COGNITIVELY COMPLEX SEMANTIC PROCESSING IN THE DOMAIN OF ASPECT

COGNITIVELY COMPLEX SEMANTIC PROCESSING IN THE DOMAIN OF ASPECT

By, MALAREE BARANIUK, B.A.

A Thesis
Submitted to the Department of Linguistics & Languages
and the School of Graduate Studies
at McMaster University
in Partial Fulfilment of the Requirements
for the Degree of
Master of Science

McMaster University MASTER OF SCIENCE (2014), Hamilton, Ontario

TITLE: Cognitively Complex Semantic Processing in the Domain of Aspect

AUTHOR: Malaree Baraniuk, B.A. (McMaster University)

SUPERVISORS: Dr. Ivona Kučerová, Dr. Elisabet Service

NUMBER OF PAGES: vii, 54

ABSTRACT

Aspect semantically denotes two concepts: eventualities and their endpoints, or the speaker's perspective of an event. These aspectual distinctions respectively refer to telicity and perfectively. Telicity specifically denotes the internal temporal organization of events in regards to their endpoint (telic) or lack thereof (atelic). The telic/atelic distinction has been studied in semantics using logical analysis, in particular mereological relations (the relation of parts to parts and parts within wholes) and scalar theory (a representation of measurement composed of degrees that are partially or totally ordered).

This thesis presents a psycholinguistic experiment investigating the cognitive nature of one component of aspectual realization, namely telicity, in English. In accordance with the predictions made by mereological and scalar theories for the semantic composition of telicity, it is hypothesized that telic constructions are semantically more complex than atelic constructions. This complexity specifically refers to maximalization; a process which Filip (2008) predicts takes place exclusively in the derivation of telic events. This prediction is made under the assumption that a correlation exists between the number of derivational steps and cognitive complexity. Specifically, constructions which involve a greater number of steps in their derivation are hypothesized to also involve a greater amount of cognitive resources in order to be successfully computed. This correlation is supported by recent research investigating the role of working memory in sentence comprehension. Recent studies have found constructions with greater syntactic derivational complexity are susceptible to interference and processing trade-offs when processed under heavy memory loads. Based on this, if telic constructions involve a greater derivational complexity compared to atelic counterparts, they therefore can be hypothesized to be more susceptible to interference and processing trade-offs from increased memory loads. The results of the experiment provide evidence supporting this prediction, ultimately suggesting telicity is a grammatically more complex process than atelicity.

ACKNOWLEDGEMENTS

I owe thanks to several people for their help and collaboration with this thesis. Firstly, I would like to thank my supervisors and mentors Ivona Kučerová and Elisabet Service. Ivona, thank you for conspiring to develop the hypothesis, for all the discussions, suggestions and insights which helped shape this thesis. This project consisted of many details and complexities. Thank you for teaching me the importance of breaking down ideas and thinking about how they work, relate, and come together to form the big picture. Lili, first and foremost, thank you for joining this thesis project. Your expertise in working memory and complex span tasks helped solidify the formulation of this thesis. I truly enjoyed this collaboration, especially the discussions surrounding the interpretations of the results.

Many thanks goes out to Victor Kuperman for his assistance in the statistical analysis of the results. As well, I would like to express my gratitude to Bridget Copley. I am forever thankful for your suggestion to design the stimuli using minimal pairs.

Special thanks to Catherine Anderson for all the academic guidance and support she has provided me with throughout the years. It was during one of our consultations which you suggested I apply to the M.Sc. program. Thank you.

I would also like to thank the members of the *Syntax Lab*, the FLA, and all of my fellow colleges at McMaster University. I truly appreciate all the feedback, and most of all, the support you have given me over these two years.

