the favoured structure, should show the shortest reading times. However, as could be seen here was that both possible structures elicit similar reading times. In other words, the predictions made by Construal did not hold up, and so the data of this experiment do not support the model.

The final model to look at is Pritchett (1992)’s Generalised Theta Attachment model. Since Pritchett (1992)’s model relies on Theta attachment, there were no real predictions to

be made based on this model[3]. A finding that could have favoured this model would be data that the other models would have a hard time explaining, under the guise that making no prediction would be more accurate than a false prediction. However, as discussed in the paragraphs above, three other models were able to explain the reading times emerging from the experiment (to varying degrees). Since those models were able to make predictions that matched the data found during the experiment, those models are more agreeable frameworks to compare sentence processing with than the Generalised Theta Attachment model.

The next paragraphs discuss what future research in the domain of noun-noun compound processing could look like. The obvious answer would be to see if a similar study could be conducted with more participants. A sample-size of ten participants simply does not hold that much power when it comes to making generalised statements on human behaviour. Having a larger body of samples that provides data on the reading times in these structures allows the statistical analysis to be more accurate, and effects that are judged to be insignificant now could turn out to be significant with more data, as errors can be filtered out more accurately. Similarly, the chance of outliers going unnoticed is smaller if there are more data to which to compare them. Effects that now seem significant might in fact be the result of outliers strongly affecting the data.

In addition to more participants, the experiment could be tweaked in several ways to scrutinise specific models more appropriately. As it stands, the experiment was a very global analysis of reading times for modified noun-noun compounds. However, items could be constructed differently in order to control for certain factors that might influence the outcome according to some models but not others, or experiments could be constructed in such a way so that more of the predictions on noun-noun compounds made by the models can be tested. One such adjustment might be adding items with ambiguous modified noun-noun compounds, rather than only having items that contain disambiguated phrases. It would be key to examine whether the reading times of ambiguous items differ from the items used in this experiment, and in what way. Perhaps these findings might strongly disfavour the models that the current experiment supports. Of course, in such a case, the ambiguous items would likely need to be constructed controlling for word-pair frequency, so that in day-to-day language, the adjective modifies the first noun as often as it modifies the second noun, so that the prediction effects based on word-pair frequency can be accounted for.

That being said, conducting an experiment similar to this one where, instead of only

[3] Assuming adjectives do not assign theta-roles. If they do, then the data do not support the model, as there is no preference for the [[AN]N] structure.

accounting for frequency of use between the adjectives of sentence-pairs, word-pair frequency differences between the two conditions are minimised, might find different results as well. It is possible that in such items, there is a preference for one structure over another. The frequency of use that is accounted for in this study did not impact the reading times, but maybe the adjectives used in the [[AN]N] condition, for example, occured more frequently with their respective nouns than the adjectives in the [A[NN]] condition. Consequently, the reading times for the [[AN]N] items might be shortened, and evidence for an effect obscured. Minimising the effects of this factor could then illuminate on a preferred structure as well. Minimising this factor would be especially important in verifying if the Constraint-based model really makes accurate predictions, as it is one of the conditions that can influence processing, according to MacDonald et al. (1994).

The effect of context would also be necessary to integrate into a study of processing noun-noun compounds. Integrating context would provide useful data for testing both the Constraint-based model, and Ferreira and Lowder (2016)'s model. A context that provides the reader with more information about what is to come, could allow researchers to induce a preference for one interpretation over the other. Examples of such items would be (6), repeated here in (10).

(10) a. I only have one porcelain item and that's my porcelain egg container.
b. I have one container for chicken eggs and one porcelain egg container.

If an effect of context is found that would indicate that context can influence the preference of an interpretation, then this would strongly support the view of MacDonald et al. (1994) and their Constraint-based model.

