Complement coercion in Polish and the role of selectional restrictions revealed in a self-paced reading study[1]

Alexandra Anna SPALEK — *Universitetet i Oslo*
Barbara Maria TOMASZEWICZ — *Universität zu Köln, Universytet Wrocławski*

Abstract So-called 'complement coercion' (*'begin a book'*), understood as a combinatorial conflict, is mainly analysed as a repair operation in composition. Experimental data has shown that there is an extra cognitive effort in the processing of event-selecting verbs with entity-denoting arguments. These results support the formal analysis of 'complement coercion' as an enriched form of semantic composition (Pustejovsky, 1995; Egg, 2003; de Swart, 2011; Asher, 2015). Recently, an alternative view has been proposed by Piñango and Deo (2015) arguing that there is no mismatch between the verb and its complement, but an ambiguity resulting from the different dimensions along which an aspectual verb in composition with its argument can be interpreted (e.g. temporal, spatial, etc.). This approach has been supported by experiments showing that aspectual verbs like *begin* incur greater processing cost in coercing contexts than psychological verbs like *enjoy* (Lai et al., 2014). We designed a self-paced reading experiment to compare the different predictions that the two approaches make for the processing of matching and mismatching verb-noun combinations. Our results are compatible with enriched composition, but not with dimensional ambiguity. We find facilitation in conditions where the selectional restrictions of the verbs are satisfied (*'begin a fight', 'see a book'*), and longer processing times in the conditions *'begin a book', 'see a fight'*, which does not support the dimensional ambiguity approach. Our experiment thus provides evidence that selectional restrictions are a fundamental property of a predicate and that they need to be understood as a graded continuum of combinatorial preferences, as also argued in Spalek (2014).

Keywords: Complement coercion, aspectual verbs, real-time processing, self-paced reading.

1. Introduction

The minimal pair in (1) presents a challenge for the compositional derivation of meaning: in (1a) the verb *start* combines with an event-denoting complement, in (1b) with an entity-denoting complement. In the first case, the event is explicitly referred to, in the second, we can only infer that an event involving the soup was started. Cases like (1b) exhibit so-called 'complement coercion' – an operation by which the verb 'coerces' its semantically mismatching object to the appropriate semantic type.

(1) a. The boy started the fight.
b. The boy started the soup.

[1]We would like to thank Joanna Błaszczak, Elsi Kaiser, Roumi Pancheva, Kjell Johan Sæbø, Petra Schumacher and the audience at the ESSLLI 2016 workshop (Referential Semantics One Step Further) for inspiring comments and discussion. We are also grateful to Joanna Błaszczak and Marcin Orszulak for assistance with conducting the experiment. The work on the experiment was supported by the FOCUS grant from the Foundation for Polish Science.

This contrast imposes the following consideration to be accounted for theoretically: given that the verb *start* presupposes an argument of eventuality type, predication in example (1b) should go wrong, because it seems that there is a conflict between the demands of the predicate on the type of its argument and the argument the predicate actually gets. Yet the sentence (1b) is interpretable and can refer to an event of soup eating or soup cooking, etc. Any compositional theory of meaning needs to account for the means by which a predicate like *start* can compose both with an eventuality-denoting noun, e.g. *fight* in (1a), and with an individual-denoting noun, e.g. *soup* in (1b). 'Complement coercion' thus raises a very general question of conceptual knowledge and how it is supposed to interact with referential semantics.

Crucially for our study, the verb *start* in the minimal pair in (1) contrasts with verbs such as *see, criticize* or *prepare*, which prima facie combine easily with both kinds of noun phrases, event nouns and individual-denoting nouns. Many studies classify verbs like *start* as 'aspectual verbs' and show that they typically allow for 'complement coercion'. The verbs in (2), in contrast, are classified as non-aspectual verbs, but not much has been said about their selectional restrictions so far.

(2) a. The man saw/ criticized/ prepared the fight.
b. The man saw/ criticized/ prepared the soup.

The present study explores the combinatorial compatibility of the two kinds of verbs, aspectual verbs (1) versus non-aspectual verbs (2), with eventuality-denoting and individual-denoting arguments. Comparing the ease of processing of the different verb-noun combinations, we obtained contrasts indicating the complexity of semantic processing during real-time comprehension, which we argue have concrete implications for the models of meaning based on the idea that predicates select their arguments. We designed two self-paced reading studies that contrast aspectual verbs with non-aspectual verbs (henceforth, 'AV(s)' and 'N-AV(s)') using Polish. Polish seemed particularly interesting for our purpose, because it has two kinds of entity-denoting nouns that differ in their morphology, one of them being morphologically non-transparent and another containing a verbal root. The latter allowed us to manipulate the 'coercion effect' in one of the experiments.

Before advancing, we introduce a terminological clarification. Since the term 'complement coercion' is widely accepted in the literature on AVs to refer to the combinations where an AV takes an individual-denoting DP, we will refer to these combinations as 'coercing contexts' and to the group of verbs as simply 'coercing verbs', without correlating these terms with any particular theoretical account.

2. On complement coercion

Coercing verbs, like *enjoy, begin, start, finish, stop*, etc., have been extensively discussed since Pustejovsky (1991), and their capacity to take an individual-denoting noun as a complement, so-called 'complement coercion' (1b), has been the object of a rich theoretical discussion (Pustejovsky, 1995; Copestake and Briscoe, 1995; Fodor and Lepore, 1998; Egg, 2003; de Swart, 2011; Asher, 2011). The common denominator of these well-known accounts has been

to take complement coercion as a semantic type mismatch between the selectional requirements of the predicate and its argument. Diverse solutions have been proposed as to how the missing specification of the event is recovered (e.g. in (1b) it could be either an eating or a cooking event depending on the context). Pustejovsky (1995) posits a complex lexical entry for the argument which is able to give access to the event reading. Fodor and Lepore (1998), on the other hand, propose that the event-reading results from post-lexical inferences, while many others advocate for some kind of enriched form of composition (Egg, 2003; de Swart, 2011; Asher, 2011). In some cases, the enriched process of composition comes together with the assumption that the lexicon is underspecified, as in Egg (2003); in other cases, certain lexical units, such as AVs, license dependent types that trigger a repair process during composition (Asher, 2011).