Lastly, thanks to all my friends and family. There truly are no words which can express my gratitude for your support.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATED SYMBOLS

∃ Existential Quantifier

∀ Universal Quantifier

⊆ Subset

∈ Set Membership

^ Conjunction

|| || The extension of…

{ } The set of

[ ] Indicating definition of…

≤ Less than or equal to

→ Defined as

↔ Equivalent to

¬ Negation

⊕ Sum of

|_| Join Operation

∏ Relation

! Exactly one

Introduction

The observation that verbs appear to denote actions which either tend toward an endpoint (telic) and those which do not (atelic) (Garey, 1957), has inspired much theoretical work in the field of semantics concerning this notion, i.e. verbal aspect. The telic/atelic distinction has been semantically explored on several levels: aspectual categorizations based on the lexical features of verbs (Vendler, 1957; Dotwy, 1979; Parsons, 1990), interval semantics (Bennet & Partee, 1972) and the properties of eventualities (Bach 1981, 1972) including those observed in mass/count distinctions (Link, 1983, 1987; Krifka, 1992), the relation between verbs and their themes (Krifka, 1992; Verkuyl, 1972; Tenny, 1987) and most recently, on the semantic/pragmatic interface using scales (Filip & Rothstein, 2005; Filip, 2008; Rappaport Hovav, 2008; Kennedy & Levin, 2008).

However, although the telic/atelic distinction has been widely investigated on theoretical grounds there exists no experimental research investigating this distinction. This thesis presents a psycholinguistic experiment which investigates the cognitive nature of telicity, i.e. the telic/atelic distinction based on a recent scalar account presented by Filip (2008). Specifically, Filip (2008) argues expressions denoting telic events are derived via the use of a maximalization operator whereas no such operator is used in the derivation of atelic events. As defined by Landman (1992) maximalization involves the comparison of a set of alternatives. This set of alternatives is lexically elicited by items such as numerals (Landman, 1992) or by the verb and its theme (Filip, 2008). Based on this, the following hypothesis can be put forth: as the derivation of telic events (i.e. those denoting an endpoint) involve a maximalization process, their overall computation should predictably be more complex compared to atelic events (i.e. those not denoting endpoints) as they involve a comparison of set alternatives. Furthermore, if it is the case that telic derivations are computationally more complex and computational complexity consumes attentional resources, we should expect such derivations to come at a higher cognitive cost than those which are less complex (i.e. atelic constructions). The Sentence-Span Task is a paradigm which is sensitive to detecting the cognitive costs (e.g. attention) associated with linguistic processing. The working hypothesis was tested using sentence-span task and revealed evidence in favour of a difference in computational cost between the two aspectual derivations. However, this evidence does not directly speak as to whether or not the difference in cost is due to maximalization. In order to better understand if maximalization is the reason for such cost differences, a separate experiment would need to be conducted. This thesis concludes by outlining such an experiment.

Overall, this thesis consists of three main chapters. The first chapter establishes the working hypothesis. Section 1.1 defines telicity, and how it is distinct from another aspectual phenomenon, namely perfectivity. Section 1.2 then presents Filip's (2008) proposal of how telicity is derived via maximalization. Sections 1.3 through 1.6 discusses in detail the theoretical assumptions of Filip's (2008) proposal. Lastly, section 1.7 concludes the first chapter of the thesis by illustrating how the derivation of telic constructions is more complex than that of atelic constructions, thus giving way to the working hypothesis.

With the first chapter having established the working hypothesis, the second chapter of the thesis discusses how to test the working hypothesis. As the thesis investigates a complexity difference between two semantic constructions, a methodology sensitive to processing complexity is required. Complex span tasks are paradigms which have been found sensitive to testing for derivational complexity associated with sentence comprehension. These paradigms are based on several assumptions stemming from the concept of working memory, which section 2.1 discusses. Section 2.2 then discusses in detail, the complex span task used to test the working hypothesis: the reading span task. Following this sections 2.3 and 2.4 converses two lines of reasoning for working memory capacity: decay and interference. Lastly, section 2.5 provides an overview of how the premise of working memory has been used to formulate predictions and interpret results of psycholinguistic experiments investigating the interaction between working memory and derivational complexities involved in sentence processing. Overall, this chapter of the thesis presents the paradigm used to test the working hypothesis as well as puts forth the predictions for the current experiment based on the literature discussed in section 2.6.

The final chapter of this thesis presents the conducted experiment in its entirety in sections 3.1 to 3.3. Section 3.4 presents the results of the experiment, followed by section 3.5 discussing the extent to which the results support the hypothesis. Finally, section 3.6 concludes.