However, context could also be used differently. It could be used in order to test the model provided by Ferreira and Lowder (2016). Ferreira and Lowder (2016) suggest that items serve different functions for discourse depending on the position of that item in a sentence. Previously established information is positioned in the front of a sentence, and new information is placed later in the sentence. Depending on where an item is positioned, it is then processed by either Good-Enough processing, or by Prediction. In analysing the data of this study, the relevant form of processing in the Ferreira and Lowder (2016) model was assumed to be Good-Enough processing. However, it is also possible that the target phrases in the sentences for this study were processed according to Prediction. By adding context to target sentences, and varying the position of the target phrases, it is possible to test which of

these processing techniques is used by the reader, if either of these is used at all. The importance of context in this set-up is that it ensures that sentences are not processed in a vacuum. This prevents readers from either glossing over every element in a sentence, as there is a discourse that information needs to be integrated into, or it prevents readers from relying on prediction for the entire sentence, simply because all the information that is presented is new.

In order to test the validity of the Syntactic Prediction Locality Theory as a model for modified noun-noun compound processing, it is necessary for the control items to be minimally paired to the target items. This way, any deviations in reading times from the [AAN] items compared to the [ANN] items can more safely be attributed to the structure. Because of the way that the items are constructed in the present study, it is possible that elements in the sentences other than the [AAN] phrase itself increase the reading times of the control sentences. This means that the [AAN] phrases could be easier to process than [ANN] phrases, but other parts of the sentences cause to reading times to level with the reading times of the target sentences. One of these factors might also be frequency of use for the second adjective in the [AAN] items compared to the first noun in the [ANN] items. This would need to be controlled for. If such items are properly constructed, a short reading time for [AAN] sentences compared to [ANN] sentences would work strongly in favour of the Syntactic Prediction Locality Theory. If the [AAN] items in this case display less processing difficulty than the [ANN] items, then that would indicate that no reanalysis is needed for the [AAN] items, and the pitfall of incrementally more lengthy phrases needing incrementally more processing power can be avoided.

What would also be intriguing to study, would be intracranial measurements during an experiment as the one conducted in this study, such as EEG measurements. Looking at the temporal properties of the processing of sentences containing [ANN] structures could provide additional insight into the sentence processing mechanism. What would be particularly informative would be to see if there are any N400 or P600 effects in the processing of modified noun-noun compounds. Presumably, an N400 effect would show in [A[NN]] structures, because here there is a semantic mismatch between the adjective and the first noun. A P600 effect would likely show in [[AN]N] structures, as in this case there is a grammatical mismatch of sorts when a noun follows another noun. However, it is also possible that only one of these effects shows (which could be indicative of a preferred structure), or neither effect shows. Further research looking into the neurolinguistics side of

sentence processing could yield some unique results, and such research would be imperative to understanding sentence processing.

There are other ways in which an experiment looking for processing effects in noun-noun compound interpretation could be set up. One such example would be pragmatic disambiguation, rather than semantic disambiguation. In this study, semantic disambiguation was used in an attempt to force an interpretation in potentially ambiguous sentences. That is to say, the lexical properties of the adjectives were used in order to guide the reader into reading the [ANN] phrase as either an [A[NN]] structure, or as an [[AN]N] structure. However, it is also possible to use pragmatic structure to guide the reader's interpretation. For example, the order in which certain types of adjectives are ranked to modify a noun can be manipulated to force an interpretation. Take for example (11).

(11) a. Jack dropped the beautiful porcelain egg container on the floor.
b. Jack dropped the porcelain beautiful egg container on the floor.

In (11a), the order of the adjectives has no effect on the interpretation. The sentence is ambiguous, as the speaker either refers to *a beautiful porcelain egg*, *a beautiful porcelain container*, or *a beautiful container for porcelain eggs*. However, the order of adjectives is switched around in (11b). Since in English, certain adjectives more comfortably precede others (Malouf, 2000), (11b) has a strong preference for only one interpretation, namely *a porcelain container for beautiful eggs*. This fact could be exploited to study the processing of modified noun-noun compounds further, although any implications that could be gathered from such a manipulation are up for future speculation.

Similarly, it could be relevant to look at the reading times of sentences where ambiguous [ANN] structures contain an adjective that is not semantically related to either noun (for example, *the purple dog breeder*). Once again there is ambiguity, where the speaker might modify either noun with the adjective, but it is difficult for the reader to gather which noun was intended to be modified. It might lead to the reader relying on different processing techniques than those that are employed in normal language processing, or a reader might utilise processing techniques more overtly, because the reader needs to put in more effort to interpret the speaker. What could be even more effective for understanding sentence processing, however, would be to look at intracranial measurements for these [ANN] structures with unrelated adjectives. It would be possible to see if effects such as the N400 effect carry over onto the second noun after it appears to be of the relevant grammatical

category to attach the adjective, but does not contain the appropriate characteristics. Perhaps an N400 effect occurs only for the second noun, or only for the first noun, or for both. Neurolinguistic research looking into this could once again provide new insights into sentence processing, if such research has not yet been conducted. While not very likely, finding an N400 effect for only the second noun, for example, might indicate that there is a neurolinguistic basis for prediction.