Abstracting away from the formal details, these approaches share the assumption that AVs determine the semantic type of their arguments and that to satisfy those semantic requirements some operation needs to be performed to 'coerce' the entity-denoting complement into the appropriate semantic type. We will refer to this as the Enriched Semantic Composition hypothesis following the terminology proposed in lexical-semantic research (Pustejovsky, 1991, 1995; Jackendoff, 1997). This hypothesis finds support in various kinds of experimental results (McElree et al., 2001; Traxler et al., 2002, 2005; McElree et al., 2006; Pylkkänen and McElree, 2007; Frisson and McElree, 2008; Kuperberg et al., 2010), which have shown that combining a coercing verb with an entity-denoting complement (1b) during real-time comprehension engenders more processing cost than combining it with an event-denoting complement (1a).

More recently, taking a special focus on entity-denoting complements of AVs, Piñango and Deo (2015) called the type-mismatch analysis into question by observing that there is a whole range of combinations of AVs and individual-denoting nouns that represent configurational relations. In combinations such as in (3), Piñango and Deo (2015) observe that there is no need for coercion and propose that these cases are prevalent, rather than exceptions. This observation leads them to develop a generalized lexical semantics for AVs that dispenses with coercion.

(3) This is the famous perch that officially begins the Appalachian Trail.
((7d) in Piñango and Deo (2015))

Based on the idea of generalized paths (Gawron, 2009), Piñango and Deo (2015) propose a unifying lexical semantics for AVs, where AVs specify relations between the sub-parts of the axis determined by their complements' denotation. On this account AVs presuppose their arguments to be structured individuals with respect to a contextually determined function in some domain. AVs thus make reference to parthood relations formalized as contextually defined functions, e.g. a spatial trace function in the case of example (3). With this analysis Piñango and Deo (2015) argue that what has been called 'complement coercion' can be explained as a case of ambiguity between the different dimensions that a particular AV can access. An interesting example to illustrate this account in more detail is (4).

(4) The little girl began the queue. ((29a) in Piñango and Deo (2015))

According to Piñango and Deo (2015), the interpretation of a sentence containing an AV, such as *begin* in (4), is dependent on determining the specific dimension along which the denotation of the complement is structured. (4) is thus ambiguous and can be paraphrased either as 'The little girl began forming the queue' (temporal dimension) or as describing the position of the little girl relative to the structure of the queue (spatial dimension). Examples of this kind lead Piñango and Deo (2015: 14) to conclude that 'any analysis of aspectual verbs that assumes that they select for event-denoting complements is not tenable'.

This approach crucially assumes that AVs do not impose restrictions on the type of their arguments, but rather establish 'parthood relations between objects along a range of familiar (and, sometimes not so familiar) dimensions' (Piñango and Deo, 2015: 10). The processing cost in experimental studies can, therefore, be attributed to the need of identifying the appropriate dimension for the interpretation of an AV and its complement. This proposal finds support in studies showing that only a subset of coercing verbs engender additional processing cost, namely strictly AVs, such as *begin, start*, etc., but not psychological coercing verbs, such as *enjoy, prefer* (Katsika et al. (2012), Lai et al. (2014)). We will refer to this approach as the Dimension Ambiguity Hypothesis.

The Enriched Composition approach predicts that during real-time sentence processing AVs create strong expectations for the semantic type of their arguments. The Dimensional Ambiguity hypothesis, on the other hand, makes no assumptions about selectional restrictions and predicts that the ease of the identification of a temporal dimension for the interpretation of the entity-denoting complement will affect the ease of the processing of coercion contexts. In the absence of ambiguity between dimensions (with N-AVs and their complements), no processing costs should be observed. We measured the processing cost in terms of reading times, and our experiment was designed on the basis of prior reading studies on complement coercion. In the next section we present prior experimental results that guided the design of our study in order to test the two different approaches to AVs and their selectional properties.

3. Differences in reading times

The processing cost of complement coercion, (1b) in contrast to (1a), was first shown in the self-paced reading and eye-tracking during reading experiments of McElree et al. (2001) and Traxler et al. (2002). In a self-paced reading experiment participants read the sentences presented on the computer screen chunk-by-chunk, advancing at their own pace (Just et al., 1982). The times of each button press to move to the next chunk are recorded, and longer reaction times (RTs) are interpreted as reflecting a higher level of processing difficulty. Using self-paced reading, McElree et al. (2001) tested the basic question whether AVs, like *start* in (5a), are more difficult to process than N-AVs, like *write, read* in (5b)-(5c). They compared the reading times at the noun complement and at the following adverbial in three conditions: coercion context (5a), preferred combination (5b) (the verb explicitly expressed the event) and non-preferred combination (5c) (the verb explicitly expressed a less frequent but plausible event).[2]

[2]Ratings for the preferred and non-preferred verb-noun combinations were obtained in a separate norming study.

(5) a. The author was starting the book in his house. – Coercion
b. The author was writing the book in his house. – Preferred
c. The author was reading the book in his house. – Non-preferred

They found that while the non-preferred verb-noun combinations elicited longer RTs than the preferred ones, there was an extra cost associated with the coercion contexts. (The reading times were measured both at the target (noun) and post-target (adverbial) regions, because the processing of the target region affects the processing of the following regions. The RTs at the target noun were significantly longer in the coerced and the non-preferred conditions than in the preferred condition. The RTs at the adverbial were significantly longer in the coerced condition than in both the preferred and the non-preferred conditions.)

Traxler et al. (2002) further tested the nature of the processing delay using self-paced reading and eye-tracking during reading methodologies. While in a self-paced reading experiment segments must be presented sequentially, in an eye-tracking experiment natural reading is possible and different measures of the processing cost are available (e.g. total reading times for a region, just like in self-paced reading, but also the time spent reading a word for the first time, re-reading it again, etc.). They investigated whether the effects found by McElree et al. (2001) could be caused by the fact that while *start* has a preference for a verbal complement, it receives a nominal complement in (5a). Their test items involved quadruples of sentences such as (6) and (7). The pair in (6) contains an AV combined with a matching event-denoting noun (6a) and with an entity-denoting noun that requires coercion (6b). In pair (7) a N-AV is combined with an entity-denoting noun (7a) and with an event-denoting noun (7b).

(6) a. The boy started the fight after school today. – No Coercion (AV, EventN)
b. The boy started the puzzle after school today. – Coercion (AV, EntityN)

(7) a. The boy saw the puzzle after school today. – No Coercion (N-AV, EntityN)
b. The boy saw the fight after school today. – No Coercion (N-AV, EventN)

The results of both experiments, the eye-tracking study and the self-paced reading experiment, indicated that the complement coercion condition, (6b), incurred extra processing cost. The self-paced reading data showed no significant differences at the target noun region like in the McElree et al. (2001) study. At the adverbial region, there was a main effect of NP-Type (EntityNs received longer RTs than EventNs) and a significant interaction: sentences with AVs and EntityNs, (6b), generated longer RTs than sentences with N-AVs and EntityNs, (7a). This means that the RTs at the adverbial were the slowest in the AV+EntityN condition, (6b). Of interest to our study, as will be explained in section 4, is also the result that sentences with AVs and EventNs, (6a), had numerically (not significantly) shorter RTs than sentences with N-AVs and EventNs, (7b).