Chapter 1: Establishing the hypothesis

1.1 Telicity vs. Perfectivity

The term *aspect* encompasses the distinction between the imperfective and perfective, or what is known as viewpoint aspect. Defined by Comrie (1976), the distinction between the imperfective and perfective is dependent on the perspective a speaker gives in regards to describing an event. Specifically, the imperfect provides an 'internal view' of an event in the sense that the speaker provides an inside perspective of the situation at the time of speech. Following Klein (1994), the imperfective the topic time of the situation is included in the event time. In contrast the perfective provides an 'external view' of the situation at the time of speech (Comrie, 1976) or where the event time is included in the topic time (Klein, 1994). These semantic distinctions are encoded morpho-syntactically as illustrated by the examples of Modern Greek in (1) and Czech in (2):

1) IMPERFECTIVE	*dhuleva "I worked"/ "I was working"*	
PERFECTIVE	*dhulepsa "I worked"*	(Filip, 2012)
2) IMPERFECTIVE	*dat* "*to give"*	
PERFECTIVE	*davit "to give"/ "to be giving"*	(Filip, 2012)

However, another component of aspect is *telicity*. Telicity is a linguistic feature, which denotes the internal temporal organization of a situation or event. In many respects telicity and perfectivity are perceived as equivalent linguistic phenomena. However, this is not entirely true for several reasons. The purpose of this section is to illustrate the distinction between telicity and perfectivity, as the current paper assumes these two phenomena are discrete.

Giorgi and Pianesi (2001) empirically established and formalized the distinction between telicity and perfectivity using the *in/for x-time* adverbial test. They illustrate how the adverbial test neatly distinguishes between telicity and atelicity in English but proves problematic for imperfectivity and perfectivity in Italian.

Following Giorgi and Pianesi (2001), the notion of *telicity* conveys a *telos* or endpoint as illustrated in (3) and (4):

3) John ate an apple
4) John reached the top

In both cases, not only is it true that the events in question are finished (i.e. the eating of the apple and the reaching of the top) but also that a goal, endpoint or telos has been attained. Specifically, in (1) the telos being that the whole apple was consumed and in (2) John made it to the top. In addition to a telos providing information of the endpoint or goal of an event, it consequently informs us that the event cannot continue further. For example, it would seem odd to continue (1) as *John ate an apple *but hasn't finished it yet*.

However, sentences such as (5) and (6) do not seem to involve any telos and thus reflect *atelicity*:

5) John ate apples
6) John pushed the cart

The crucial difference between (3)-(4) and (5)-(6) is that the events of eating apples and pushing the cart are able to continue on for any given duration. Thus the notion of atelicity does not semantically specify a telos.

Furthermore telic and atelic events can be distinguished using the *in/for x time* adverbial test. Sentences which convey telic events are only able to be modified by *in-x time* adverbial phrases:

7) John ate an apple in/*for ten minuts
8) John reached the top in/*for ten minutes

In contrast, atelic sentences only yield felicitous interpretations when combined with the *for-x time* adverbial phrase:

9) John ate apples *in/for ten minutes
10) John pushed the cart *in/for ten minutes

Interestingly, the adverbial test used to distinguish between telicity and atelicity in English does not produce the same discrete correspondence across other languages or even tenses other than the simple past. Giorgi and Pianesi (2000) illustrate this using examples from

Italian. The imperfect tense when combined with event predicates renders an ambiguous interpretation that can either be perceived as habitual or factual (i.e. continuous):

11) (Alle tre) Mario mangiava una mela
(At three o'clock) Mario ate(IMPF) an apple

Therefore depending on the context, (11) could be interpreted as Mario being involved in an ongoing event of eating an apple (factual) or that on a regular basis Mario eats an apple at three o'clock (habitual). This ambiguity can be resolved by using a when-clause and including a fixed time location to render habitual and factual readings respectively. However, when the imperfective tense is combined with either type of adverbial phrase the factual (continuous) reading becomes unavailable[1]:

12) Mario mangiava (una mela)/ *in/*per un'ora
Mario ate(IMPF) (an apple) in/for an hour

This observation also holds for the present tense which is another imperfective tense in Italian:

13) Mario mangia (una mela) (*in/ *per un'ora)
Mario eats (an apple) in/for an hour