Finally, it might be worthwhile to look into modified noun-noun compounds in combination with word category ambiguity. As studied by MacDonald (1993, 1994), compounds can often be subject to word category ambiguity. However, looking at it in combination with an ambiguously modifying adjective, it might be possible to find preferences in interpreting modified compounds. For example, in the phrase *the Bavarian German professor*, *German* can either be a noun referring to the language, or an adjective referring to the heritage of the professor. The categorically unambiguous adjective might affect how the ambiguous second element is interpreted, or instead, that the second element in the phrase affects how the adjective is modifying the phrase. This once again could provide insight into how readers process a given structure, perhaps showing hierarchy in grammatical categories, or simply a preference of one interpretation over another, when two highly different interpretations are available.

Since this phenomenon has been relatively unstudied so far, it is possible to conceive even more follow-up studies. Many other studies could be conducted utilizing the modified noun-noun compound ambiguity, and hopefully the above provided some interesting suggestions that will lead into further research of the modified noun-noun compound phenomenon.

6. Conclusion

In sum, a self-paced reading task that involved minimally paired sentences containing either an [A[NN]] structure or an [[AN]N] structure showed no significant difference in reading times between these two groups. In addition, reading times of these sentences were compared to reading times of sentences containing [AAN] structures, which also showed no significant difference. This led to the conclusion that there is no difference in how taxing any of the constructions is on the processer. The statistical analysis showed an effect on reading times between British and non-British subjects, as well as an effect of bilingualism between subjects, but these effects did not tell us anything on the processing of these items, as there was no interaction-effect with the used structure. That being said, the sample size was rather

small (n=10), which means that any generalised statements made as a result of this study need to be studied further in order to hold any weight.

After the experiment, the findings of this study were compared to the outcomes that existing models would predict. Out of this comparison, it seemed that Ferreira and Lowder (2016)'s model and MacDonald et al. (1994)'s model seemed to make the most accurate predictions on the reading times of the [ANN] and [AAN] items. Gibson (1998)'s model also made accurate predictions with regards to the [ANN] reading times, but the fact that there was no difference between the reading times of the [ANN] sentences and the [AAN] sentences could have certain implications that should preferably be avoided. The data gathered in this study did not support Frazier and Clifton (1997)'s Construal model, or Pritchett (1992)'s Generalised Theta Attachment model as well as they supported the other three models, as these models made no (strong) predictions on the absence of an effect of sentence-type. Finally, this paper discussed some ideas for future research, for instance alternative set-ups that could further investigate the validity of some of the models discussed. Most interesting, however, would be research studying intracranial measurements of modified noun-noun compound interpretation, in order to see what neurolinguistic effects (if any) are brought about by such constructions (e.g. N400 or P600 effects). All in all, much more research is needed on the topic of modified noun-noun compound processing in order to see how existing models hold up when accounting for this form of ambiguity. This has only been a first step in fully understanding the phenomenon, which has hopefully been able to provide a stepping-stone for future research into this so-far neglected aspect of sentence processing.

Acknowledgements

The author of this dissertation would like to thank a number of people. First of all, I would like to thank my supervisor, Dr Olaf Koeneman, for his guidance and insight throughout the writing of this dissertation. Additionally, I would to express my gratitude to the second reviewer, Prof Dr Ans van Kemenade. Furthermore, I would like to express a great amount of appreciation towards the Max Planck Institute of Psycholinguistics, especially, Prof Dr Caroline Rowland, Patricia Manko, Nienke Rulkens-Dijkstra, Maarten van den Heuvel, and Nanjo Bogdanovicz for facilitating my experiment. I would also like to thank Dr Anouschka Foltz for inspiring me to research this topic. Finally, I would like to thank Sophie Fennema, for some last-minute proofreading.

