Competing Analyses of Complement Coercion: New Evidence from Behavioral and Electropsychophysiological Methods

Laura DiNardo

Advisor: Maria Mercedes Piñango

Submitted to the faculty of the Department of Linguistics
in partial fulfillment of the requirements for the degree of
Bachelor of Arts

Yale University
20 April 2015

Abstract

The term "coercion" has been used in linguistics to define a wide range of phenomena in which semantic mismatches are resolved through extra-syntactic processing (Jackendoff, 1997; Klein & Sag, 1985). The present study investigates one particular manifestation: complement coercion.

In complement coercion, a coercion verb is combined with an entity-denoting complement, resulting in a semantically underspecified meaning (Piñango, 2014). To explain the semantic processes underlying these types of sentences, the literature has traditionally proposed the Type-Shifting Hypothesis. This hypothesis is grounded in the observation that coercion verbs (such as *begin* and *enjoy*) and entity-denoting complements combine to result in an eventive interpretation of the complement (Jackendoff, 1997; Pustejovsky, 1991, 1995). Previous work (Baggio et al., 2010; Kuperberg et al., 2010) presents evidence consistent with this hypothesis, demonstrating that coercion verbs elicited higher processing costs, attributed to type-shifting, in real-time comprehension.

However, recent real-time processing investigations have challenged this traditional definition: converging evidence from Katsika et al. (2012) and Lai et al. (2014) suggest that a reexamination of coercion verbs is needed. While previous work has considered both aspectual and psychological verbs to be part of one homogenous set of coercion verbs, new evidence demonstrates that the increased processing cost of these verbs is elicited by only the aspectual (*begin*) type. These findings not only challenge the interpretations from the experimental literature, they also call for a reanalysis of the source of complement coercion's processing cost.

This study uses offline and electrophysiological methodologies to arbitrate between three hypotheses of complement coercion processing. The findings, consistent only with the Structured Individual Hypothesis, both refine our understanding of the complement coercion phenomenon and provide insight into the real-time integration of syntactic and semantic information in the context of a parallel architecture.

Contents

List of Tables

List of Figures

1 Introduction

1.1 The Architecture of the Language Faculty and Its Relation to Coercion

An introduction begins my senior thesis.

In its colloquial use, the term "coercion" is defined as "the practice of persuading someone to do something by using force" (Merriam-Webster, 2014). When used in linguistics, this term describes a group of phenomena in which meaning is coerced, or forced, to express itself in a non-morphosyntactic manner. Therefore, an understanding of any coercion phenomenon must be grounded in the notion that the meaning of an entire sentence can be built at some deeper level of semantic structure, not apparent to the morphosyntax. Such an assumption, however, challenges the traditional view of meaning composition, which assumes that meaning is not built in parallel with syntactic structure but rather that syntactic structure contains semantic structure (Jackendoff, 2002). Consequently, coercion phenomena are intriguing in that they present a challenge to this traditional view.

To understand how coercion, and particularly complement coercion, is conceptualized and processed, we must first invoke a cognitive model to understand the way that individual lexical items and stored and then recalled to build larger structures in the language faculty. The present study is grounded in the Tripartite Parallel Architecture model (Jackendoff, 1997), henceforth referred to as simply the Parallel Architecture. In his chapter, *The Parallel Architecture and Its Place in Cognitive Science*, Jackendoff defines the Parallel Architecture as the notion that "the generative capacity of language is invested in multiple components...Each component has its own distinctive primitives and principles of combination and generates its own structure." (Jackendoff, 2010, p. 647). Crucially within this model, phonological, syntactic, and semantic structures exist separately but are interfaced with one another via the linking of their parts to express language.

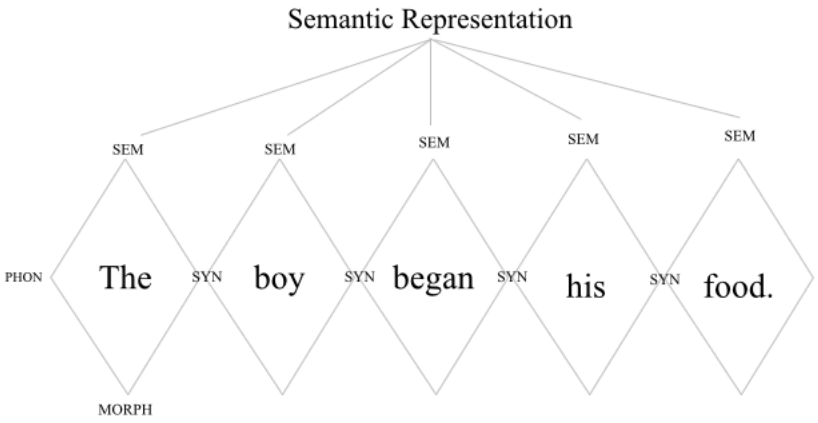
The Parallel Architecture is a direct contrast to a syntax-centered model, in which the "recursive rules of the syntactic component provide the generative capacity of language." (Jackendoff, 2009, p. 646) Therefore, a syntax-centered approach states that meaning arises through the mapping of syntactic structures onto semantic forms. This type of model, however, cannot readily account for cases such as coercion, in which the meaning of a sentence cannot arise from its syntactic structure alone. The Parallel Architecture, on the other hand, provides a framework in which syntax and semantics can be built at least semi-independently and then interfaced to give rise to a sentence's

meaning.

With the idea of the Parallel Architecture in mind, one way to understand phenomena such as coercion is through the notion of enriched composition (Jackendoff, 1997). Enriched composition is a direct contrast to the standard view of semantic composition in which an interpretation is determined by combining lexical representations sequentially in the order of their surface syntactic structure. Rather, in enriched composition, all lexical components are built in parallel and then combined such that the semantic representation of a sentence reflects an integration of the semantic structures of individual items to form one larger semantic structure. As in the model of the Parallel Architecture, the key aspect of enriched composition is that the meaning of a sentence does not always need to be expressed in the surface form of a lexical item, but can be contained in the deeper semantic representations and combined to form a composite structure (Jackendoff, 1997).

To help visualize such a model, an example sentence built from a series of lexical items has been provided. Critical to this model, illustrated Figure (1), is the fact that the semantic representation of each individual lexical item contributes to the larger semantic representation of the sentence and is concatenated independently from the sequential building of syntactic structure.

Figure 1: A Model of the Parallel Architecture



The phenomenon of coercion is a relevant example of enriched composition in that fundamental semantic elements of coercion sentences are not present in the surface structure and yet are still understood by the processor. As mismatches between what is expressed at the syntactic level and what is expressed at the semantic level are integrated for sentence comprehension, coercion is able

to further corroborate the Parallel Architecture by demonstrating that syntax and semantics are not united via a direct one-to-one link. While the ultimate goal of the Parallel Architecture and enriched composition is to maintain the notion of one-to-one correspondence between form and meaning, much like the syntax-centered approach, these models demonstrate that the processes of building syntax and connecting it to meaning do not need to happen unidirectionally (Jackendoff, 1997).

1.2 A Linguistic Analysis of Complement Coercion

Complement coercion is defined as a phenomenon in which a coercion verb is combined with an entity-denoting complement, resulting in a semantically underspecified meaning (Piñango, 2014). This underspecified semantics requires a composed meaning, such that an eventive interpretation of the sentence is obligatory. To understand this phenomenon we must first define what a coercion verb is and what an underspecified semantics is.

Traditionally, the study of complement coercion has focused on the properties of two particular verb classes, aspectual verbs and psychological verbs. Aspectual verbs are verbs whose meaning is temporal, in that they introduce quantification over some event whose subpart is referred to by the verb (Katisika et al., 2012). Some quintessential examples of aspectual verbs are *begin, finish, start, complete, continue, end*. Psychological verbs, on the other hand, are verbs whose external argument is an experiencer and whose complement is the target of an emotion (Katsika et al., 2012). Examples of psychological verbs are *endure, tolerate, enjoy, resist, favor, prefer*.

In particular, Pustejovsky (1995) and Jackendoff (1997) proposed that aspectual (e.g. *begin*) and psychological (e.g. *enjoy*) verbs carry a selection restriction such that they require combination with event-denoting complements. It was observed, however, that when these verbs are combined with an ordinary (entity-denoting) complement, an eventive interpretation still occurs. This eventive interpretation has been thought to be attributed to some sort of extra-syntactic resolution, e.g. a type-shifting operation/mechanism (Pustejovsky, 1995). Consider the example:

(1) *The tourist began the Freedom Trail in Boston.*[1]

As a well-formed expression of English, most native speakers would readily agree that this sentence means that the tourist began walking, running, studying, etc. the Freedom Trail in Boston.

[1] Example borrowed from Lai et al. (2014) experimental stimuli.

Whatever action is interpreted, it is understood that some sort of event took place in the expression, such that the tourist began the event and the event was located on the Freedom Trail. Nowhere in the surface structure of the sentence, however, is this interpretation directly supplied. In fact, the only semantic structure provided for the processor is:

(2) *The tourist began the Freedom Trail in Boston.*
Underspecified semantic structure: [AGENT[EVENT$_{begin}$ [THING$_{entity}$]]]

where the intended event is inserted as:

(3) Complement coercion reading: [AGENT[EVENT$_{begin}$[EVENT[THING$_{entity}$]]]]
Resulting Meaning: *The tourist began* [*walking, running, etc.*] *the Freedom Trail in Boston.*

Similarly, the same sort of underspecified structure has been posited in complement coercion sentences using psychological verbs.

(4) *The tourist enjoyed the Freedom Trail in Boston.*
Complement coercion reading: [AGENT[EVENT/STATE$_{enjoy}$[EVENT[THING$_{entity}$]]]]
Resulting Meaning: *The tourist enjoyed* [*walking, running, etc.*] *the Freedom Trail in Boston.*

1.3 The Starting Point of the Present Study

It has been shown that complement coercion engenders a higher processing cost in real-time comprehension. This cost has been observed across a variety of methodologies such as self-paced reading (McElree et al, 2001), eye-tracking (Pickering et al., 2005; Traxler et al., 2002), focal brain lesion (Piñango & Zurif, 2001), MEG (Pylkkänen & McElree, 2001), ERP (Baggio et al., 2010; Kuperberg et al., 2010), and fMRI (Husband et al., 2011; Lai et al., 2014).