Therefore examples such as (12) and (13) illustrate that the telicity distinction does not apply to imperfective predicates. However, Giorgi and Pianesi (2000) suggest that this is too extreme of a generalization. Specifically, they argue that a morphological counterpart associated with perfective verb forms enforces *terminativity* but the morphology associated with the imperfective does not impose such a requirement. Thus perfective verbal forms refer to events as terminated in (13) whereas imperfective verbal forms may refer to events as terminated or non-terminated, as seen in (14) and (15a,b):

14) ? Questa mattina Mario ha spinto il carretto, e lo sta spingendo tutt'ora
This morning Mario pushed(PERF) the cart, and he is still pushing it
PERF≠ continual

15a) (Alle tre) Mario mangiava una mela (e la sta mangiando tutt'ora).
(At three) Mario ate(IMPF) an apple (and he is still eating it)
IMPF=continual

15b) Tre ore fa, Messner raggiungeva la vetta (*e la sta ancora raggiungendo).
Three hours ago, Messner reached(IMPF) the top (*and he is still reaching it)
IMPF≠ continual

[1] Giorgi and Pianesi note that "…depending on the actional nature of the verbal predicate, the habitual reading might still be there" (p.g. 20).

These examples thus illustrate the non-committal nature of imperfective verb forms. Thus the morphemes encoding aspect here are neutral in the sense of their semantic interpretation being imperfect. Now consider the following example:

16) Ieri Gianni raggiungeva la vetta in tre ore
Yesterday Gianni reached(IMPF) the top in three hours
IMP≠ continual

Although the imperfective is present, the *in-x time* adverbial is allowed and thus illustrates the imperfective is compatible with terminative readings. Therefore, when the imperfective provides a terminative reading, only then is it able to be modified by the *in-x time* adverbial.

The examples presented by Giorgi and Pianesi (2001) are important for the current experiment for the following reasons. Firstly, Giorgi and Pianesi (2001) illustrate how the adverbial test detects telicity, and not necessarily perfectivity. Traditionally, the adverbial test was used to distinguish between the semantic readings associated with the imperfective and perfective. However, as made evident by Giorgi and Pianesi (2000) the examples from Italian confound the clear divide the adverbial test is assumed to make. Instead, the *in/for-x time* adverbial test demonstrates that English semantically conveys telicity while Italian morpho-syntactically conveys perfectivity. The test provides explicit telic and atelic readings for English, but proves problematic for Italian due to the imperfect being incompatible with either adverbial phrase. Based on this evidence, it is clear that not only are telicity and perfectivity separate aspectual concepts, but the manner in which these concepts are encoded is also unique. As the current experiment is an investigation of telicity, it assumes the distinction asserted by Giorgi and Pianesi (2001). In addition, as the adverbial test renders absolute telic and atelic readings for English, it serves as a crucial component in the compositional of the stimuli for the present experiment[2]; a point which is discussed in detail in section 3.2.1.

Secondly, although Giorgi and Pianesi (2001) propose an account for the telic/atelic distinction using the notion of terminativity, their theory is not adopted by the current paper. The reason being is that the current paper investigates the difference in derivational complexity hypothesized to exist between telic and atelic constructions. Giorgi and Panesi's (2001) account provides no such grounds to establish a hypothesis based on complexity. Instead, the theory of telicity this paper assumes is that proposed by Filip (2008) presented in the next section.

1.2 Telicity via Maximalization

With the concept of telicity established by Giorgi and Pianesi's (2000), we now turn to a theory which predicts *how* an endpoint or telos is reached. The current study formalizes its hypothesis based on the approach presented by Filip (2008).

[2] I would like to thank Bridget Copley for her helpful insight and proposal to use minimal pairs based on the adverbial test for the stimuli.

According to Filip (2008) telicity is derived via a maximalization operator. This operator was originally proposed by Filip and Rothstein (2006) and is defined in (17) below:

17) Telicity corresponds to the maximalization operator MAX_E. It is a monadic operator, such that

MAX_E (∑), which maps sets of partially ordered events ∑ onto sets of maximal events MAX_E(∑)

Thus the manner in which telic events are rendered, according to (17), is a MAX_E operator applies to a set of a sum of events and maps this set of events (which are partially ordered) onto a set which reflects a maximal event. The maximal event here is similar to a telos, as it, in a sense, reflects a point of discontinuation or finality. However, the derivation of telicity proposed in (17) is based on several theoretical assumptions: maximalization, event semantics, grammar of measurement and scalar semantics. The next section discusses each of these assumptions in turn to provide a full understanding of how telicity is derived via maximalization.