The eye-tracking data revealed reading differences already at the noun region, with the coercion condition (6b) being the hardest. Crucially, the measures that are typically considered to reflect

integrative processing in eye-movement data ('second pass time'[3] and 'total time') revealed that at the noun region EntityNs were harder to process with AVs – there was a main effect of NP-Type and a significant interaction. Second pass time and total time are later measures which can explain the lack of effects on the noun in their self-paced reading study. Moreover, at the noun, in the second pass time data, the RTs in the AV+EventN condition, (6a), were numerically lower than in the N-AV+EventN condition, (7b) – though this difference was not significant, just like in the self-paced reading experiment, we take this finding into account for the design of our study, as explained below (section 4).

The Traxler et al. (2002) study provided evidence that AVs with EntityNs are costly to process. It did not yield statistically significant results indicating that AVs create an expectation for EventNs during processing, but in both of their experiments there was clearly such a trend (EventNs received faster RTs after AVs than N-AVs). The goal of our study was to confirm that coercion contexts require extra processing cost in Polish, and to find out what causes this extra cost: (i) the verb and its selectional restrictions (Enriched Semantic Composition) or (ii) the noun interpreted as a structured individual (Dimension Ambiguity). We based our experimental design on Traxler et al. (2002), but structured the whole experiment in such a way that we could take advantage of two different classes of nouns in Polish that allowed us to manipulate both coercion contexts and the properties of the complement nouns, as presented in the next section.

4. The present study

We designed a self-paced reading experiment to compare the different predictions that the Enriched Composition approach and the Dimension Ambiguity approach make for the processing of matching and mismatching verb-noun combinations. The two approaches make different predictions for (i) the processing AV+EventN combinations in contrast to N-AV+EventN, and for (ii) the processing of those EntityNs that are morphologically non-transparent and those that contain a verbal root.

The first contrast follows from the fact that, as discussed in section 2, the Enriched Composition approach, but not the Dimension Ambiguity approach, assumes that AVs select for event-denoting complements. This predicts that during incremental processing, when the parser encounters an AV, an expectation for an event-denoting complement is created. In contrast, N-AVs, which can select for both entity- and event-denoting arguments, should not create such an expectation during parsing. Accordingly, AVs but not N-AVs should display a bias for EventNs in processing. The Dimension Ambiguity, on the other hand, predicts the absence of such a bias, because AV+EventN combinations are simply unambiguous, just like N-AV+Entity/EventN combinations. As noted above in section 3, in the two experiments in Traxler et al. (2002) there was a trend towards AVs facilitating EventNs (not statistically significant), therefore we hypothesized that we could increase the sensitivity of the parser to the contrast between AVs and N-AVs by increasing our participants' exposure to coercing contexts.

[3]Second pass time includes all of the time spent in the region following first-pass fixations and the time at second access after exiting the region to the left or right

The second set of predictions where the two theories diverge follows from the assumption of the Dimension Ambiguity approach that EntityNs as complements to AVs are interpreted as structured individuals. During processing, the parser must choose the contextually relevant dimension along which the individual is structured, so when several dimensions are available the parser must resolve the ambiguity. We know from reading experiments on lexical ambiguity that 'balanced' and 'biased' ambiguous words are processed differently (Rayner and Duffy, 1986). Biased ambiguities are those where the two meanings are asymmetric in their likelihood, such that one meaning is dominant and the other subordinate (e.g. *ball* 'a spherical object', 'a social gathering for dancing'), though the context can reverse which of the meanings is the dominant one. In balanced ambiguities both meanings are equally available. Rayner and Duffy (1986) found that within a sentence balanced ambiguous targets were read slower than biased ambiguous targets, but in the post-target region there was an additional cost with biased ambiguous words. Thus, the two types of ambiguities received clearly different processing profiles. Accordingly, we hypothesized that we should find such different processing profiles with EntityNs biasing a temporal dimension and with EntityNs that are balanced. Polish provided us with two types of EntityNs that differed in the likelihood of a temporal interpretation: morphologically simple nouns denoting physical objects and morphologically complex nouns containing a verbal root.

The two sets of predictions were tested as two sub-experiments of one study. This combined presentation allowed us to expose our participants to a wide variety of coercion contexts, which, as noted above, was hypothesized to increase the parser's sensitivity to the different selectional requirements of AVs and N-AVs. In effect, 25% of all the sentences in the whole experiment involved coercion contexts. We discuss the predictions, materials and results for each of the sub-experiments in turn.

4.1. Sub-Experiment 1

The goal of the experiment was two-fold. First, we wanted to replicate the results of Traxler et al. (2002) by finding evidence for the processing cost of AV+EntityN combinations in Polish. Second, we wanted to test whether AV+EventN combinations facilitated processing relative to N-AV+EventN combinations, as predicted by the Enriched Composition approach. If AVs create an expectation for an eventuality-denoting complement, AV+EventNs should be easier to process than N-AVs which do not create expectations for one specific type of complement. Accordingly, we predict a three-way distinction in the processing cost:

(8) AV+EntityN > N-AV+EntityN/EventN > AV+EventN
'begin book' > 'see book/fight' > 'begin fight'
type mismatch > no specific requirement > requirement satisfied

The Dimensional Ambiguity approach, on the other hand, predicts a two-way contrast – between the AVs+EntityN combination, which requires the resolution of dimension ambiguity, and the other three combinations, where there is no ambiguity to resolve:

(9) AV+EntityN > N-AV+EntityN/EventN, AV+EventN
'begin book' > 'see book/fight', 'begin fight'
dimension ambiguity > no ambiguity

Since an AV merely requires that its argument is a structured individual with respect to a contextually determined dimension, in the combination AV+EventN, the noun is unambiguously interpreted along the temporal dimension. Therefore, the Dimensional Ambiguity analysis does not predict differences in the processing of AVs with EventNs and N-AVs with either kind of complement.

4.1.1. Materials and procedure

We used the Polish adaptations of the 24 items from Traxler et al. (2002) in the same four conditions as in (6) and (7): AV+EntityN, AV+EventN, N-AV+EntityN, N-AV+EventN. We used the AVs listed in (10)[4] and the N-AVs listed in (11).