Three unique hypotheses have been proposed to account for this observed processing cost: The Type-Shifting Hypothesis (McElree et al., 2001; Pickering et al., 2005), the Structured Individual Hypothesis (Piñago & Deo, 2012), and the Surprisal Hypothesis (Delogu, 2013). In the following section, each hypothesis will be presented along with its supporting experimental evidence. Then, three new studies will be presented to help determine which of the hypotheses best explain complement coercion's processing cost.

1.4 A General Overview of Methodologies

Three common methodologies used in this study and in psycho- and neurolinguistic literature are presented below: the lexical decision task, Cloze test, and EEG. Other methodologies are mentioned throughout the paper, however, these three are the most relevant to the present study.

1.4.1 Lexical Decision Task

This online methodology is comprised of an experimental task in which participants are asked to classify a phonologically permissible string of letters as either a word or a non-word in their language, and to do so as quickly and accurately as possible (Meyer & Schvaneveldt, 1971). The participants' reaction times, measured in milliseconds (ms), while performing the task are thought to be representative of the word string's meaning and familiarity. In particular, lexical decision tasks have been successfully been used as a proxy for determining the processing difficulty of various verb classes when presented in isolation (Rayner & Duffy, 1986).

1.4.2 Cloze Test

A Cloze test is an offline methodology used to assess the readability of a sentence or passage. (Oiler & Conrad, 2006). In a Cloze test, sentences are presented with words removed and the task of the participant is to fill in the missing words. Thus, this test measures the participants ability to understand context to correctly fill in a blank. Cloze tests have been reliably shown to demonstrate the comprehension of sentences by subject (Taylor, 1953).

1.4.3 EEG

A commonly used methodology for studying the time course of processing is electroencephalography (EEG). This methodology measures the neurophysiological activity of neuronal clusters through metal electrodes placed on the surface of the scalp (Bornkessel & Schlesewsky, 2009). Measurements of this activity are time-locked to critical stimuli (e.g. a specific word in a sentence). These neurophysiological responses are then averaged together to make event-related potentials (ERPs). A benefit of EEG experimentation is that it provides high temporal resolution of this neurophysiological activity, however it provides only low spatial resolution.

Within ERP experimentation, two important language-related components are N400 and P600. The first component, the N400, is defined as a negative waveform of electrical activity, with a peak somewhere between 300 and 500 ms after the onset of a word. Widely studied in the literature, this component has been implicated in lexical activation particularly in lexical retrieval and access (Brown & Hagoort, 1993; Lau et al., 2008). The P600 component is a positive waveform with a peak somewhere between 500-700 ms after the onset of a word and is associated with ambiguity resolution and syntactic-semantic integration (Kaan & Swaab, 2003; Friderici & Weissenborn, 2007). Given that the processing of complement coercion involves lexical access, ambiguity resolution, and syntactic-semantic integration, it is no surprise that these two components are the main focus of real-time studies in this phenomenon.

2 The Competing Hypotheses in Complement Coercion

2.1 The Type-Shifting Hypothesis

The first proposed hypothesis is the Type-Shifting Hypothesis. The claim at the basis of the Type-Shifting Hypothesis is that coercion verbs select for event-denoting complements (McElree et al., 2001; Pickering et al., 2004). When a coercion verb, which requires an event-denoting complement, is instead paired with an entity-denoting complement, a mismatch repair must occur through a type-shifting operation. This type-shifting operation coerces the semantic type of the entity-denoting complement into the appropriate event denoting type (see McElree et al., 2001 for a summary). Returning to example (1) above, *The tourist began the Freedom Trail in Boston*, the complement is shifted from referring to an entity to referring to some activity associated with that entity (Pickering et al., 2005). The event associated with the entity is often one that is strongly linked to the object, such as *walking* in the example above.

Experimental evidence from two previous neurolinguistic studies on complement coercion are particularly relevant to the present one: Baggio et al. (2010) and Kuperberg et al. (2010). These studies have corroborated the Type-Shifting Hypothesis by observing a higher processing cost after the onset of the entity-denoting NP-complement when combined with a coercion verb, attributable to type-shifting operations. The findings of each experiment are summarized in the section to follow.

2.2 Experimental Evidence for the Type-Shifting Hypothesis

2.2.1 Baggio et al., 2010

The aim of Baggio et al. (2010) was to understand the ERP components of complement coercion by comparing the experimental stimuli of three conditions – coercive verbs, anomalous verbs, and neutral verbs. In this study, the authors grouped both aspectual and psychological verbs into their set of coercion verbs. Each participant read 159 experimental stimuli, 53 per condition, and 150 fillers. Results demonstrated that the NP-complement noun in coercion sentences evoked a larger N400 than NP-complement nouns in the control condition. A negative-going shift was also observed 700-1000 ms after the onset of the NP-complement noun, with a central scalp distribution. Interpretations of such an effect were unclear, although it was hypothesized that the processing of coercion verbs may be associated with downstream integration costs.

2.2.2 Kuperberg et al., 2010

The aim of Kuperberg et al. (2010) was also to examine the electrophysiological components of complement coercion. In this study, ERP signals were measured as participants read and made acceptability judgements about three sentence types – plausible coerced, plausible non-coerced, and highly implausible animacy violated sentences. Once again, the set of coercion verbs used in this study contained both aspectual and psychological verbs. Each participant saw a total of 338 sentences. Results demonstrated that coerced sentences and animacy violated sentences evoked a similar N400 at the NP-complement noun onset. Therefore, Kuperberg et al. (2010) suggest that the N400 effect could be attributed to a mismatch in types between the semantic restrictions of the verb and the semantic properties of the complement noun. Additional effects were seen in the coercion sentences at the sentence final words. Here, a sustained left anterior positivity was seen, which Kuperberg et al. (2010) suggests to reflect the delayed attempts of the processor to retrieve the specific events associated with the complement noun.

In the context of the present investigation, Baggio et al. (2010) and Kuperberg et al. (2010) present two major issues. First, although both studies implemented very similar design, their results were not completely in line with one another. While both found an N400 at the NP-complement noun following a coercion verb, Baggio et al. (2010) reported a long-lasting negative shift after the noun in central regions, while Kuperberg et al. (2010) saw an anterior positivity at the sentence final word. Second, as will be explained in the section to follow, both studies consider coercion verbs

to be a homogeneous set, rather than a heterogeneous one of aspectual and psychological verbs, leading to possible problems in the interpretations of their results.

2.3 Challenges to the Type-Shifting Hypothesis

One major theoretical challenge to the Type-Shifting Hypothesis is found in sentences where the subject of the coercion sentence is non-agentive. Looking back at example (1), *The tourist began the Freedom Trail in Boston,* an agentive subject combined with a coercion verb and an entity-denoting complement to result in a meaning that was only able to be composed once an event was inserted between the verb and the complement. However, consider an example such as:

(5) *A long coda finished the music recorded for the 1984 Olympics.*[2]
Complement coercion reading: **A long coda finished (singing, writing, etc.) the music recorded for the 1984 olympics.*

Here, the subject of the sentence (*a long coda*) provides a constitutive reading, meaning that it denotes a subinterval of some larger interval identified by the complement (Piñago & Deo, 2012). Henceforth, sentences of this type will be referred to as aspectual verbs with a constitutive subject. The insertion of an event after the coercion verb in a sentence with a constitutive subject renders that sentence nonsensical, unlike the examples we have seen thus far of coercion sentences with an agentive subject.

Sentences such as (5) present a problem for the Type-Shifting Hypothesis in that the foundation on which type-shifting is built states that all entity-denoting complements must shift into event-denoting ones. However, as seen in example (5), an eventive interpretation of sentences whose subjects are constitutive is impossible. Thus, not only does type-shifting not predict linguistic data similar to example (5), it cannot explain the processing cost observed in these sentences.

A second challenge to the Type-Shifting Hypothesis has been observed in recent real-time experiments of complement coercion processing. In the Type-Shifting Hypothesis, the set of coercion verbs contains both aspectual and psychological verbs. However, a recent eye-tracking study by Katsika et al. (2012), comparing aspectual to psychological verbs in the context of complement coercion, found that a higher processing cost was observed only with respect to aspectual verbs, and not to psychological verbs. These results indicate that the set of coercion verbs examined in

[2]Example borrowed from Lai et al., 2014 experimental stimuli.

prior literature may actually be uniting two heterogeneous verb classes under one phenomenon, when in fact they do not actually contain the same processing profile. Not only does this finding present a problem for the theoretical basis of type-shifting, it also points to possible weakness in previous experimental evidence and implementation associated with type-shifting.

2.4 The Structured Individual Hypothesis

Piñango & Deo (2012) propose an alternative hypothesis to explain the real-time processing costs associated with coercion verbs. This hypothesis states as a fundamental that the set of coercion verbs must be redefined as that of only aspectual verbs.

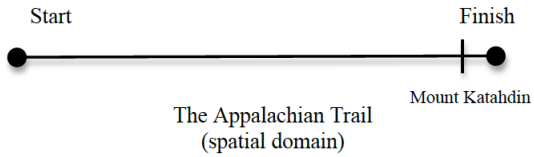
Aspectual verbs distinguish themselves from psychological verbs in their semantic representation within the mental lexicon. In their semantic representation, aspectual verbs lexically select for structured entities, which can be conceptualized as ordered structures along some dimension (e.g. temporal, spatial, eventive, etc.) (Lai et al. 2014). These dimensions are mapped onto axes that consist of ordered subparts, such that the aspectual verb specifies a relation between a structured individual and a subpart thereof, relative to the axis over which the individual extends (Piñango & Deo, 2012). One key aspect of this definition, in contrast to the explanation provided by the Type-Shifting Hypothesis, is that aspectual verbs do not exclusively require an event as their complement.

To illustrate this definition, consider example (6):

(6) *Mount Katahdin finishes the Appalachian Trail*.[3]

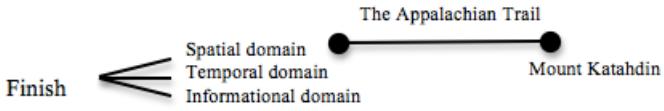
In the example above, the structured individual is the complement, *the Appalachian Trail,* which extends along some sort of spatial axis. This axis can be broken up into subparts – perhaps different trails, different states, or different mountains such as Mount Katahdin – such that the Appalachian Trail consists of a beginning point, middle points, and an end point. The coercion verb *finishes* specifies the relation between the subpart *Mount Katahdin* and its location along the structured individual. In this example in particular, it specifies the point at the end of the spatial axis, representing the spatial dimension of the *Appalachian Trail*, where Mount Katahdin is located.