References

Cinque, G. (1990). Ergative Adjectives and the Lexicalist Hypothesis. *Natural Language & Linguistic Theory*, 8(1), 1-39.

Cuetos, F., & Mitchell, D. C. (1988). Cross-Linguistic Differences in Parsing: Restrictions on the use of the Late Closure Strategy in Spanish. *Cognition*, 30(1), 73-105.

Ferreira, F., Bailey, K. G. D., Ferraro, V. (2002). Good-Enough Representations in Language Comprehension. *Current Directions in Psychological Science*, 11(1), 11-15.

Ferreira, F., & Lowder, M. W. (2016). Prediction, Information Structure, and Good-Enough Language Processing. *Psychology of Learning and Motivation*, 65, 271-247

Foltz, A. (2016). Lecture 1 [In-class experiment]. Retrieved from Bangor University, Models in Sentence Processing. Blackboard: https://blackboard.bangor.ac.uk/

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance* (559-586). Hiltsdale, NJ: Ertbauin.

Frazier, L., & Clifton, C. Jr. (1996). *Construal*. Cambridge, MA: MIT P.

Frazier, L., & Clifton, C. Jr. (1997). Construal: Overview, Motivation and Some New Evidence. *Journal of Psycholinguistic Research*, 26(3), 277-295.

Gibson, E. (1998). Linguistic Complexity: Locality of Syntactic Dependencies. *Cognition*, *68*(1), 1-76.

MacDonald, M. C. (1993). The Interaction of Lexical and Syntactic Ambiguity. *Journal of Memory and Language*, 32, 692-715.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9(2), 157-201.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review*, 101(4), 676-703.

Malouf, R. (2000). The order of prenominal adjectives in natural language generation. *ACL '00 Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, 85-92.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: U. of Chicago P.

The Center for Reading Research. (2013). *Subtlex-UK* [Data file]. Retrieved from http://crr.ugent.be/archives/1423

Appendix A

[A[NN]]-sentences	**[[AN]N]-sentences**	**$f_{[A[NN]]}$**	**$f_{[[AN]N]}$**	**f-ratio**
Jacob threw the simple earring packaging in the trash.	Jacob threw the silver earring packaging in the trash.	23743	20779	1.142644016
Mary gave the kindly artwork supplier a nice ring.	Mary gave the cardboard artwork supplier a nice ring.	1331	1362	0.9772393539
Lisa heard the savvy record collector from next door.	Lisa heard the vinyl record collector from next door.	376	373	1.008042895
Michael got the smartest backpack warranty at the shop.	Michael got the nylon backpack warranty at the shop.	250	253	0.9881422925
People thought the handsome clothing retailer to be nice.	People thought the cotton clothing retailer to be nice.	3112	2544	1.22327044
Gavin pays the lively comic library handsomely.	Gavin pays the printed comic library handsomely.	1812	1682	1.077288942
Lindsey thought the measured content quality was quite high.	Lindsey thought the toxic content quality was quite high.	1454	1540	0.9441558442
Aaron planned the unfair beauty election to a T.	Aaron planned the naked beauty election to a T.	3736	3516	1.062571104
Students meet the depressed career counsellor all the time.	Students meet the naval career counsellor all the time.	1575	1633	0.9644825475
Clients loved the thankful software producer from the start.	Clients loved the buggy software producer from the start.	512	522	0.9808429119
Geoffrey told the eager sofa designer to help him.	Geoffrey told the bouncy sofa designer to help him.	1537	1485	1.035016835
Ryan asked the cautious swimming instructor for advice	Ryan asked the outdoor swimming instructor for advice	1814	1789	1.013974287
Danny did the ghastly physics exercise in a day.	Danny did the thermal physics exercise in a day.	594	593	1.001686341
Countries asked the brilliant Baroque orchestra to play there.	Countries asked the early Baroque orchestra to play there.	43705	42423	1.030219456
Abby gave the cheery flower consultant a warm hug.	Abby gave the bridal flower consultant a warm hug.	330	333	0.990990991
Artists sued the savvy music editor for his theft.	Artists sued the choral music editor for his theft.	376	367	1.024523161
Police did a routine chamber inspection in our street.	Police did an upper chamber inspection in our street.	4954	3811	1.299921281
Sarah found the concrete movie theatre on her trip.	Sarah found the silent movie theatre on her trip.	3332	3553	0.9377990431