(10) **Aspectual verbs** selecting event denoting complements:
zacząć (begin)×4; *rozpocząć* (begin)×4; *skończyć* (finish)×4; *ukończyć* (finish)×2; *zakończyć* (finish)×2; *przerwać* (pause)×2; *wytrzymać* (endure)×2; *oczekiwać* (await) ×2

(11) **Non-aspectual verbs** taking both entity- and event-denoting complements:
zobaczyć (see)×4; *skrytykować* (criticize)×2; *przygotować* (prepare)×2; *pochwalić* (praise)×4; *zignorować* (ignore)×2; *obejrzeć* (watch)×2; *opisać* (describe)×4; *wspomnieć* (mention)×2

In addition to the 24 experimental items, the participants (*n*=36, all native speakers of Polish, students from the University of Wrocław) saw the 24 items from Sub-Experiment 2 and 24 unrelated fillers containing the comparative construction (e.g. *'The daughter downloaded more games on the home computer than the son.'*)

The self-paced reading experiment was conducted using the Linger program (Rohde (2001), `http://tedlab.mit.edu/~dr/Linger/`) in a moving-window paradigm. The whole sentence was initially presented with all the words masked by dashes. Participants pressed the space bar to reveal a phrase (consisting either of a single word, a compound word or a preposition and a word), and when the new phrase appeared, the previous phrase was masked again (non-cumulative presentation). Half of the sentences in the experiment were followed by a 'yes/no' question testing comprehension. The Linger program distributed the items (from both sub-experiments) and the fillers into four lists and randomized the order of presentation of sentences within each list for each participant. Each list contained 72 test and filler sentences. The experiment took about 30 min to complete.

[4]Following Traxler et al. (2002) we included the verb *forget*, but after the experiment was conducted, we realized that it can create a coercion context. When you forget an article, you either forget what it was about or you forget the action of taking/bringing it. As a result, we did not include the two items with *forget* in our statistical analysis.

4.1.2. Data analysis

Data from participants with at least 70% accuracy on the comprehension questions, (n=35), was analyzed. None of the participants' mean reading time was more than 2.5 standard deviations from the participant mean. For reading time data analysis, we used the two-step model of Jaeger et al. (2008), Hofmeister (2011); Hofmeister et al. (2013), where first residual reading times are obtained and then those are used for the statistical analyses.[5] We removed the outlier RTs below 200ms and above 4000ms, which represented 0.33% of the experimental and filler data, and used the Box-Cox procedure (Box and Cox, 1964) to determine that the reading times should be log-transformed to meet the assumption of the linear model that residuals be normally distributed (see Baayen and Milin (2010) and Vasishth et al. (2013) a.o. for arguments that the normality assumption is important for reaction time data, whose distribution is positively skewed). For each subject we computed residual reading times to account for the differences in: (a) kind of stimulus (Sub-Experiment 1 and 2, fillers), (b) word length (because there is no linear relationship between the number of characters in a word and the required reading time), (c) word position in a sentence (because it also has no linear effect on reading times) and (d) the log-transformed trial number (because readers speed up as the experiment progresses). The residual reading times correct for these individual differences between participants' reading speeds (e.g. Ferreira and Clifton (1986), Trueswell and Tanenhaus (1994)). Statistical analyses were carried out over the residual reading times without further trimming using linear mixed-effects models in R (version 3.3.2; R Core Team (2016)) with the lme4 package (version 1.1-12; Bates et al. (2015)). We report the results of linear mixed effects models with a fully specified random effects structures (including random intercepts and for all fixed effects by participants and by items, Barr et al. (2013)); in case of convergence failures the random effects structure was simplified following Baayen et al. (2008). The p-values were obtained using the Satterthwaite approximation implemented in the lmerTest package (Kuznetsova et al., 2016).

4.1.3. Results

We found no differences in reading times at the subject and verb regions (see the plot in Figure 1). At the object noun region, there was a significant interaction between Verb-Type and NP-Type ($\beta = -.11$, $SE = .047$, $t = -2.357$, $p = .02$). AVs were faster with EventNs ($-.009$ vs. .03) while N-AVs were faster with EntityNs ($-.025$ vs. .046).

At the region of the adverbial, we found both a highly significant interaction ($\beta = -.13$, $SE = .04$, $t = -3.283$, $p = .001$) and main effects of Verb-Type ($\beta = .01$, $SE = .028$, $t = 3.547$, $p = .0004$) and NP-Type ($\beta = .075$, $SE = .028$, $t = 2.668$, $p = .008$). As can be seen in the plot, at the adverbial, N-AVs were slower than AVs and EntityNs were slower than EventNs. A four-way comparison revealed that the AV+EventN condition (the red line in the plot) was significantly faster than both the AV+EntityN condition (green) ($\beta = .074$, $SE = .028$, $t = 2.6$, $p = .014$), and the N-AV+EventN condition (blue) ($\beta = .1$, $SE = .03$, $t = 3.348$, $p = .002$). Additionally, the N-AV+EventN condition (blue) was marginally slower than the N-AV+EntityN

[5]For the implementation in R see Jaeger's blog entry `https://hlplab.wordpress.com/2008/01/23/modeling-self-paced-reading-data-effects-of-word-length-word-position-spill-over-etc/`

condition (purple) ($\beta = -.0564$, $SE = .031$, $t = -1.802$, $p = .07$), but not significantly slower than the AV+EntityN condition (green) ($\beta = -0.025$, $SE = .033$, $t = -0.762$, $p = .45$).

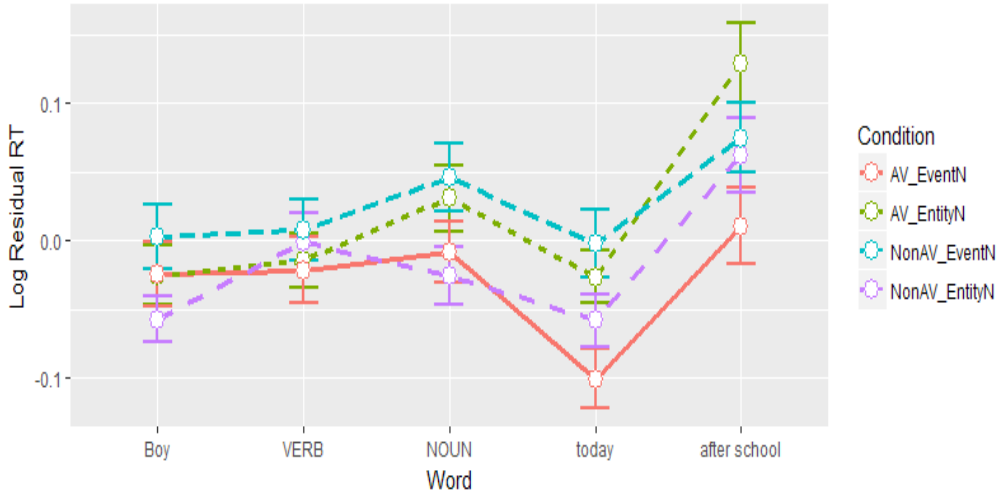


Figure 1: Sub-Experiment 1. Average Log-Transformed Residual Reading Times.