[3]Example borrowed from Lai et al., 2014 experimental stimuli.



In real-time processing, however, ambiguity arises at the onset of the aspectual verb, as the processor builds a sentence sequentially. This ambiguity is due to the fact that there are several dimensions along which any aspectual verb can extend. These possible dimensions are lexically associated with the entries of each aspectual verb in the lexicon. Because several dimensions are possible for each aspectual verb, the interpretation of the aspectual verb is underdetermined at the onset of the verb and depends on which dimension is relevant in the context of utterance (Lai et al., 2014). Therefore, a dimension ambiguity arises when each set of possible dimensions is recalled, and can only be resolved when the proper one is selected after the entity-denoting complement has been reached.

Returning to the example above, the verb *finish* is one whose lexical entry can be associated with the spatial, temporal, eventive, and informational dimensions, among others, and thus when it is recalled by the processor, it is not until the complement, *the Appalachian Trail*, that the processor can select the spatial dimension to properly interpret the sentence.



It is this recall of all dimensions associated with the verb and then the selection of the correct one that is thought to be the source of the extra processing costs seen in real-time comprehension. A critical strength of this hypothesis is that it works for complement coercion sentences with either constitutive or agentive subjects. Thus, a similar sort of analysis can be done on sentences with an agentive subject:

(7) *Jimmy Carter began the anthology of 20th century inaugural speeches.*[4]

Similar to example (6), the verb *began* is associated with a set of dimensions over which some type of ordered axis can be extended. In the case of example (7), *began* can again be associated with

[4]Example borrowed from Lai et al. (2014) experimental stimuli.

the spatial, temporal, or informational dimensions, along whose axis *Jimmy Carter* is located at the start. Dimension ambiguity arises at the verb onset when all possible dimensions associated with *began* are recalled. Once the complement, *the anthology of 20th century inaugural speeches*, is reached, the processor knows to select for a temporal dimension, in which *Jimmy Carter*, the person, is located at the beginning of a point in time where he began doing something with a collection of speeches.

Therefore, examples (6) and (7) demonstrate that the Structured Individual Hypothesis is able to account for sentences whose subject is either agentive or constitutive. The strength of the Structured Individual Hypothesis lies in the fact that it provides an explanation for sentences with a constitutive subject, precisely the types of sentences that the Type-Shifting Hypothesis could not explain.

2.5 Evidence for the Structured Individual Hypothesis

2.5.1 Katsika et al., 2012

Katsika et al. (2012) represents a breakthrough study in the understanding of complement coercion. Using eye-tracking, this study aimed to resolve the confound found in complement coercion literature – that of a heterogeneous group of verbs being tested as coercion verbs in previous real-time experiments. Participants were provided with a context sentence and then one of three types of target sentences, containing either an aspectual, psychological, or control verb. Aspectual predicates were analyzed in comparison to psychological predicates and evidence demonstrated that the aspectual ones triggered coercion and a processing cost while psychological ones did not. These findings had direct implications on the way the study of complement coercion approached aspectual and psychological verb types, particularly with the way they are lexically encoded. Moreover, this study called into question the involvement of type-shifting or restructuring in the composition of psychological verbs with entity-denoting complements (Katsika et al., 2012). It was from this study that the Structured Individual Hypothesis began to be developed.

2.5.2 Lai et al., 2014

The aim of Lai et al. (2014) was to test the Structured Individual Hypothesis in real-time processing. The study contained two parts, first a self-paced reading task and then an fMRI task. The purpose of the first task was to investigate the time-course of the processing cost associated with

aspectual and psychological verbs in a complement coercion context. These costs were compared against one another to directly challenge all other studies who conflated the two verb types into one class under type-shifting. Each participant saw 150 test sentences and 150 filler sentences, for a total of 300 stimuli. The three verb types tested in this study were aspectual verbs, psychological verbs of the *enjoy*-type (traditionally considered coercion verbs within type-shifting), and psychological verbs of the *love*-type (controls). The results of this study demonstrated that aspectual verbs induced longer reading times than both *enjoy*-type and *love*-type psychological verbs, and moreover that the reading times between *enjoy*-type and *love*-type verbs did not significantly differ. The higher processing cost associated with aspectual verbs, as compared to psychological verbs, provides support for the Structured Individual Hypothesis, since the processor has to recall all possible dimension axes along which an aspectual verb can lie and then determine the correct axis for interpretation, while psychological verbs do not carry the same restriction.

The fMRI study used the same stimuli as the self-paced reading task. Each sentence was divided into two events – Event 1 including the subject and verb, and Event 2 including the complement until the sentence final verb. Thus, Lai et al. (2014) hypothesized that if the Structured Individual Hypothesis was invoked, activation of an aspectual verbs functions should occur at Event 1 and dimension ambiguity resolution should occur at Event 2. Indeed, results demonstrated that during Event 1, areas BA40 and BA7 were activated. BA7 is especially important support for the Structured Individual Hypothesis in that it is related to spatiotemporal tasks. At Event 2, the left inferior frontal gyrus (LIFG) was activated, which was attributed to the "process of determining the dimension along which the complement axis is structured" (Lai et al., 2014).

Therefore, the results of both Lai et al.'s (2014) self-paced reading study and fMRI study provided evidence consistent with the Structured Individual Hypothesis.

2.6 Challenges to the Structured Individual Hypothesis

A challenge to the Structured Individual Hypothesis was presented in the Surprisal Hypothesis (Delogu, 2013), by suggesting that the real-time processing cost observed in complement coercion is not grounded in the semantic representations of coercion verbs themselves, but rather in the fact that entity-denoting complements are merely unexpected after coercion verbs. This hypothesis will be discussed further in the section to follow.

2.7 The Surprisal Hypothesis

A third hypothesis for the processing profile of coercion sentences is surprisal. Surprisal as a linguistic phenomenon is the notion that improbable events carry more information than expected ones, and thus that "surprising" linguistic items will carry a heavier processing cost. Therefore the Surprisal Hypothesis states that difficulty at each newly encountered word should be equal to its level of surprisal (Levy 2008). To apply linguistic surprisal to complement coercion, the hypothesis states that the higher processing cost observed at the entity-denoting complement is a result of an unexpected linguistic item (Delogu, 2013). The processor is expecting some type of event to follow an aspectual verb, which usually modifies a temporal reference (Lai et al., 2014), but when an entity is provided instead, that surprise carries a higher processing cost.

2.7.1 Challenges to the Surprisal Hypothesis

To challenge the surprisal hypothesis, both a lexical decision task and a Cloze test were conducted by DiNardo (2015). These studies will be discussed in detail below. Looking at the Cloze test in particular, this study will demonstrate that subjects are not, in fact, surprised by aspectual verbs combining with entity-denoting complements, as complement coercion sentences with aspectual verbs receive as high average ratings as those with psychological and event-denoting verbs. Moreover, when given the option to write in preferred verbs for complement coercion sentences, aspectual verbs are ranked in the top ten verbs used for sentences with both animate and inanimate subjects. The findings from these studies demonstrate that the processing cost observed in complement coercion can thus be attributed to factors other than linguistic surprisal.

3 The Present Study: Testing the Hypotheses

Given the three hypotheses, their previous evidence, and their challenges, the present study aims to use three different methodologies to arbitrate between them. With this purpose in mind, all three experiments in this study address on two central questions:

1. Which of the three hypotheses provides the most explanatory analysis of the complement coercion effect in both behavioral and real-time data?
2. How does the separation of coercion verbs into two classes, aspectual and psychological,

affect the ERP profile of complement coercion?

3.1 Aims

To answer the above research questions, one can choose to provide evidence either consistent with or inconsistent with the different hypotheses. Study 1 and Study 2 aim to assess the Surprisal Hypothesis by testing the availability of aspectual verbs as compared to psychological verbs in both reaction time and in offline acceptability judgement tasks. Based on the opposing theories, we expect that the evidence will be inconsistent with the Surprisal Hypothesis.

Study 3 aims to provide evidence consistent with the Structured Individual Hypothesis by analyzing complement coercion in two novel ways using ERP. First, the experiment will separate the traditional coercion verb class into two distinct classes, aspectual verbs and *enjoy*-type psychological verbs, with *love*-type psychological verbs as a baseline, similar to Lai et al. (2014). In doing this, Study 3 contrasts with the previous work of Baggio et al. (2010) and Kuperberg et al. (2010), who grouped aspectual and psychological verbs into one coercion condition. Second, the study will look at the ERP components not only at the NP-complement onset, as Baggio et al. (2010) and Kuperberg et al. (2010) did, but also at the verb onset. The motivation to look at the ERP components at the point of the verb onset in particular is to better understand the processing profile of the different verb classes themselves – aspectual, *enjoy*-type psychological, and *love*-type psychological – before the complement has been reached.

4 Study 1: Testing the Surprisal Hypothesis

4.1 Goals

This study was designed as a lexical decision task. The motivation for this task was to control for the possible costs placed on the processor when processing the verbs present in the complement coercion stimuli. If no effect of verb type was observed on response time, then the manifestation of higher costs during complement coercion could be attributed to the coercion construction rather than the verb type itself.

4.2 Methods

4.2.1 Participants

Twenty native speakers of American English between the ages of 18-30 were recruited from the Yale community to participate in the study. All participants were right handed, had no history of neurological disease, brain injury, psychological illness, or learning disability, and had normal or corrected vision by self-report. All participants gave verbal consent to participate in this project, and no personal information was collected or retained about them after the conclusion of the task.

4.2.2 Materials

The materials for this experiment consisted of 21 experimental items and 108 filler items for a total of 129 words. Stimuli were divided into two scripts of 65. All subjects saw both scripts, with half the subjects seeing one script first, and the other half seeing the other script first. Experimental stimuli (aspectual, love, and psychological verbs) were taken from the stimuli used in Katsika et al. (2012) and Lai et al. (2014). These items were chosen because they represent the verbs typically associated with and used in pervious studies on complement coercion (see Katsika et al., 2012 for further explanation). The list of verbs used in each experimental condition is provided below, along with their frequency of occurrence in the study.