1.3 Maximalization as defined by Landman (1998)

The first assumption Filips's (2008) proposal makes is that telicity is derived via a maximalization operator. The notion of maximalization she adopts for telicity is based on the definition developed by Landman (1998). This section discusses in detail Landman's (1998) maximalization and concludes by illustrating how the premises of his account transfer to the operator proposed by Filip (2008).

Landman's (1998) proposal for maximalization extends the basic account provided by Scha (1981) regarding cumulative readings. For example, (18) below possesses three readings: cumulative, plural and exactly.

18) Three boys invited four girls

According to Scha (1981) the cumulative reading of (18) is that the total number of boys that did the inviting is 3 and the total number of girls who were invited were 4. This reading is based on a semantic binary operation between the determiners "three…four", the nouns "boys…girls" and the verb "invited". In addition, Scha (1981) also argues that (18) entails an exactly reading: Not more than three boys invited a girl, and not more than four girls were invited by a boy.

Landman (1995) argues against this in two respects: first, the cumulative reading and account provided by Scha (1981) can be instead determined by semantic plurality and second, the exactly reading Scha (1981) assumes is entailed by (18) is incorrect. According to Landman (1995), the cumulative reading of (18) results from the semantic composition formally noted in (19):

19) CUM. = $\exists x \in$ *INVITE: $\exists x \in$ *BOY: $|x| = 3 \wedge$ *Ag(e) = x $\wedge \exists y \in$ *GIRL: $|y| = 4 \wedge$ *Th(e) = y

Thus, in the case of *Three boys invited four girls* there exists a set of inviting events which includes all singular *and* plural sums of a set of boys and a set of girls, which are agent and theme respectively. Where each of the four boys invites at least one of the three girls, and where each of the three girls is invited by at least one of the four boys. Thus the cumulative readings of both Scha (1981) and Landman (1995) are derived via the semantic composition of the sentence.

However, Landman (1998) argues there is nothing within the semantics of (18) that is able to denote an exactly reading: *exactly 3 boys invited exactly 4 girls*. If the exactly reading was derived from semantic composition then according to standard tests, it would not be cancelable. Interestingly, this does not hold as Landman (1998) illustrates in example (7), here re-stated in (20):

20) In the shyness-therapy group the therapist has set an award for each group where three boys dare to invite four girls. We go up to the therapist and say: "We claim the award, because in our group three boys invited four girls. In fact," we proudly add, "in our group, three boys invited six girls."

As (20) shows, the exactly reading is not part of the semantic meaning of the sentence as it can be cancelled with the "in fact" addition. Thus, the exactly reading is only becomes available through the addition of context.

Based on this, Landman (1998) considers that it is possible for the exactly reading to come in through the Gricean maxim of quantity: if the speaker knew there were more boys or girls then he/she should have said so, and since this is not the case the exactly reading arises due to a conversational implicature. However, Landman (1998) does not follow this line of reasoning and instead argues the exactly reading becomes available via a scalar implicature (Horn, 1972; Fauconnier, 1975).

Assuming that numerals, such as those present in (18), are lexically associated with a scale, when these numerals are part of an argument NP (Kadmon, 1987), they induce a scale of numeral propositions:

.
.
n. there were n boys at the party
3. There were 3 boys at the party
2. There were 2 boys at the party
.
.

The exactly implicature is thus calculated using the above scale to select the highest level which is true. Crucially, the implicature is not directly calculated via the scale but rather through the entailments this scale induces. To clarify this, consider the chain of logic progressing as such: *lexical item* → *numeric scale* → *entailment scale* → *scalar* *implicature*. Therefore Landman (1998) further assumes in order for this to be so, the grammar involves a mechanism which builds a scale of propositions from the scale which is lexically induced.