At neither the noun nor the adverbial region was the AV+EventN condition (red line) significantly different from the N-AV+EntityN condition (purple) ($\beta = .044$, $SE = .028$, $t = 1.572$, $p = .12$, $\beta = .04$, $SE = .031$, $t = 1.385$, $p = .18$). For the four-way comparison we used simple contrast coding with AV+EventN as the baseline because the theory predicts that AVs select for EventNs, hence the satisfaction of selectional requirements in this condition can account for the faster reading times seen in the results. The fact that EventNs were significantly faster with AVs than with N-AVs suggests that a processing cost was associated with the N-AV+EventN combination. This indicates that N-AVs in our experiment biased EntityN complements just like AVs biased EventN complements.

4.1.4. Discussion

At the region of the adverbial, we found the two-way contrast in (12). This result is compatible with the predicted three-way contrast in (8), but not the two-way contrast in (9) predicted by Dimensional Ambiguity hypothesis, because the two conditions that receive longer reading times, AV+EntityN and N-AV+EventN, do not form a class on the Dimensional Ambiguity theory.

(12) AV+EntityN , N-AV+EventN > AV+EventN, N-AV+EntityN
‘begin a book’, ‘see a fight’ > ‘begin a fight’, ‘see a book’

The faster reading times with both, AV+EventN and N-AV+EntityN, suggest that in both cases processing might be facilitated by the fact that the parser encounters what it expects, that is, the complement whose type matches the selectional restrictions of the predicate. What we can conclude from the bias for EntityNs after N-AVs is that N-AVs select for EntityNs as their ‘preferred’ arguments in contrast to EventNs.

Notably, the predicted 3-way contrast in (8) also involves a slow-down for the AV+EntityN condition, which we did not find, as opposed to Traxler et al. (2002). The AV+EntityN condition (green) and the N-AV+EventN condition (blue) did not differ: they were both significantly slower than the AV+EventN condition (red). This null result means that in our experiment coercion was no more costly than the combination of N-AVs with EventNs. We have just said that N-AVs combine with EventNs as their 'less preferred' complements, so does our null result suggest the same for AVs with EntityNs? Given the wide range of prior experiments showing extra cost for coercion contexts beyond the cost of the dispreferred combinations (e.g., McElree et al. (2001) discussed in section 3), we speculate that the absence of the extra cost in the AV+EntityN condition in our experiment results from the fact that we increased our participants' exposure to coercion contexts by combining the two sub-experiments. The processing of coercion became easier, at the same time, however, new differences became exposed, namely that both AVs and N-AVs create a bias for a particular semantic type of their nominal complement.

4.2. Sub-Experiment 2

The goal of this sub-experiment was to test the assumption of the Dimension Ambiguity approach that the EntityNs as complements to AVs are ambiguous between different dimensions along which the event could be structured. This predicts that the ease of identification of a particular dimension will affect the ease of processing of AV+EntityN combinations and that in the absence of ambiguity between dimensions, i.e., with N-AVs, no processing costs should be observed. As explained at the beginning of section 4, we compared two kinds of EntityNs available in Polish that differed in the likelihood of a temporal interpretation and thus should result in two different processing profiles. The first kind were morphologically simple nouns (SimpleNs) which denoted physical objects (e.g. *kolekcja* 'collection'), and thus should give rise to dimension ambiguity as complements to AVs vs. N-AVs. The ambiguity should elicit longer RTs already at the noun region.

The second kind were morphologically complex nouns (ComplexNs) (e.g. *zbiór* 'set', 'collection') containing verbal roots (verb: *zbier-ać* 'to collect') which also denoted entities, but in some cases could have a subordinate event-reading (we should note here that these nouns were not nominalizations, cf. *zbier-anie*). We hypothesized that ComplexNs might contrast with SimpleNs in that they would more readily allow the selection of the temporal dimension in the context of AVs. What is special about AVs on the Dimensional Ambiguity approach is that they require the identification of a particular dimension, which needs to happen with both ComplexNs and SimpleNs. With ComplexNs, however, this could be easier because the temporal dimension could be easier to access due to the verbal root (and the secondary eventive reading for some of the items). Accordingly, we should find the following three-way contrast:

(13) AV+SimpleN > AV+ComplexN > N-AV+ComplexN, N-AV+SimpleN

The alternative approach, Enriched Composition, makes two kinds of predictions: nouns with verbal roots may be more ready to participate in the entity-to-event semantic shift, (14); or

alternatively, the semantics of the verbal root may not be able to override the semantics resulting from the nominal morphology, (15) (this is a more compositional version of the approach).

(14) AV+SimpleN > AV+ComplexN, N-AV+ComplexN, N-AV+SimpleN
type-mismatch > type-match (assuming access to the event semantics of the verbal root of ComplexNs)

(15) AV+SimpleN, AV+ComplexN > N-AV+ComplexN, N-AV+SimpleN
type-mismatch > type-match (assuming **no** access to event semantics of the verbal root of ComplexNs)

4.2.1. Materials and procedure

We constructed 24 items in four conditions: AV+SimpleN, AV+ComplexN, N-AV+SimpleN, N-AV+ComplexN. We selected 24 SimpleN–ComplexN pairs that minimally differed in meaning. For instance *zbiór* 'set', 'collection' is usually used in reference to the mathematical object while *kolekcja* is not; both, however, can be used to refer to a gathering of valuable items, e.g., in a museum. We were able to construct 24 such pairs, and tried to match their members in terms of frequency, although this was not always possible. For example, a corpus can provide the frequency of use in the formal register, but some members in the pairs were very colloquial and frequent in everyday contexts (e.g., *jedzonko* 'goodies', 'food' has only 571 hits in the monitor corpus of Polish (`http://monco.frazeo.pl/`), whereas *potrawa* 'dish', 'plate', 'meal' has 4033 hits because it is frequently used in written recipes). In the appendix, we provide the list of the ComplexN–SimpleN pairs indicating if there is a difference in their usage frequency.