Table 1: Study 1 Experimental Items and Distribution

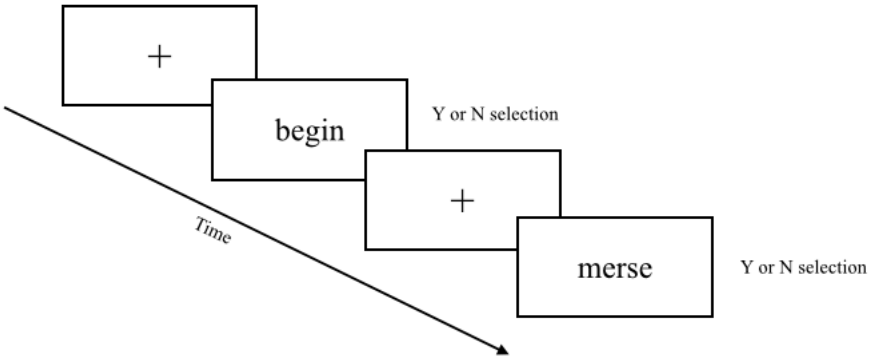
Condition	**Verb type**	**Number of Items**	**Experimental items**
1	Aspectual	5	begin, start, finish, complete, continue
2	"For-ad" fillers	44	–
3	Love	10	detest, love, like, disapprove, dislike, celebrate, approve, respect, to be fond of, hate
4	Non-words	64	–
5	Psychological	6	enjoy, tolerate, prefer, resist, favor, endure

4.2.3 Procedure

Given the goals of the present study, a lexical decision task was selected to measure the processing difficulty of aspectual, psychological, love, and filler verbs. This task was designed and implemented in E-Prime. Two scripts were presented to subjects, each containing 65 pseudo-randomized verbs, controlling for a mixture of all five condition types. Stimuli were presented in white text on a black screen for up to 500 ms at a time. Participants were asked if the string of letters presented on the screen represented a possible word in the English language. Participants were instructed to press *Y* or *N* as soon as they had decided upon an answer.

Before beginning the main task, all participants performed a practice run, consisting of a lexical decision task involving four verbs. After completing the practice task, participants were asked if they understood the task, and the same directions were repeated for the main task. As stated above, in the main task each participant was presented with both scripts in alternating orders, so that every participant saw a total of 129 words. Reaction times from the onset of the word presentation were measured. Reaction times were then used as a measure of the availability of each verb class in the mental lexicon.

Figure 2: Presentation of the Lexical Decision Task



4.3 Results

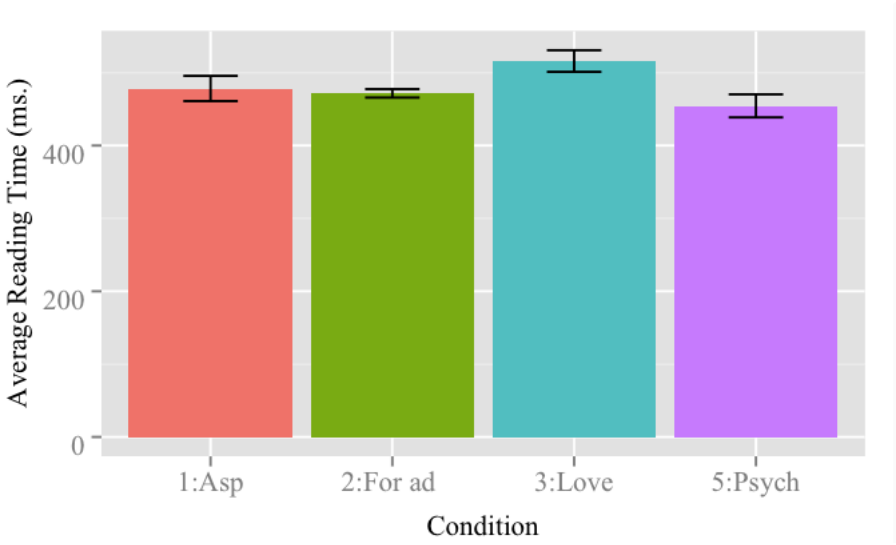
Reaction times were analyzed to determine an average response time to each verb type. Table (2) presents the average response times and standard deviation for all verb types. The key finding from this norming task is that aspectual verbs had similar reaction times (465.77 ms) compared to their psychological counterpart (454.46 ms). Response times to love verbs were higher on average, compared to all other verb types. However, this effect was predicted due to the inclusion of the phrase "to be fond of" within the love verb class. It was expected that this item would taken longer to process due to the fact that it must be syntactically composed, unlike the other single word items presented in the study. Indeed, when "to be fond of" is excluded from the response time data for love verbs, the average response time of this class lowers from 516 ms to 488 ms, thus showing that this item affected the love verbs' higher average response time.

Table 2: Study 1 Average Response Time by Verb Type (in ms)

Condition	**Verb type**	**Average Response Time (ms)**	**Standard Deviation**
1	Aspectual	478.317	189.907
2	"For-ad" fillers	471.639	172.724
3	Love	516	209.665
4	Non-words	496.740	174.833
5	Psychological	454.464	187.291

Response times to Conditions 1, 2, 3, and 5 are shown below in Figure (3). Condition 4 (non-words) was excluded from the pairwise condition comparison due to its high number of items, making it a non-proportional comparison.

Figure 3: Study 1 Average Response Time by Very Type (in ms)



Pairwise comparisons were performed between aspectual and psychological verbs, aspectual and love verbs, love and psychological verbs. This comparison indicated no significant difference between aspectual and psychological verbs ($p = 0.311$), aspectual and love verbs ($p = 0.0997$), however it did indicate a significant difference between love and psychological verbs ($p = 0.00484$). The significant difference observed between love and psychological verbs is unsurprising given the prediction that love verbs would have a higher average response time due to the inclusion of one four word item in that class, as noted above.

Thus, a second pairwise comparison was made excluding the experimental item "to be fond of." This item was removed from the comparison because of the prediction that it would have a higher processing time *a priori*. Results showed that when "to be fond of" was excluded from the comparison, all three conditions now demonstrated no significant difference between them: aspectual versus psychological ($p = 0.311$), aspectual versus love ($p = 0.0637$), love versus psychological ($p = 0.0947$). Since no significant difference was observed by experimental class, these results confidently demonstrate that reaction times were not affected by verb type.

It is important to note that "to be fond of" was not excluded from the stimuli of the second or third study, since it appeared that the only reason its response time was higher was due to the obligatory syntactic composition of the four words, rather than difficulty of processing from a semantic perspective.

4.4 Implications

Results from this first study demonstrated that there is no effect of verb type on reaction time. The findings from this study are important moving forward with the subsequent two studies, since the purpose of the lexical decision task was to control for possible costs placed on the processor due to verb type alone during the Cloze test and the ERP experiment. Since no effect was seen across classes, we were able to move forward with the next two studies using these same aspectual, psychological, and love verbs. Moreover, these results indicated that whatever results would be seen in the Cloze test and ERP experiment could be safely attributed to the complement coercion phenomenon, and not to verb type.

5 Study 2: Testing the Surprisal Hypothesis

5.1 Goals

This study was designed as a Cloze test, using Qualtrics software. The purpose of this particular Cloze study was to test the availability of various subject and verb types within previously used complement coercion stimuli (See 5.2.2. Materials) to eliminate the possibility of surprisal as a processing cost.

5.2 Methods

5.2.1 Participants

Twenty native speakers of American English were recruited from the Yale community to participate in the study (11 males, ages 18-29, mean age 21.4). All participants were right handed, had no history of neurological disease, brain injury, psychological illness, or learning disability, and had normal or corrected vision. All participants gave written informed consent to participate in this project, in accordance with the guidelines set by the Yale University Human Subjects Committee. Each subject was compensated $20 for their participation.

5.2.2 Materials

Fifty experimental sets were created for this experiment, each containing five conditions, for a total of 250 experimental items. These experimental sets were expanded from the complement coercion stimuli used in Lai et al. (2014). The five conditions of test sentences are:

Condition 1: Subject Multiple Choice

E.g. _____*finished the last row of the flower exhibition.*

Condition 2: Verb Multiple Choice with an Animate Subject

E.g. *The celebrated florist* _____ *the last row of the flower exhibition.*

Condition 3: Verb Multiple Choice with an Inanimate Subject

E.g. *The tiger orchid* _____ *the last row of the flower exhibition.*

Condition 4: Fill-in-the-Blank with an Animate Subject

E.g. *The celebrated florist* _____ *the last row of the flower exhibition.*

Condition 5: Fill-in-the-Blank with an Inanimate Subject

E.g. *The tiger orchid* _____ *the last row of the flower exhibition.*

The motive for creating each of these condition types was to manipulate the pairing of a word deletion with either a subject or verb. For the Subject Multiple Choice, the purpose was to test the acceptability of an animate versus an inanimate subject when verb type was manipulated. For the Verb Multiple Choice, the purpose was to test the acceptability of different verb type when subject animacy was manipulated. The purpose of the two Fill-in-the-Blank conditions was to see what verbs were the most readily available for participants, given an animate or an inanimate subject. Sixteen filler sets of sentences from another psycholinguistic experiment by the Yale Language & Brain Lab, each with three conditions, were added for a total of 298 items in the survey. From the total stimuli, twenty unique pseudo-randomized scripts were created.

For multiple choice questions, participants were asked if the options provided for each condition were acceptable on a scale of "'Definitely," "Yes," "Sure," "No," and "Absolutely Not." The options provided for Subject Multiple Choice questions contained an abstract noun, an animate option, an inanimate option, and an unconnected noun. For Verb Multiple Choice (both animate and inanimate), options provided included an aspectual verb, a love verb, a manner of motion verb, a plausible but infrequent verb, and a psychological verb. Subjects were able to rate each option

provided, as demonstrated in the Procedure below. Fill-in-the-Blank conditions were presented with a response box underneath them. Subjects were asked to write the first two verbs that came to their mind for each fill-in-the-blank stimulus. An example of one set of stimuli with their multiple choice options is provided in Table (3).