Appendix B

Group 1	Group 2
A beard is a great way to look more mature.	A beard is a great way to look more mature.
Larry spent all night working.	Larry spent all night working.
The mother cow believed the calf was hers.	The mother cow believed the calf was hers.
Sarah found the silent movie theatre on her trip.	Sarah found the concrete movie theatre on her trip.
Matt has a new cat.	Matt has a new cat.
There is a cute guy working at the new coffee shop.	There is a cute guy working at the new coffee shop.
Lindsey thought the toxic content quality was quite high.	Lindsey thought the measured content quality was quite high.
Dave is currently petitioning for a visa.	Dave is currently petitioning for a visa.
The nurse took the candy from the box that is by the window.	The nurse took the candy from the box that is by the window.
Gregg doesn't read the news anymore.	Gregg doesn't read the news anymore.
Pianos are the perfect extra instrument for this song.	Pianos are the perfect extra instrument for this song.
I think I will get some sushi tonight.	I think I will get some sushi tonight.
Jerry is going to need some help with getting a promotion.	Jerry is going to need some help with getting a promotion.
Police did a routine chamber inspection in our street.	Police did an upper chamber inspection in our street.
The student stared at the assistant of the teacher who was at the school.	The student stared at the assistant of the teacher who was at the school.
Rick was mumbling to himself during his break.	Rick was mumbling to himself during his break.
The patient has not spelt in days.	The patient has not spelt in days.
Patrick feared the heavy polished statuette hit the ground.	Patrick feared the heavy polished statuette hit the ground.
David read the sign in the library.	David read the sign in the library.
The girl teased her brother with the lollipop	The girl teased her brother with the lollipop
Gaby saw the shady teenaged hooligans by the park.	Gaby saw the shady teenaged hooligans by the park.
The man gave his girlfriend a ring.	The man gave his girlfriend a ring.
Social skills are something you work on.	Social skills are something you work on.
Mary interviewed the chauffeur of the actress who had an accident.	Mary interviewed the chauffeur of the actress who had an accident.
Mary gave the kindly artwork supplier a nice ring.	Mary gave the cardboard artwork supplier a nice ring.
The dog chased the cat with its owner.	The dog chased the cat with its owner.
Children are learning to stand up for themselves.	Children are learning to stand up for themselves.
People thought the handsome clothing retailer to be nice.	People thought the cotton clothing retailer to be nice.
Politics greatly interested George.	Politics greatly interested George.
The lady was observing the toy of the baby that was on the bed.	The lady was observing the toy of the baby that was on the bed.
The shepherd took the sheep to the field that was by the road.	The shepherd took the sheep to the field that was by the road.
Barry showed the funny online video to his friends.	Barry showed the funny online video to his friends.
Rachel's son has picked up smoking.	Rachel's son has picked up smoking.
Grandpa has started using Snapchat.	Grandpa has started using Snapchat.
Clients loved the buggy software producer from the start.	Clients loved the thankful software producer from the start.