Furthermore, five of the ComplexNs have a dominant entity-reading, as well as a secondary event-reading. The latter is hardly available for those words presented in isolation, but can be made explicit by adding some minimal context (e.g., *zbiór truskawek* 'strawberry picking' or *'Zbiór twarł 2 godziny'* 'The gathering lasted 2 hours'). Those five nouns were: *zbiór* 'set', *wydruk* 'printout', *przesyłka* 'package', *okop* 'trench', *opowieść* 'story'. There were six ComplexNs which cannot have event readings, *budynek* 'building', *zapałka* 'matchstick', *rysunek* 'drawing', *rzeźba* 'sculpture', *napój* 'drink', *mrożonki* 'frozen food', as evidenced by the fact they cannot appear in the frame *'X lasted 2 hours'*. The rest of the ComplexNs denote the results of activities, e.g. *napar*, 'infusion', and if they appear in our test frame, they start denoting events but those readings are odd unless more supporting context is present.

Crucially, ComplexNs contained roots that are related to the verbs implicit in the event recovered in the coercing contexts and not any other verbs (compare how in English *'John began the queue'* refers to a queueing event, whereas *'John began the pullover'* means that John started making a garment named *pullover* and not that he started pulling the garment over himself).

We used the same set of verbs as in Sub-Experiment 1.[6] The procedure and the treatment of the data for statistical analysis was as described in section 4.1.1.

4.2.2. Results

The plot of the reading data in Figure 2 shows that already on the verb there were differences in reading times, with the N-AV+SimpleN condition being the slowest (the purple line in the plot). We found a main effect of NP-Type in this region ($\beta = .073$, $SE = .03$, $t = 2.442$, $p = .015$) and a marginal interaction ($\beta = -.08$, $SE = .043$, $t = -1.748$, $p = .08$). The main effect (the N-AV+SimpleN and N-AV+ComplexN conditions are significantly slower than the AV+SimpleN and AV+ComplexN conditions) is clearly driven by the slow reading times for the the N-AV+SimpleN condition. A four-way comparison reveals that N-AV+SimpleN (purple) condition is significantly slower than the N-AV+ComplexN (blue) condition ($\beta = -.073$, $SE = .03$, $t = -2.393$, $p = .019$) but is not significantly slower than the other two conditions. This result is mysterious because the subject and the verb in the N-AV+SimpleN and N-AV+ComplexN conditions were identical, so the reading times should be no different (as is the case in Sub-Experiment 1 – the sets of verbs in both experiments was identical).[7]

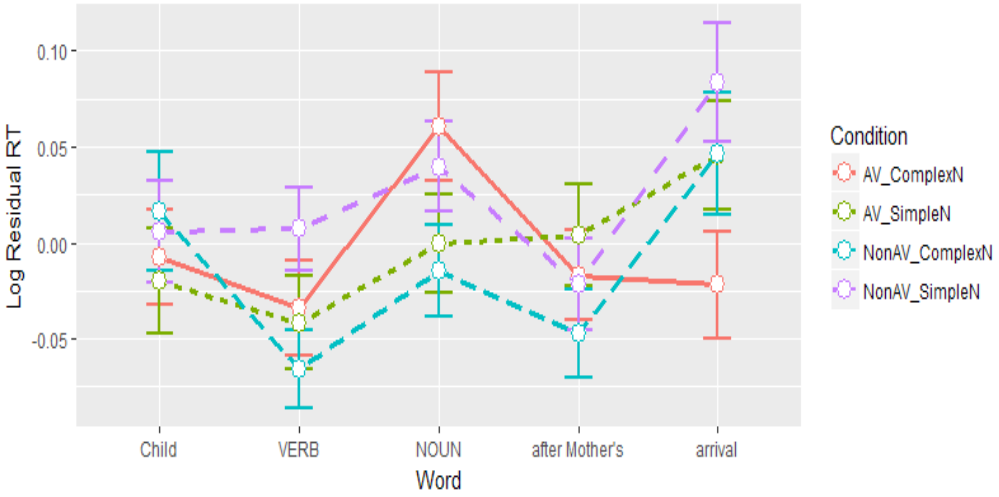


Figure 2: Sub-Experiment 2. Average Log-Transformed Residual Reading Times.

At the noun region there was a significant interaction between Verb-Type and NP-Type ($\beta = -.113$, $SE = .051$, $t = -2.212$, $p = .031$) and a main effect of Verb-Type ($\beta = .073$, $SE = .035$, $t = 2.072$, $p = .04$), but no effect of NP-Type ($\beta = .053$, $SE = .039$, $t = 1.354$, $p = .18$). A four-way comparison revealed that the AV+ComplexN condition (the red line in the plot) was significantly slower than the N-AV+ComplexN condition (blue) ($\beta = -.074$, $SE = .039$, $t = -2.129$, $p = .036$) and marginally slower than the AV+SimpleN condition (green) ($\beta = -.06$,

[6]Therefore, we needed to remove two items with the verb *forget* as explained in footnote 4. We also removed one item with a typo in the verb.

[7]We excluded the possibility that an outlier was driving this effect or that the slow-down accumulated over the course of the experiment for some reason: in Appendix A we present a boxplot and the distribution of RTs over the course of the trials showing that the RTs in the N-AV+SimpleN condition are slower at the beginning of the experiment.

$SE = .034$, $t = -1.683$, $p = .09$). This means that ComplexNs are harder with AVs than with N-AVs and that after reading an AV, a ComplexN is harder to process than a SimpleN.

At the regions of the subject and the adverbial, there were no significant differences in reading times.

4.2.3. Discussion

The experiment showed that ComplexNs are harder with AVs than with N-AVs, while SimpleNs are equally hard with both kinds of verbs (though the result that N-AV+SimpleN condition is so slow may be an artifact given the oddly high RTs at the verb). Additionally, ComplexNs are harder to process than SimpleNs following AVs. The differences in reading times can be summarized as follows:

(16) AV+ComplexN, (N-AV+SimpleN) >
(N-AV+SimpleN), AV+SimpleN, N-AV+ComplexN

The greatest processing cost occurred in the AV+ComplexN condition (red), and this condition was significantly slower than the N-AV+ComplexN condition (blue), and marginally slower than the AV+SimpleN (green). We predicted that AVs+ComplexNs would elicit the processing profile of 'biased' ambiguities (as identified in Rayner and Duffy (1986); see introduction to section 4), such that the RTs at the noun region would be faster than with SimpleNs and slower RTs would emerge at the post-target region. This prediction was not borne out, hence, we conclude that the ambiguity was not processed, possibly because the morphological complexity prevents access to the verbal root during semantic composition. Since there was no significant difference between the N-AV+ComplexN (blue) and the AV+SimpleN (green) conditions with respect to the AV+ComplexN condition (red), we can infer that morphological complexity without coercion (blue) is as hard to process as regular nouns with coercion (green).