Table 3: Study 2 Example Set with Experimental Stimuli

Condition	**Sentence**	**Choice 1**	**Choice 2**	**Choice 3**	**Choice 4**	**Choice 5**
1	_____ finished the last row of the flower exhibition.	The tiger orchid	An ear piercing	Modernity	The celebrated florist	–
2	The celebrated florist _____ the last row of the flower exhibition.	enjoyed	detested	finished	poured	perused
3	The tiger orchid _____ the last row of the flower exhibition.	finished	funneled	disrupted	detested	enjoyed
4	The celebrated florist _____ the last row of the flower exhibition.	n/a	n/a	n/a	n/a	n/a
5	The tiger orchid _____ the last row of the flower exhibition.	n/a	n/a	n/a	n/a	n/a

The distribution of experimental items was controlled to ensure equal distribution. The distribution of each verb type in its class is as follows:

- **Aspectual:** begin (33), start (32), finish (33), continue (32), complete (38)
- **Psychological:** enjoy (25), tolerate (22), prefer (22), resist (20), favor (25), endure (3)
- **Love:** detest (9), love (14), like (24), disapprove (11), celebrate (18), approve (26), respect (13), to be fond of (8), hate (6)

5.2.3 Procedure

The present Cloze study was implemented using Qualtrics survey software. Twenty unique scripts were created containing all 300 test sentences. One practice test and one set of instructions were created for all participants. Administration of the Cloze survey took place in the Yale Language & Brain lab. At least one experimenter was present in the room at all times and was monitoring

the progress of the subject. Measures were taken to make sure outside noises were minimized for maximum concentration. The study required approximately 1.5 hours to complete, with a short break given at the halfway point. Each half consisted of 149 stimuli, for a total of 298 stimuli. Six questions were presented on each page of the survey (one of each condition and a control sentence) and subjects were not allowed to advance to the next page until all questions were answered in full.

Participants were instructed to rate each option provided, based on how good it sounded in the sentence. The goal of rating sentences based on how good they sounded was to steer participants away from judging stimuli on how they thought the sentence should be said, and instead judged it on acceptability of conveying meaning. The rating scale was described as follows:

- **Definitely:** I fully understand the sentence and I would say it.
- **Yes:** I can make sense of the sentence and I may or may not say it.
- **Unsure:** It sounds okay but I'm not sure I get a meaning.
- **No:** It looks okay but I am sure I don't get a meaning.
- **Definitely not:** I don't know what you're talking about.

A practice run of non-experimental items was given before the beginning of each testing session to ensure comprehension of the directions. Figure (4) provides an example of one experimental item exactly as seen by participants. Color coding was added to the rating system to make it as comprehensible as possible.[5]

[5]The aim of comprehensibility here was clarity of our directions. We provided visual color coding to our rating system to group the five possible categories into three larger ones – options that sounded good, maybe, or not. We hoped these colors helped participants respond to the stimuli based on immediate response rather than prescriptive grammar.

Figure 4: Study 2 Example Presentation of Stimuli

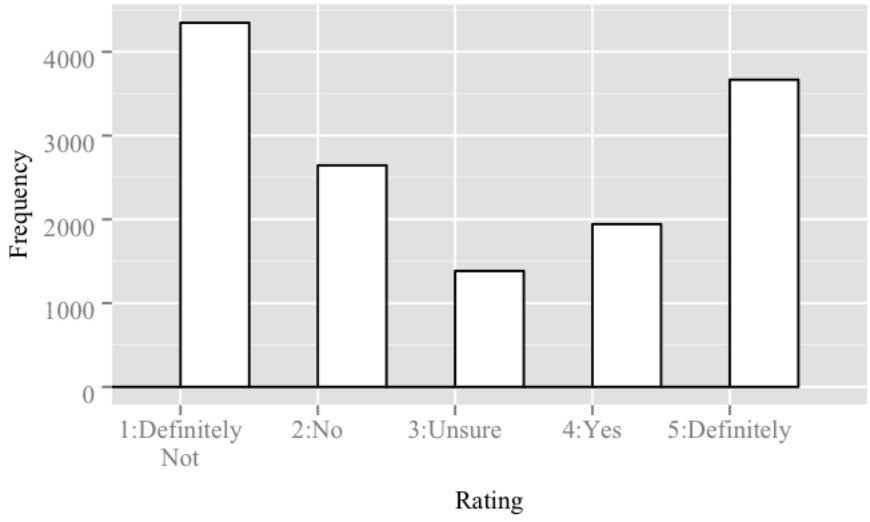
F. Scott Fitzgerald _____ the anthology of 20th century American literature.
(14, 31, 2)

	Definitely	Yes	Unsure	No	Definitely Not
hated	○	○	○	○	○
bought	○	○	○	○	○
completed	○	○	○	○	○
resisted	○	○	○	○	○
shoved	○	○	○	○	○

5.3 Results

To analyze the data of the Cloze survey, ratings of "Definitely," "Yes," "Unsure," "No," and "Definitely not" were converted to a five point numeric scale, in which 5 represents "Definitely" and 1 represents "Definitely not." Average ratings for each condition are provided in the tables and their corresponding figures below. In each condition, certain options were predicted to be floors or ceilings – that is we expected that subjects would reliably either choose the lower or the upper limits of the rating. The total average rating (all subjects by all responses) was 2.853. A distribution of the ratings across all sentence types is provided in the histogram in Figure (5).

Figure 5: Study 2 Frequency of Ratings Across All Conditions

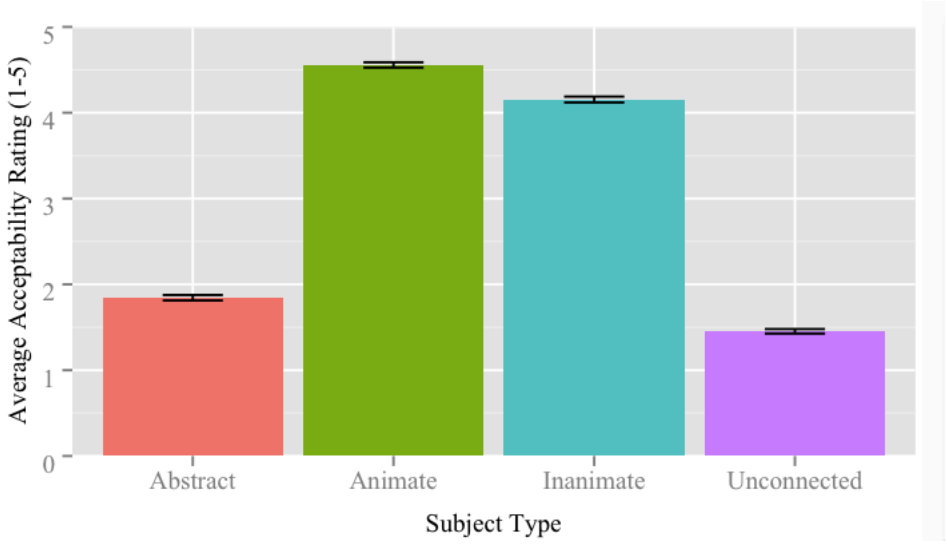


In condition 1 (e.g. _____*completed the anthology of 20th century American literature*), Subject Multiple Choice, the subject types of interest were animate and inanimate. Floor conditions were the abstract and the unconnected options. The average rating across all responses in Subject Multiple Choice was 3 and the median response was 3. As stated above, the critical comparison in the ratings of Subject Multiple Choice were the animate and the inanimate condition. An example of such a comparisons: *Adele started the CD containing 2011's best known pop songs* versus *Rolling in the Deep' started the CD containing 2011's best known pop songs.* Both animate and inanimate subject types were rated at higher than a 4, somewhere between "yes" and "definitely" on the acceptability scale.

Table 4: Study 2 Average Rating by Subject Type in Subject Multiple Choice Condition

Subject Multiple Choice			
Subject Type	**Average Rating**	**Standard Deviation**	**Standard Error**
Abstract	1.844	1.001	0.032
Animate	4.557	0.752	0.031
Inanimate	4.154	1.072	0.034
Unconnected	1.453	0.832	0.026

Figure 6: Study 2 Average Rating by Subject Type in Subject Multiple Choice Condition



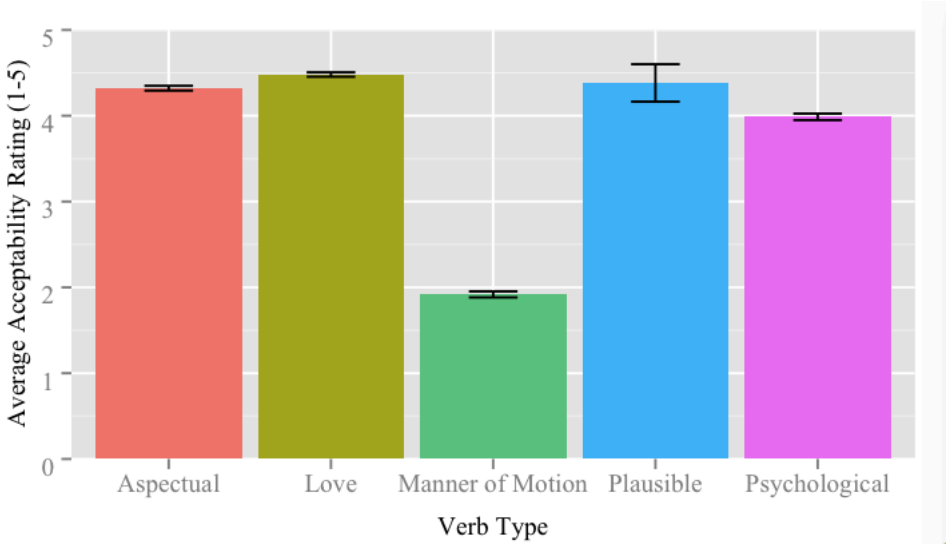
In condition 2 (e.g. *F. Scott Fitzgerald* _____*completed the anthology of 20th century American literature*), Verb Multiple Choice with an Animate Subject, the verb types of interest were the

aspectual, love, and psychological verbs. The manner of motion verb type was expected to be given a floor rating, as it made the sentence nonsensical (e.g. *The Canadian tourist rolled the Freedom Trail in Boston*) and the plausible but less frequent type was expected to be given a ceiling rating, as it was perfectly grammatical and sensical (e.g. *The Canadian tourist inspected the Freedom Trail in Boston.*) The average rating across all responses in Verb Multiple Choice with an Animate Subject was 2.195 and the median response was 2. Interestingly, when an animate subject was paired with possible verb types, the aspectual verbs were given a higher average rating (4.321) than their psychological counterparts (3.987). Love verbs were rated a little bit higher (4.480) than aspectual verb.