This latest assumption then leads to the question: if scales are built by the grammar does this mean scalar implicatures are as well? Landman (1998) argues, in line with Gazdar (1979) and Kadmon (1987), that scalar implicatures are calculable and done so via grammar. Specifically, Landman (1998) argues that an implicature is calculated not at the lexical level (as this is where the scale is calculated) but from the meaning of the sentence at particular scope domains:

21) *Implicature Construction Principle (ICP)*

> The exactly-implicature of a numerical noun phrase is calculated at the scope domain that the noun phrase is in, and inherited from there on, built up following the semantic composition of the sentence. It becomes a scalar implicature at the sentence level if it is neither incompatible with nor entailed by the semantic meaning (Landman, pg 244).

In addition to where the implicature is calculated, Landman (1998) asserts that the scalar implicature is calculated by a maximal operation. Following the assumption that numerals induce a numeric scale, it also follows that this numeric scale is associated with a pragmatic scale. This pragmatic scale is a set of ordered meanings based on the numbers, and this set enters into the maximalization calculation as a *comparative*. That is, the maximalization is carried out in the sense of “no number higher on the scale than me” which provides the exactly or maximal reading *Exactly 3 boys invited exactly 4 girls* (Landman, p. 247). For example, the implicature is calculated by the comparison of “four girls being invited vs. three girls being invited etc.” and “three boys inviting vs. two boys inviting etc.”. As this scale is built on the meaning of a numeral and the verb, this is considered to be a complex scale.

Landman (1998) follows Rooth (1985) and assumes complex scales are built by a *p*-mechanism. Where a p-set is a set of alternative meanings of a similar type and are associated with the meaning of an expression. These alternatives are built at a semantically or pragmatically later stage of the derivation which is why they are considered complex, and are unordered. Based on this, Landman (1998) assumes we associate the numerical NP with a scale of alternatives. Rooth’s p-mechanism builds this scale based on the semantic composition until a scope domain is reached. At this point, under a Davidsonian theory existential closure takes place, however Landman instead replaces this operation with an operation of *maximalization* on scope domain. Thus the maximalizing is relative to the scale of alternatives and is what calculates the core implicature (i.e. exactly reading).

To summarize, Landman (1998) proposes that the exactly reading associated with the example *Three boys invited four girls* is not provided by the semantic meaning of the sentence, contrary to Scha (1981). Instead this reading is derived via a scalar implicature which is rendered possible through a maximalization operation and scalar semantics.

Based on this, Filip’s proposal of maximalization for telicity follows Landman’s (1998) account in the following respects. Firstly, if telicity is derived via maximalization there needs to be a lexical item within the derivation which induces a scale of propositions, with meanings similar to the original expression. This assumption is discussed in detail in the next section. Secondly, this numeric scale must also have an associated pragmatic scale which contains a set of alternatives that the maximalization operator compares and selects from. As we will also see in the next section, these alternatives are not based upon meanings of a nominal domain, but rather a verbal domain. Lastly, in order for maximalization to be applied to telicity Filip (2008) also assumes, like Landman (1998), that there is a mechanism in the grammar responsible for the construction of scales. This point will also be discussed in more detail in section 1.5.

Landman's (1998) proposal is relevant to the current paper because it is one of the many assumptions Filip's (2008) definition for the derivation of telicity is based on. Crucially, it offers the *source of complexity* which the hypothesis is based upon: the derivation of semantic constructions which denote definite endpoints (i.e. telic) are more *complex* compared to constructions which do not denote endpoints (atelic). Specifically, if Filip (2008) assumes telicity is derived via a maximalization operator, then it follows that her operator (MAX_E) computes a comparison between a set of alternative meanings in order to return the unique largest event. This largest event is the maximalization equivalent of a telos in the sense of Giorgi and Pianesi (2000), and is therefore a feature associated exclusively with telic events. Filip (2008) states that MAX_E is an operator present only in the derivation of telic events, and so no such comparison process is used to derive atelic events. Therefore, due to the fact that the MAX_E operator is absent for the derivation of atelic events and no comparison process takes place, these constructions are less complex compared to their telic counterparts.