The RTs in the N-AV+SimpleN condition (purple) are difficult to interpret, due to the mysteriously high RTs at the verb region which carry over to the noun region. It needs to be pointed out, however, that the difference between N-AV+SimpleN condition (purple) and N-AV+ComplexN condition (blue) is not significant, suggesting that in the absence of an AV there is no special cost for morphological complexity (while the cost in the AV+ComplexN condition (red) is evidenced by significantly longer RTs than in the N-AV+ComplexN condition (blue)). Accordingly, we place the N-AV+SimpleN condition in parentheses in (16).

Since the items from Sub-Experiment 2 were intermixed with the items from Sub-Experiment 1 we can conclude that the participants were sensitive to the semantic differences between the entity-denoting and event-denoting complements and the need for coercion in the context of AVs. Therefore, the fact that Sub-Experiment 2 showed the largest processing cost in the coercion condition with morphologically complex nouns seems to indicate that real-time comprehension semantic composition has no access to the morphological structure below the word level. This result is compatible with the predictions of the Enriched Composition approach in (15) where direct compositionality does not allow access to the semantics of the root inside a

complex noun (recall that ComplexNs were not nominalizations). The predictions in (15) distinguish between the conditions with type-match and those with type-mismatch, but we found that morphological complexity adds to the cost of type-mismatch.

The result is also compatible with the predictions of the Dimensional Ambiguity approach, because the fact that the AV+SimpleN condition was not the hardest to process as predicted in (13), could be due to the easier resolution of dimension ambiguity in this condition than in the AV+ComplexN condition. As opposed to what we hypothesized, morphological complexity does not facilitate the selection of a dimension appropriate to the semantics of the verb.

5. General discussion

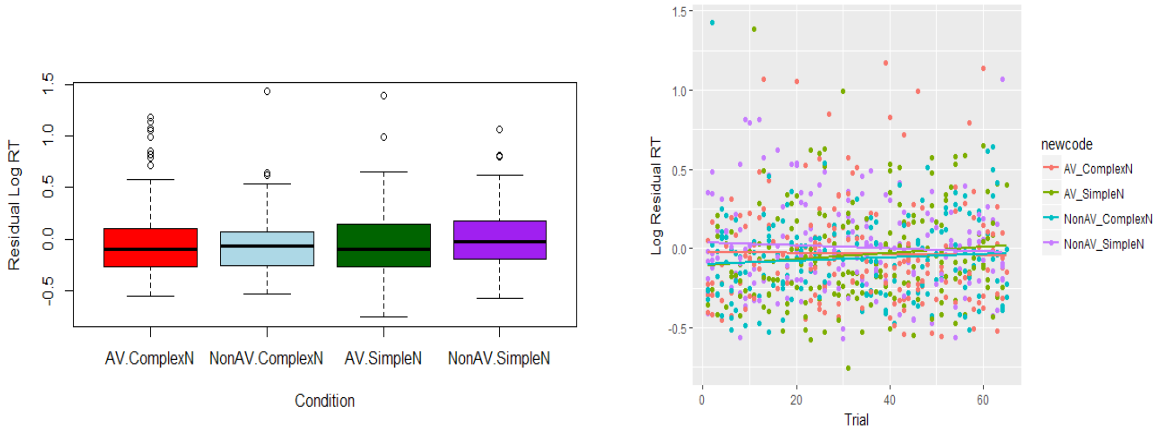
Our experiment showed that AVs with EventNs (*'begin a fight'*) and N-AVs with EntityNs (*'see a book'*) were easier to process than AVs with EntityNs (*'begin a book'*) and N-AVs with EventNs (*'see a fight'*). This result indicates that processing is easier when the parsing expectations created by the predicate are met. AVs create an expectation for EventNs, which is unsurprising if they specifically select for them. N-AVs, on the other hand, are semantically compatible with both EntityNs and EventNs as their argument, yet EntityNs appear to be preferred over EventNs (at least in Polish). The assumption of the Dimensional Ambiguity hypothesis, that what is at stake for AVs in coercion contexts is the difficulty of identifying the specific dimension along which the complement is construed as a structured individual, cannot account for the fact that the AVs+EventNs and N-AVs+EntityNs are easier to process. However, a theory based on selectional restrictions allows us to explain why these combinations facilitate processing. We should note here that our experimental setup with the increased exposure to coercion contexts (25% of the sentences in the whole experiment) resulted in the absence of a special processing cost for the AV+EntityN condition. Yet, the obtained results clearly show that predicates impose semantic restrictions on their arguments, which supports the Enriched Semantic Composition hypothesis. A theory based on selectional restrictions can also accommodate the result that N-AVs such as *see* combine with both EntityNs and EventNs but are easier to process with the former. This result is in line with the findings in Spalek (2014), who has shown that verbs select for a variety of semantic types of complements, where some of those types are the primary ones while others can be considered secondary but still not anomalous.

Selectional restrictions have been the subject of a longstanding discussion at least since Katz and Fodor (1963) and later Chomsky (1965), who incorporated them as part of grammar in an attempt to express limitations on the applicability of predicates to arguments. Two major questions have been at stage ever since: 1) whether selectional restrictions should be treated as a lexical semantic property or as a matter of world knowledge, and 2) how rich and rigid the inventory of selectional restrictions should be. Our results contribute to this discussion by showing that the idea of incorporating conceptual knowledge into compositional semantics through the notion of selectional restrictions of predicates is fundamentally right, but should be refined by taking into account combinatorial preference patterns. On the basis of our findings we can conclude that AVs differ form N-AVs in that they impose a strong selectional constraint for event nouns, whereas N-AVs have merely a preference for entity-denoting nouns but also

combine with event nouns. Our results suggest that a graded notion of selectional satisfaction conditions as mentioned in the work of Wilks (1978) is more appropriate. We propose that the satisfaction conditions of a predicate should be understood as a graded continuum of highly typical arguments and less probable arguments and should thus be called more appropriately 'selectional preferences'. Accordingly, the degree to which various selectional (mis)matches will be judged as semantically anomalous will vary since they do not result from a violation of hard constraints (Resnik, 1996).