Table 5: Study 2 Average Rating by Verb Type in Verb Multiple Choice-Animate Condition

Verb Multiple Choice - Animate			
Verb Type	**Average Rating**	**Standard Deviation**	**Standard Error**
Aspectual	4.321	0.896	0.028
Love	4.480	0.845	0.026
Manner of Motion	1.918	1.121	0.035
Plausible	4.382	1.004	0.218
Psychological	3.987	1.208	0.038

Figure 7: Study 2 Average Rating by Verb Type in Verb Multiple Choice-Animate Condition

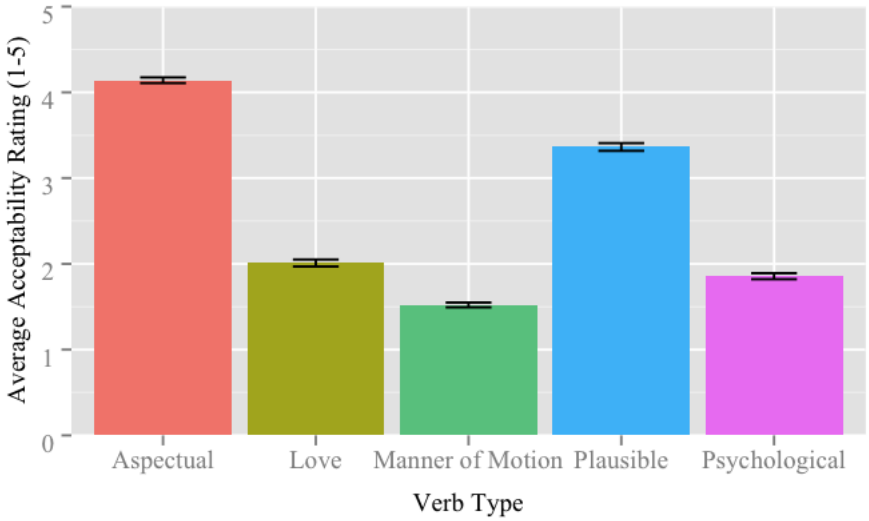


In condition 3 (e.g. *"The Great Gatsby" _____completed the anthology of 20th century American literature*), Verb Multiple Choice with an Inanimate Subject, the verb type of interest was aspectual. Love, psychological, and manner of motion verb types were all expected to be given floor ratings, however verb types were kept the same to maintain continuity with Verb Multiple Choice with an Animate Subject. The average rating across all responses in Verb Multiple Choice with an Inanimate Subject was 3.391 and the median response was 4. In this condition, an important comparison is between the aspectual verb type and the plausible verb type. An example of such a contrast is: *Starry Night began the collection of impressionist oil paintings* versus *Starry Night grounded the collection of impressionist oil paintings.* Interestingly, aspectual verbs (4.141) were rated higher than their plausible counterpart (3.363), implying that aspectual verbs are not only unsurprising in such a context but some times even preferred.

Table 6: Study 2 Average Rating by Verb Type in Verb Multiple Choice-Inanimate Condition

Verb Multiple Choice - Inanimate			
Verb Type	**Average Rating**	**Standard Deviation**	**Standard Error**
Aspectual	4.141	1.030	0.033
Love	2.010	1.277	0.041
Manner of Motion	1.521	0.869	0.028
Plausible	3.363	1.407	0.045
Psychological	1.856	1.109	0.035

Figure 8: Study 2 Average Rating by Verb Type in Verb Multiple Choice-Inanimate Condition



A comparison of conditions 2 and 3, Verb Multiple Choice with an Animate Subject versus Verb Multiple Choice with an Inanimate Subject, demonstrates that aspectual verbs were given a rating of a little over 4, or a little above "Yes," in either context. This finding further indicates that aspectual verbs are not a surprising verb type to follow either agentive or constitutive subjects, as they are rated as acceptable and understandable.

For conditions 4 (e.g. *F. Scott Fitzgerald _____completed the anthology of 20th century American literature*) and 5 (e.g. *"The Great Gatsby" _____completed the anthology of 20th century American literature*), Fill-in-the-Blank responses were tallied by condition across all subjects. A list of the top responses is provided along with their number of occurrences. For the Animate Fill-in-the-Blank condition, the action verbs *wrote* and *read* were the most frequent filler. This finding is not surprising, however, if we take Gricean maxims to be guiding our word choice. In this case, the Maxim of Manner particularly explains our findings by stating that "one tries to be as clear, as brief, and as orderly as one can in what one says, and where one avoids obscurity and ambiguity" (Grice, 1975). An action verb is much more clear than an aspectual verb in any of the experimental stimuli in that it circumvents enriched composition, and thus a more obvious choice for Fill-in-the-Blank.

Of the verb types of interest in this study (aspectual, psychological, and love), love verbs were used most frequently in the Animate Fill-in the-Blank condition than aspectual verbs. However, two aspectual verbs were in the top ten animate responses (*started* and *began*). Aspectual verbs were the most frequent verb type used in the Inanimate Fill-in-the-Blank condition, with six out of the top nine verbs used being aspectual.

Table 7: Study 2 Most Used Verbs for Fill-in-the-Blank Conditions

Fill-in-the-Blank			
Top Animate	Number	**Top Inanimate**	Number
Wrote	256	Began	211
Read	129	Concluded	131
Loved	50	Completed	109
Hated	50	Started	95
Enjoyed	42	Ended	81
Watched	40	Opened	49
Compiled	38	Finished	38
Published	34	Improved	32
Started	32	Enhanced	28
Began	30		
Liked	27		

5.4 Implications

Results from the Cloze survey demonstrated that aspectual verbs are as acceptable or more acceptable than other traditional complement coercion verbs, based on average rating. Crucially, these results present a challenge to the hypothesis that surprisal is the cause for a higher processing cost in complement coercion sentences. The Verb Multiple Choice with an animate or inanimate subject demonstrated that an aspectual verb is rated at or higher than its psychological and love counterparts. Moreover, the Fill-in-the-Blank results demonstrated that even when given the freedom to select the first verb that was available, subjects still preferred aspectual verbs in coercion sentences. Since aspectual verbs were not an unexpected verb type for all subjects, we can hypothesize that surprisal plays no role in complement coercion processing. With these results, we were able to move forward with an ERP experiment, having sufficiently challenged both the Type-Shifting and the Surprisal Hypotheses.

6 Study 3: Testing the Structured Individual Hypothesis using ERP

6.1 Goals

On the basis of the results from Study 1 and Study 2, the third experiment was created. Returning to the overarching goals of the present paper, the purpose of the third study was to help identify the most explanatory hypothesis of complement coercion processing, and to see how the separation of coercion verbs into two classes, aspectual and psychological affected the ERP profile of complement coercion. The third study aimed to answer both of these questions.

To answer the first question, a new approach was taken from the previous two studies. While Study 1 and Study 2 attempted to adjudicate between the three hypotheses by providing evidence inconsistent with the Surprisal Hypothesis, Study 3 aimed provide evidence consistent with the Structured Individual Hypothesis.

To answer the second question, the experimental conditions of Study 3 were created with the intention of separating aspectual verbs from psychological verbs. Doing so allows us to see how the processing profile of complement coercion changes in comparison to the previous ERP work of Baggio et al. (2010) and Kuperberg et al. (2010), who grouped aspectual and psychological verbs

into one coercion condition.

6.2 Why is this study novel?

This study is novel in two particular ways. First, as stated above, it constrains our definition of coercion verbs to include only aspectual verbs, analyzing them in contrast to psychological verbs rather than with. Second, it places a tag at the verb onset as well as the NP-complement onset. This is contrasted with the two previous ERP studies (Baggio et al., 2010; Kuperberg et al., 2010) who only look at the electrophysiological responses at the NP-complement. The motivation for looking at the verb onset is to see if coercion verbs (i.e. aspectual verbs) engender a higher processing cost before even arriving at the complement, based on the predictions of the Structured Individual Hypothesis.

6.3 Predictions

In this study, stimuli were tagged in two locations – the verb onset and the NP-complement onset. To review the findings of Baggio et al. (2010) and Kuperberg et al. (2010), both of these previous ERP studies found an N400 effect at the onset of the NP-complement. In this study, we also predict to observe an N400, however in contrast to the two previous studies, we expect the N400 effect to be observed at the onset of the verb in a complement coercion sentence, as a measure of lexical access and retrieval. A P600 effect is predicted to be observed at the onset of the NP-complement as a measure of ambiguity resolution.

These two predictions are grounded in the Structured Individual Hypothesis – First, a larger N400 effect should be observed at aspectual verbs, as opposed to psychological verbs. This prediction is based on the fact that the recall of an aspectual verb should have a higher processing cost than the psychological or baseline love verbs because all axes over which the aspectual verb can be extended must be recalled as well. Second, a larger P600 component should be observed as well for aspectual verbs at the NP-complement onset. As stated in the "General Methodologies" section, the P600 component has been associated with effects of ambiguity resolution and semantic-syntactic integration within the literature, and thus a larger P600 observed for aspectual verbs would likely reflect the dimension ambiguity resolution once the proper axis over which the verb extends is chosen at the onset of the NP-complement.

6.4 Methods

6.4.1 Participants

Eight native speakers of American English were recruited from the Yale community to participate in the study (5 females, ages 18-23, mean age 21.125). All participants were right handed, had no history of neurological disease, brain injury, psychological illness, or learning disability, and had normal or corrected vision. All participants gave written informed consent to participate in this project, in accordance with the guidelines set by the Yale University Human Subjects Committee. Each subject was compensated $20 for their participation.

6.4.2 Materials

Fifty experimental sets, each with four conditions, were used in this experiment for a total of 200 stimuli. The fifty sets were taken from Lai et al. (2014), and were the same stimuli used in the design of the Cloze study. The four conditions in each experimental set were Aspectual Verb-Agentive Subject, Aspectual Verb-Constitutive Subject, Love Verb (LoveV), and Psychological Verb (PsychV). Thus, in each condition either the subject or the verb was manipulated, however the complement remained the same in all four experimental items of the set. Fifty sets of a filler condition were added, each with two conditions. The total number of stimuli seen by all participants was 300.

Scripts were created in E-prime, with one word placed in each presentation window. A tag was placed at the onset of the verb and the onset of the noun in the NP-complement (after the article). An example of the presentation for all four conditions of one experimental set is provided in Table (8) with the critical words (words 4 and 7) bolded. These critical word windows were epoched together in EEGLAB during data analysis.