The next sections discuss in detail, the remaining assumptions of Filip (2008)'s proposal. Specifically, section 1.4 addresses the alternatives that are compared by Filip's maximal operator: event stages. Event stages are semantically composed in a similar fashion as mass/count nouns. This similarity is illustrated by comparing Link's (1983) proposal for mass/count nouns to Krifka's (1992) proposal for telic and atelic events. Following this section, I will discuss how a scale is formed, as maximalization is an operation which applies to scales. This will be illustrated again by Krifka (1992) and his account of which lexical items constitute a scale reflecting strictly incremental relations.

1.4 Event Structures

This section discusses the second assumption Filip's (2008) proposal for telicity via maximalization makes: that telic and atelic events are semantically composed using event semantics. This section starts by establishing the semantic distinction between mass and count nouns proposed by Link (1983). The reason for this is that the composition of events has several striking parallels to the composition of mass and count nouns. These parallels are emphasized and discussed further in section 1.4.2 where Krifka's (1992) framework for telic and atelic events is presented. Together these two sections will provide the foundation for which the maximalization operator MAX_E can apply. Specifically, section 1.4.2 provides the alternatives which the operator compares and section 1.5 illustrates how the scale to which these alternatives reside on is lexically induced.

1.4.1 Link (1983)

An interesting parallel exists between the semantic composition of mass/count nouns and telic/atelic events. Namely, the semantic structuring of eventualities and mass/count nouns both make use of the *part-of* binary relation "≤" introduced by Sharvy (1980) and is similar to the account made by Scha (1981) for cumulative readings. Link (1983) illustrates how the part-of relation is the essential semantic component which distinguishes the difference between mass and count nouns. To illustrate, consider the following examples from Link (1983):

22) The cards
23) The deck of cards
24) The water gathers in pools

Arguably, (22) and (23) can both reflect collections of cards, and thus be referred to as collective plurals. In each case both collections consist of the same portions of matter – cards. However, (22) can also reflect a pure plural term, as when compared to (23) *the cards* appears to express individual portions rather than a cumulative entity or just one individual. Example (24) is also a collective predicate, but is distinct from (23) as it is a mass term. Unlike mass terms, the individual entities which comprise the set of a collective plural can be singled out (e.g. the Queen of Hearts from the deck). What entity then can be singled out from water? Unless a quantifier is used (pool/drop of water), mass nouns appear to denote something quite different than other collective predicates.

Link argues the difference between examples (22-24) can be captured using mereological relations or lattice structures. Mereological structures are essentially homomorphic mappings of an object and a specified property of that object. For example, in (22) the mapping would include how the individual cards of a deck comprised the deck of cards. Formally, Link (1983) accounts for the different mappings between examples (22) and (23) using the following variables and operators.

Let *P* denote a 1-place predicate and the operator (*) generate the individual sums of members of an extension of *P*. Therefore **P* reflects the same cumulative property as a mass noun. Furthermore, *E* represents the domain of the discourse which here concerns predicates of individuals (*P* and **P*).

The difference between predicate types is their mapping relation with the set of atoms *A* residing in *E*. Set *A* reflects singular objects (e.g. a card) and follows that there are multiple atoms within the set such as atoms *a* and *b* in set *A*. The relation between atoms can be either singular fusion (a + b) or individual sum/plural (a ⊕ b).

The sum relations are subject to two types of ordering relations. The first is an intrinsic ordering called the individual part relation. This relation is denoted via "$\leq_i$" which is expressed using a 2-place predicate " ∏" and is defined below in (25):

25) *Individual Part Relation*
$a \prod b \leftrightarrow a \oplus b = b$

Semantic Equivalent: $||a|| \leq_i ||b||$ iff $||a|| \sqcup_i ||b|| = ||b||$

Essentially (25) depicts the individual parts which comprise a set of the predicate. For example, *Tom and Dick carried the piano* (Link, 1983) where there exists an individual *a* (Tom) and an individual *b* (Dick) such that both individuals, when joined through the join operation of individuals create a sum of the individuals (the set) who carried the piano. The individual part relation captures *P* with the corresponding extension **P* plural predicate.

Now consider $^{*}P$ which is a non-atomic predicate extension of **P*, or the proper plural predicate of *P*. Thus we further distinguish our mapping relations between sums and proper sums, respectively defined in (26) and (27) below.