Appendix A

Sub-Experiment 2. The Distribution of Log-Transformed Residual Reading Times at the Verb Region summarized in a Boxplot and a Scatterplot over Trials 1-65.



Appendix B

ComplexN – SimpleN pairs in Sub-Experiment 2 and their frequency in everyday use:

zbiór 'set' = *kolekcja* 'collection'
wydruk 'printout' = *odbitka* 'photo print'
przesyłka 'package' = *paczka* 'package'
układanka 'puzzle' = *puzzle* 'puzzle'
opowieść 'story' = *historia* 'story'
budynek 'building' = *dom* 'house'
rysunek 'drawing' = *obrazek* 'drawing, painting'
rzeźba 'sculpture' = *posąg* 'statue'
nasyp 'earth mound' = *sterta* 'heap'
napój 'drink' = *sok* 'juice'
mrożonki 'frozen food' = *lody* 'ice-cream'
wiązanka 'bouquet' = *bukiet* 'bouquet'
przekąska 'snack' = *przystawka* 'appetizer'
naklejka 'sticker' = *ozdoba* 'ornament'
zapałka 'matchstick' =? *świeczka* 'candle'
okop 'trench' < *tunel* 'tunnel'
napar 'infusion' < *herbata* 'tea'
wyszywanka 'embroidery' < *haft* 'embroidery'
czytanka 'school text' < *powieść* 'novel'
malowanka 'picture, colouring' < *ilustracja* 'illustration'
plecionka 'wickerwork, weaved object' < *koszyk* 'basket'
wycinanka 'cutout' < *wzór* 'design'
pismo 'letter, writing' > *list* 'letter'

References

Asher, N. (2011). *Lexical Meaning in Context*. Cambridge: Cambridge University Press.

Asher, N. (2015). Types, meanings and coercions in lexical semantics. *Lingua 157*, 66–82.

Baayen, R. H., D. J. Davidson, and D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language 59*(4), 390–412.

Baayen, R. H. and P. Milin (2010). Analyzing reaction times. *International Journal of Psychological Research 3*(2), 12–28.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language 68*(3), 255–278.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1–48.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological) 26*(2), 211–252.

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

Copestake, A. and T. Briscoe (1995). Semi-productive polysemy and sense extension. *Journal of Semantics 12*(1), 15–67.

de Swart, H. (2011). Mismatches and coercion. In C. Maienborn, K. von Heusinger, and P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, Volume 1, pp. 574–597. Berlin: de Gruyter.

Egg, M. (2003). Beginning novels and finishing hamburgers: Remarks on the semantics of *to begin*. *Journal of Semantics 20*(2), 163–191.

Ferreira, F. and C. Clifton (1986). The independence of syntactic processing. *Journal of Memory and Language 25*(3), 348–368.

Fodor, J. and E. Lepore (1998). The emptiness of the lexicon: Critical reflections on James Pustejovsky's *The Generative Lexicon*. *Linguistic Inquiry 29*(2), 269–288.

Frisson, S. and B. McElree (2008). Complement coercion is not modulated by competition: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition 34*(1), 1–11.

Gawron, M. (2009). The lexical semantics of extent verbs. Ms. San Diego State University.

Hofmeister, P. (2011). Representational complexity and memory retrieval in language comprehension. *Language and Cognitive Processes 26*(3), 376–405.

Hofmeister, P., I. Arnon, T. Jaeger, I. Sag, and N. Snider (2013). The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes 28*(1), 48–87.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jaeger, T., E. Fedorenko, P. Hofmeister, and E. Gibson (2008). Expectation-based syntactic processing: Anti-locality outside of head-final languages. Oral presentation at CUNY.

Just, M. A., P. A. Carpenter, and J. D. Woolley (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General 111*(2), 228–238.

Katsika, A., D. Braze, A. Deo, and M. Piñango (2012). Complement coercion: Distinguishing between type-shifting and pragmatic inferencing. *The Mental Lexicon 7*(1), 58–76.

Katz, J. J. and J. A. Fodor (1963). The structure of a semantic theory. *Language 39*(2), 170–210.

Kuperberg, G., A. Choi, N. Cohn, M. Paczynski, and R. Jackendoff (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience 22*(12), 2685–2701.

Kuznetsova, A., P. Bruun Brockhoff, and R. Haubo Bojesen Christensen (2016). *lmerTest – Tests in Linear Mixed Effects Models*. R package version 2.0-33.

Lai, Y.-Y., C. Lacadie, T. Constable, A. Deo, and M. Piñango (2014). Complement coercion as the

processing of aspectual verbs: Evidence from self-paced reading and fMRI. In *Proceedings of the Cognitive Science Society Meeting*, Volume 36, pp. 2525–30.

McElree, B., L. Pylkkänen, M. J. Pickering, and M. J. Traxler (2006). A time course analysis of enriched composition. *Psychonomic Bulletin and Review 13*(1), 53–59.

McElree, B., M. J. Traxler, M. Pickering, R. E. Seely, and R. Jackendoff (2001). Reading time evidence for enriched composition. *Cognition 78*(1), B17–B25.

Piñango, M. and A. Deo (2015). Reanalyzing the complement coercion effect through a generalized lexical semantics for aspectual verbs. *Journal of Semantics 33*(2), 1–50.

Pustejovsky, J. (1991). The generative lexicon. *Journal of Computational Linguistics 17*(4), 409–441.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Pylkkänen, L. and B. McElree (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience 19*(11), 1905–1921.

R Core Team (2016). *R – A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rayner, K. and S. A. Duffy (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition 14*(3), 191–201.

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition 61*(1–2), 127–159.

Rohde, D. (2001). Linger. `http://tedlab.mit.edu/~dr/Linger/`.

Spalek, A. (2014). *Verb Meaning and Combinatory Semantics: A Corpus Based Study of Spanish Change of State Verbs*. Ph. D. thesis, Universitat Pompeu Fabra, Barcelona.

Traxler, M. J., B. McElree, R. S. Williams, and M. J. Pickering (2005). Context effects in coercion: Evidence from eye movement. *Journal of Memory and Language 53*(1), 1–25.

Traxler, M. J., M. J. Pickering, and B. McElree (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language 47*(4), 530–547.

Trueswell, J. and M. Tanenhaus (1994). Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, Jr., L. Frazier, and K. Rayner (Eds.), *Perspectives on sentence processing*, pp. 155–179. Hillsdale, NJ: Erlbaum.

Vasishth, S., Z. Chen, Q. Li, and G. Guo (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS ONE 8*(10), e77006.

Wilks, Y. (1978). Making preferences more active. *Artificial Intelligence 11*(3), 197–223.