Table 8: Study 3 Stimuli Presentation

Fixation	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10	Word 11
+	The	celebrated	florist	**finished**	the	last	**row**	of	the	flower	exhibition
+	The	tiger	orchid	**finished**	the	last	**row**	of	the	flower	exhibition
+	The	celebrated	florist	**hated**	the	last	**row**	of	the	flower	exhibition
+	The	celebrated	florist	**enjoyed**	the	last	**row**	of	the	flower	exhibition

The baseline for this study was the LoveV condition, following the example of Lai et al. (2014). In

this condition, no coercion effect, that is no obligatory eventive interpretation, is seen when a love verb is paired with an animate subject. Because coercion is not obligatory, no extra processing cost should be observed.

6.4.3 Procedure

A standard ERP procedure was followed during testing. Subjects were tested in the EEG lab at the Anlyan Center at the Yale Medical School. A cap with 64 electrodes was was placed on the head of each subject, with four electrodes placed in areas around the eyes and on the earlobes. Conductive gel was applied to each electrode. Ten pseudo-randomized scripts, each containing thirty items, were presented to each participant for a total of 300 stimuli. In each run, stimuli were presented on the screen one word at a time. Each word appeared on the screen for 500 ms. A fixation cross signaled the beginning of each stimulus. A comprehension question followed each stimulus.

Participants were instructed to remain still and not blink while stimuli were presented on the screen. Participants were monitored for accuracy in comprehension questions, as well as for movement.

6.5 Results

Behavioral Results Subjects' answers to the comprehension questions following the stimuli were recorded individually. Of the eight total subjects, all subjects scored above 89% accuracy on this task. The average accuracy of the task was 91.63%. No significant effect of condition or subject was observed on the accuracy ($p > 0.01$), and thus no subject was excluded from the ERP analysis.

Electrophysiological Results Average ERPs were calculated by condition across subjects for each electrode. The averaged epoch extended from 200 ms before constituent onset to 1000 ms after constituent onset. Once the data sets were epoched, contaminated epochs were rejected by calculating statistically abnormal values and improbable data, as well as visual inspection. The corrected results presented were analyzed using MatLab, with a correction based on the control of False Discovery Rate (FDR) ($p<0.05$).

In analyzing this data set, ten total comparisons were made – contrasting both Aspectual-Agentive and Aspectual-Constitutive conditions with LoveV and PsychV conditions at the verb onset and the NP-complement onset. Aspectual-Agentive and Aspectual-Constitutive conditions were also contrasted against one another at the verb onset and the NP-complement onset.

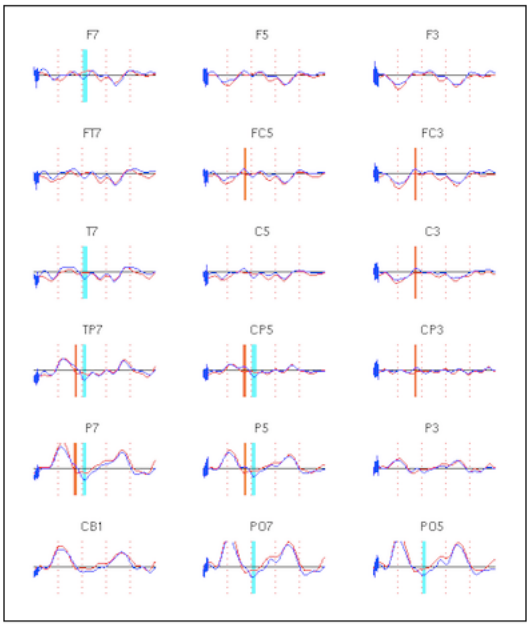
For the purposes of this study, four critical comparisons will be discussed: (1) Aspectual-Agentive versus PsychV at verb onset, (2) Aspectual-Constitutive versus PsychV at verb onset, (3) Aspectual-Agentive versus PsychV at NP-complement onset, and (4) Aspectual-Constitutive versus PsychV at NP-complement onset. These four comparisons were chosen to help directly support one of the three competing hypotheses. Previous ERP work has considered aspectual and psychological verbs to be unified in coercion, and thus in comparing them against one another we are able to measure the results of this study against those in support of the Type-Shifting Hypothesis.

I will also briefly contrast both Aspectual conditions against the LoveV condition at the verb and NP-complement onset. A comparison between aspectual verbs and love verbs is relevant to determine that there is not only a signifiant difference in processing cost observed between aspectual and psychological verbs, but also that the same pattern differentiates aspectual verbs from the baseline of love verbs as well.

First we will look at the effects seen at the verb onset and then we will turn to the NP-complement onset. In all figures provided, red indicates the Aspectual condition (either Agentive or Constitutive) and blue indicates the PsychV condition. It is important to note that all ERP figures provided are only a relevant sample of the electrodes. These electrodes have been selected because they demonstrate the effect described.

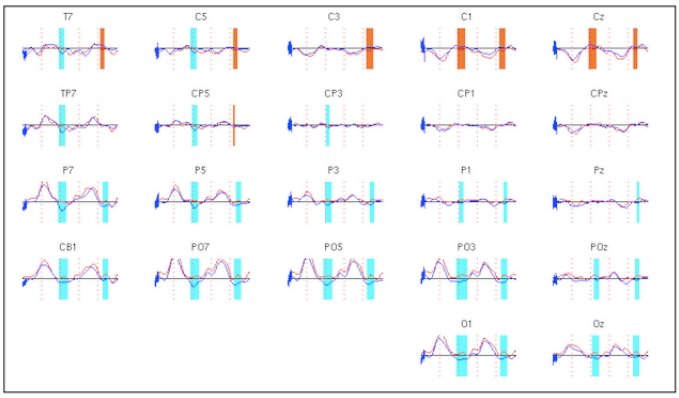
In the Aspectual-Agentive versus PsychV comparison at verb onset (Figure (9)), a significant difference was observed between the Aspectual (red) condition and the PsychV (blue) condition at 400 ms after the onset of the verb, with the Aspectual condition evoking a larger negative waveform than the PsychV condition. This effect was observed in the left posterior regions (T7, TP7, P7, CP5, P5,PO7, PO5) as well as left anterior regions (F7).

Figure 9: Study 3 Aspectual-Agentive vs. Psychological at Verb Onset



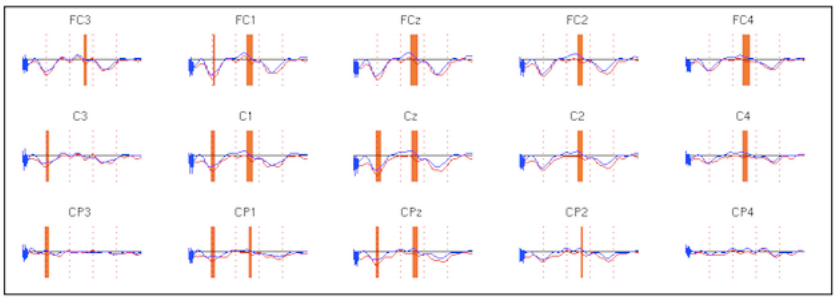
A similar effect was observed in the Aspectual-Constitutive versus PsychVl comparison at the verb onset (Figure (10)). Again, a a significant difference between Aspectual (red) and PsychV (blue) was observed at 400 ms after the onset of the verb, with the Aspectual condition evoking a larger negative waveform than the Psychological condition. Moreover, the effect was once again observed in the left posterior regions (e.g. T7, TP7, P7, CB1, C5, CP5, P5, PO7, CP3, P3, PO5, P1, P03, O1)

Figure 10: Study 3 Aspectual-Constitutive vs. Psychological at Verb Onset



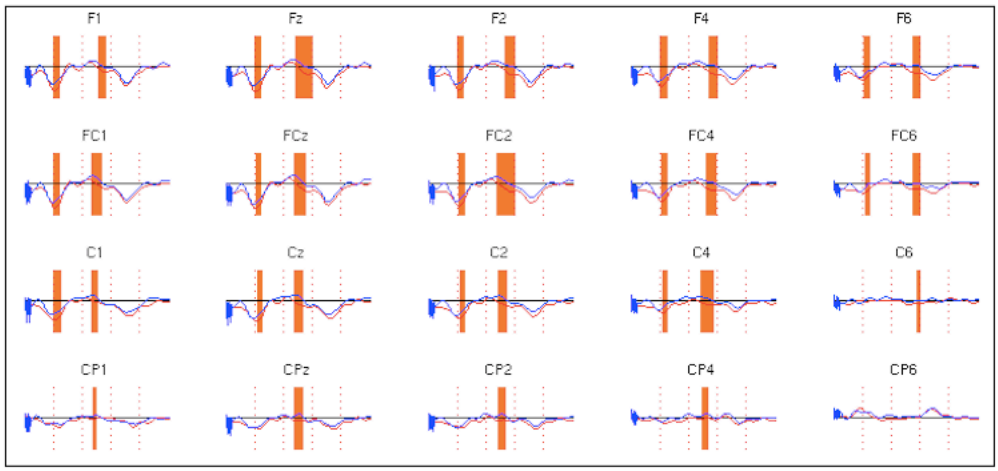
In the Aspectual-Agentive versus PsychV comparison at the NP-complement onset (Figure (11)), a positive waveform was seen around 500 ms in the anterior regions (FC3, FC1, C1, CP1, FCz, Cz, CPz, FC2, C2, P2), with a larger positive effect seen in the Aspectual-Agentive condition as compared to the PsychV condition.

Figure 11: Study 3 Aspectual-Agentive vs. Psychological at NP-Complement Onset



Once again, a similar effect was observed in the comparison of Aspectual-Constitutive versus PsychV conditions at the NP-complement onset (Figure (12)): a positive effect was seen around 500 ms in the anterior regions (F1, FC1, C1, CP1, FCz, Cz, CPz, FC2, C2, CP2), with a larger effect seen in the Aspectual-Constitutive condition as compared to the PsychV condition.