Somebody shot the servant of the actress who was on the balcony.	Somebody shot the servant of the actress who was on the balcony.
The bus took an unexpected left turn.	The bus took an unexpected left turn.
Jack met the friend of the teacher who was in Germany.	Jack met the friend of the teacher who was in Germany.
Police searched the empty rural residence for more clues.	Police searched the empty rural residence for more clues.
Ryan found Gavin together with Michael.	Ryan found Gavin together with Michael.
Bruce was not trying to scare anyone.	Bruce was not trying to scare anyone.
George was holding the toy of the dog that was by the window	George was holding the toy of the dog that was by the window
Jacob threw the simple earring packaging in the trash.	Jacob threw the silver earring packaging in the trash.
The soldiers had a refreshing drink at the spring.	The soldiers had a refreshing drink at the spring.
The pig's house was made of bricks.	The pig's house was made of bricks.
Michael got the nylon backpack warranty at the shop.	Michael got the smartest backpack warranty at the shop.
She is from a different time.	She is from a different time.
Ali had made an appointment with the doctor.	Ali had made an appointment with the doctor.
Aaron planned the unfair beauty election to a T.	Aaron planned the naked beauty election to a T.
Kevin had to go clean his room.	Kevin had to go clean his room.
The good Samaritan found the missing child.	The good Samaritan found the missing child.
The child was crying in the back of the BMW.	The child was crying in the back of the BMW.
Hikers use special tempered equipment on their hikes.	Hikers use special tempered equipment on their hikes.
The number of black widows is on the rise.	The number of black widows is on the rise.
The student approached the secretary of the manager who was at the mall.	The student approached the secretary of the manager who was at the mall.
Danny did the thermal physics exercise in a day.	Danny did the ghastly physics exercise in a day.
Herbert sometimes spends an entire day in bed.	Herbert sometimes spends an entire day in bed.
Olaf decided to turn down the music.	Olaf decided to turn down the music.
The man thought about the news in the bar.	The man thought about the news in the bar.
Artists sued the savvy music editor for his theft.	Artists sued the choral music editor for his theft.
Peter was looking at the book from the shelf that is by the desk.	Peter was looking at the book from the shelf that is by the desk.
The rabbit chased Corrie around the room.	The rabbit chased Corrie around the room.
David heard the gorgeous annual performance from his house.	David heard the gorgeous annual performance from his house.
Frank's rolls taste amazing.	Frank's rolls taste amazing.
The green car is parked further down the street.	The green car is parked further down the street.
Ella spent all day playing with the child.	Ella spent all day playing with the child.
Lisa heard the vinyl record collector from next door.	Lisa heard the savvy record collector from next door.
William is not a smooth talker.	William is not a smooth talker.
Snakes are excellent swimmers.	Snakes are excellent swimmers.
Alfred drinks bubbly sugared beverages on the job.	Alfred drinks bubbly sugared beverages on the job.
Robert threw the match to the ground.	Robert threw the match to the ground.
I have added some examples.	I have added some examples.
My daughter wanted me to bring some incense.	My daughter wanted me to bring some incense.

Geoffrey told the eager sofa designer to help him.	Geoffrey told the bouncy sofa designer to help him.
In order to help the environment, we planted trees.	In order to help the environment, we planted trees.
Tim has slept like a log on his new matrass.	Tim has slept like a log on his new matrass.
Ryan asked the outdoor swimming instructor for advise	Ryan asked the cautious swimming instructor for advise
John saw the man with the binoculars.	John saw the man with the binoculars.
Blood donations save people every day.	Blood donations save people every day.
If you want to impress people, you need to put in more effort.	If you want to impress people, you need to put in more effort.
Students meet the naval career counsellor all the time.	Students meet the depressed career counsellor all the time.
The noise started to agitate Sean.	The noise started to agitate Sean.
The intelligent student had earned her degree.	The intelligent student had earned her degree.
Letters from the bankrupt legal company must be tossed.	Letters from the bankrupt legal company must be tossed.
Susan dried the child with the towel.	Susan dried the child with the towel.
He found a turtle behind the barn.	He found a turtle behind the barn.
The man ran over the child with the bicycle.	The man ran over the child with the bicycle.
Abby gave the bridal flower consultant a warm hug.	Abby gave the cheery flower consultant a warm hug.
Dad put the coffee on the table.	Dad put the coffee on the table.
The black television cost two thousand pound.	The black television cost two thousand pound.
Countries asked the brilliant Baroque orchestra to play there.	Countries asked the early Baroque orchestra to play there.
The company did not feel I was qualified for the position.	The company did not feel I was qualified for the position.
Rob has always been a great conductor.	Rob has always been a great conductor.
The guest proposed to the girl with the ring.	The guest proposed to the girl with the ring.
Gavin pays the lively comic library handsomely.	Gavin pays the printed comic library handsomely.
The cat looks at the bird in the garden.	The cat looks at the bird in the garden.
The notches on the wall show the height of the kids.	The notches on the wall show the height of the kids.
Receipts by the clumsy senior employee are all wrong.	Receipts by the clumsy senior employee are all wrong.
Make sure to eat fruit twice per day.	Make sure to eat fruit twice per day.
The dogs do not bark at the mailman.	The dogs do not bark at the mailman.