Figure 12: Study 3 Aspectual-Constitutive vs. Psychological at NP-Complement Onset



In the case of Aspectual-Agentive and -Constitutive conditions versus the LoveV condition at both the verb onset and the NP-complement onset, similar patterns were observed to that of both Aspectual conditions versus PsychV. At the verb onset, the Aspectual-Constitutive versus LoveV contrast demonstrated a similar negative wave form at 400 ms after the verb onset, with the aspectual verb condition showing a larger negative amplitude. At the NP-complement onset, a positive wave form was observed 500 ms after the complement onset, with the aspectual verb having a larger amplitude in both the Aspectual-Agentive and Aspectual-Constitutive conditions. These results demonstrate that aspectual verbs not only had a higher processing cost in the predicted directions as compared to psychological verbs, but also higher in comparison to the baseline.

6.6 Discussion

The ERP results demonstrate a higher processing cost of aspectual verbs at both the verb onset and the NP-complement onset. To return to the predictions of Study 3, it was hypothesized that an N400 effect would be observed at the verb onset, and a P600 effect would be observed at the NP-complement, with a larger effect being seen in Aspectual Verb conditions as opposed to Psychological or Love Verb conditions. Results from the eight subjects show exactly that.

As stated above, at the verb onset, a negative going wave form was observed 400 ms after the verb onset in the left posterior electrodes, attributable to the N400 component. Based on the hypothesis that an N400 would be observed as a measure of lexical retrieval of aspectual verbs, it is unsurprising that this effect was seen, since lexical representations are thought to be stored in the lower left areas of the brain (Lau, 2008; Simos et al., 1997). At the NP-complement onset, a positive wave form was observed at 500 ms in the anterior electrodes. We attribute this to the P600 component, since semantic P600 effects have been observed between 500 and 900 ms after critical word onset in anterior regions of the brain (Kaan & Swaab, 2003; Regal et al., 2013). Therefore, in all four contrast, the condition types behaved exactly as we had predicted, with aspectual verbs showing a larger effect than psychological verbs.

The findings from this study not only provide support for the Structured Individual Hypothesis, but also give further evidence to challenge both the Type-Shifting and Surprisal Hypotheses. I will first review the results found in real-time experiments on type-shifting and surprisal and then provide a comparison to the results found in the current study.

In Baggio et al. (2010) and Kuperberg et al (2010), an N400 effect was observed at the NP-complement onset, with coercion sentences showing a larger effect than non-coercion sentences.

As for the Surprisal Hypothesis, Delogu (2013) reports that "relative to predictive controls, coercion expressions initially pattern with equally surprising non-coercion expressions, suggesting that the coercion cost is initially driven by surprisal for the complement noun [...]"

Thus, results from the current study present evidence inconsistent with the findings of all three aforementioned studies. By separating the traditional set of coercion verbs into aspectual and psychological verbs, we found that a P600 effect, rather than an N400 effect or no effect, was observed at the NP-complement. This component is consistent with the Structured Individual Hypothesis in that it is linked to ambiguity resolution. Moreover, the presence of an N400 at the verb onset only further supports the Structured Individual Hypothesis by linking the larger N400 effect observed at the onset of an aspectual verb to the lexical retrieval of the verb and all the possible dimensions along which it can extend.

7 Conclusion

The aim of the present study was to arbitrate between the three possible hypotheses for the observed processing cost in complement coercion – the Type-Shifting Hypothesis, the Structured Individual Hypothesis, and the Surprisal Hypothesis. Recent findings from Katsika et al. (2012) and Lai et al. (2014) challenged the Type-Shifting Hypothesis in two ways: First, by demonstrating that only complement coercion sentences with aspectual verbs displayed the costs typically associated with the phenomenon and second, by providing stimuli not easily explained by type-shifting (e.g. *A long coda finished the music recorded for the 1984 Olympics*).

Having found adequate challenges to the Type-Shifting Hypothesis in previous literature, Studies 1 and 2 of the present investigation aimed to challenge the Surprisal Hypothesis. Study 1 provided evidence inconsistent with surprisal by demonstrating that the observed cost associated with complement coercion cannot be attributed to the processing difficulty of one verb class compared to another. Study 2 provided evidence inconsistent with surprisals by demonstrating that pairings of an aspectual verb with an entity-denoting complement are not rated any lower that of psychological or control verbs in complement coercion sentences. Moreover, when given the option to choose a verb to complete a coercion sentence, aspectual verbs are often one of the first to be chosen – especially when the sentence contains a constitutive subject. Finally, Study 3 provided evidence consistent with the Structured Individual Hypothesis by demonstrating that the ERP components observed in the comprehension of complement coercion sentences can be implicated in the recall

of aspectual verbs and their dimensional axes, as well as in dimension ambiguity resolution at the point of the NP-complement.

In the context of recent literature (Katsika et al., 2012; Lai et al., 2014), this group of three studies readily suggests that the Structured Individual Hypothesis is best explanation for the processing cost observed in complement coercion sentences. Moreover, these studies further demonstrates that the cost typically associated with aspectual and psychological verbs should be attributed only to aspectual verbs within this context. These findings not only refine our understanding of the complement coercion phenomenon, they also provide insight into the real-time integration of syntactic and semantic information in the context of the Parallel Architecture. In particular, the use of ERP in Study 3 provides direct insight into the resolution of mismatches between what is expressed at the syntactic level and what is expressed at the semantic level of sentence structure. The dimension ambiguity resolution observed at the point of the NP-complement in the Structured Individual Hypothesis corroborates the notion that syntax and semantics do not need to be united via a direct one-to-one link, but rather that their integration is dynamic and relies on a complex semantic representation to inform its ultimate interpretation.

8 Acknowledgements

First and foremost, I would like to thank my advisor, Maria Mercedes Piñango, and the co-director of the Yale Language & Brain Lab, Ashwini Deo, for their support and guidance on this project for the past year and a half. Thank you, as well, to Dr. Michiro Negishi for his help in ERP data processing and analysis. The ERP results could not have been included in this study without his generous help.

I would also like to thank the members of the Language & Brain Lab. In particular, I would like to thank Andy Zhang, for his friendship, teaching, and mentorship since I have joined the lab and for answering every one of my questions with patience, as well as Yao-Ying Lai for educating me on all things related to complement coercion. Lastly, I would like to thank all the friends who helped make this project possible, either through participating in my studies or providing support during the writing process. A special thanks goes to Connor Durkin for taking the time to help me in my statistical analyses.

9 References

Baggio, G., Choma ,T., van Lambalgen, M., and Hagoort, P. (2010). Coercion and Compositionality. *Journal of Cognitive Neuroscience*, 2131-2140.

Bornkessel-Schlesewsky, I. and Schlesewsky, M. (2009). *Processing Syntax and Morphology: A Neurocognitive Perspective (Vol. 6)*. Oxford: Oxford University Press.

Brown, C. and Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5(1).34-44.

Delogu, F. (2013). Teasing Apart Coercion and Surprisal: Evidence from ERPs and Eye-movements. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.409.630&rep=rep1&type=pdf.

Chomsky, N. (2006). *Language and Mind.* Cambridge University Press.

Friederici, A. D. and Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. Brain research, 1146, 50-58.

Grice, P.H. (1975). Logic and conversation. In Cole, P. and Morgan, J.L., *Syntax and Semantics III: Speech Acts*. Academic Press, New York, NY, pages 41?58.

Husband, E. M., Kelly, L. A., and Zhu, D. C. (2011). Using Complement Coercion to Understand the Neural basis of Semantic Composition: Evidence from an fMRI Study. *Journal of Cognitive Neuroscience*, 23(11), 3254-3266.

Jackendoff, R. (1997). *The Architecture of the Language Faculty.* Cambridge: MIT Press, 28.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford University Press.

Jackendoff, R. (2010). In Heine, B. and Narrow, H. (Eds.), *The Oxford Handbook of Linguistic Analysis, 645-668.* New York: Oxford University Press.

Kaan, E. and Swaab, T. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Cognitive Neuroscience*, Journal of, 15(1), 98-110.

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

Klein, E., and Sag, I.A. (1985). Type-driven Translation. *Linguistics and Philosophy*, 8(2), 163-201.

Kuperberg, G.R., Choi, A., Cohn, N., Paczynski, M., and Jackendoff, R. (2010). Electrophysiological Correlates of Complement Coercion. *Journal of Cognitive Neuroscience*, 2685-2701.

Lai, Y.Y. (2014). Complement Coercion as the Processing of Aspectual Verbs [PowerPoint Slides].

Lai, Y.Y., Lacadie, C., Constable, T., Deo, A., and Piñango, M.M. (2014). Complement Coercion as the Processing of Aspectual Verbs: Evidence from Self-Paced Reading and fMRI. Cogs 2014 Proceedings, 2525-2530.

Lau, E.F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920-933.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

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

Meyer, D.E. and Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology,* 90(2), 227.

Monsalve, I.F., Frank, S.L., and Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. 398-408.

Oiler, J.W. and Conrad, C.A. (1971). The Cloze technique and ESL proficiency. *Language Learning*, 21(2), 183-194.

Partee, B. (1987). Noun phrase interpretation and typeshifting principles. In Groenendijk, J., de Jongh, D., and Stokhof, M. (Eds.), *Studies in discourse representation theory and the theory of generalized quantifiers.* Dordrecht: Foris.

Partee, B. and Rooth, M.. (1983). Generalized conjunction and type ambiguity. In Bauerle, R., Schwarze, C., and von Stechow, A., *Meaning, use and interpretation of language.* Berlin, Germany: De Gruyter.

Pickering, M. J., McElree, B., and Traxler, M.J. (2005). The difficulty of coercion: A response to de Almeida. *Brain and Language*, 93(1), 1-9.

Piñango, M.M. (2014). Constraints on the language architecture. Talk given to the Yale Cognitive Science Senior Seminar, New Haven.

Piñango, M.M. and Deo, A. (2012). Aspectual verbs and the coercion effect. Handout from SALT

22, University of Chicago, May 18-20 2012.

Piñango, M. M. and Zurif, E. B. (2001). Semantic Operations in Aphasic Comprehension: Implications for the Cortical Organization of Language. *Brain and Language*, 79(2), 297-308.

Pustejovsky, J. (1991). The Generative Lexicon. *Computational linguistics*, 17(4).409-441.

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

Pylkkänen, L. and McElree, B. (2007). An MEG Study of Silent Meaning. *Journal of Cognitive Neuroscience*, 19(11), 1905-1921.

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

Simos, P.G., Basile, L.F., and Papanicolaou, A.C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain research*, 762(1), 29-39.

Taylor, W.L. (1953). "Cloze procedure:" a new tool for measuring readability. *Journalism Quarterly*, 30, 415-433.

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