26) *Plural Predicate*

$||{*P}|| := \{ x \in E \mid \exists X \subseteq ||P|| \wedge X \neq 0 \text{ such that } x := \sup_i X\}$

27) *Proper Plural Predicate*

$||^{*}P|| := ||{*P}|| \setminus A$

Where a plural predication is the set with a variable *x* which is within the domain of individuals or there exists a set *X* which is included in the denotation of *P* and set *X* is not empty such that the variable *x* is defined as the supremum of the set *X* ($\sup_i X$). In other words, the mapping here reflects a predicate where all individuals residing in the set reflects a cumulative reading (e.g. a deck of cards).

In contrast, a proper plural predicate essentially is an atomic subset of **P*. Thus the mapping reflects one individual from the set being represented rather than the collective unit of all entities (e.g. a card).

We have now discussed the distinguishing property between pure and collective plural terms; the mapping relations of individual parts. Let us now turn to mass terms, which introduces the second type of ordering between sum relations.

The material part relation, denoted by "T" is also intrinsic but unlike the individual part relation, it reflects the portions of matter constituting the individuals themselves. Ultimately this mapping provides us with the lattice structure for mass nouns.

28) *Material Part Relation*

$a \prod b \rightarrow aTb$

$x \leq y$ iff $x \sqcup y = y$

In addition, *D* denotes a set joined with *E* through the joining operation to create a semi-lattice between *D* and *E*. Therefore *D* represents the set of individual portions of matter, which is a subset of *A*, the set of atoms within *E*. Link then posits every model consists of a semilattice homomorphism (*h*) from $E\setminus \{0\}$ to *D*. If *P* is a mass term it is denoted as in (29).

29) *Mass Predicate*

$||^{m}P|| := \{ x \in D \mid x \leq \sup h[||P||]\}$

$^{m}Px \rightarrow \vee y\ (^{*}Py \wedge xTy)$

Where a mass term is defined as the set of *x* within the set of *D* or *x* is in a partial relation with the supremum ordering of the extension *P*. To illustrate this, again consider (24) restated here as (30).

30) The water gathered in pools

The mapping of water is thus the material part relation of individual portions of the 'water-matter' which is denoted as a collective indivisible unit. However, each pool of water can reflect the individual parts/atomic elements (each pool of water) within the set of *E* (the set of all gathered pools of water).

As illustrated in the above framework, it is the part relations themselves which create a set reflecting of plural/sum or singular/atomic predicates. With this basic outline in place of how mereological mappings construct sets for mass/count nouns, let us now turn to event structures and see how they are constructed using a similar framework.

1.4.2 Krifka (1992): Sums and Parts of Events

Several interesting semantic similarities can be drawn between nominal references such as the ones discussed above and temporal structures. As noted by Krifka (1992) there exists a structural analogy between nominal references and events: count nouns ~ telic verbs and mass ~ atelic verbs.

The framework proposed by Krifka (1992) illustrates these approximations by assuming the following variables and operators: *O* denoting objects, events are represented by *E* and times characterized by *T*. All three variables extend to complete join semi-lattice structures via the same join operator used in Link (1983). With this in place, the following relations can be made to reflect the various event types. As stated below in (31), Krifka defines telic predicates to hold a *quantization* property which mirrors the properties of (25) the Individual Part Relation above.

31) $\forall P\ [QUA(P) \leftrightarrow \forall x,y\ [P(x) \wedge P(y) \rightarrow \neg y\ [x]]]$

A nominal (*x*) or verbal (*e*) predicate *P* is quantized (QUA) iff some *x* or *e* falls under *P* as a non-proper part. Quantized predicates, whether nominal or verbal, cannot be a proper part *x'* or *e'* under *P*.

Thus quantized NPs denote objects with precise limits (e.g. an apple) and coincide with telic events which also denote limitations (e.g. arrive). In addition, Krifka assumes events are structured via an atomic time lattice (T_a), a temporal ordering relation $\leq$, the temporal trace function τ which is the extension from *E* to *T*, and finally *TP* which maps an event to the last time point in its run time.

Telic events predicates are characterized by possessing a *set terminal point* (STP) whereas atelic events lack this property.

32a) *Terminal Point*

$\forall e,t\ [TP(e) = t \leftrightarrow T_a(t) \wedge t \subseteq \tau(e) \wedge \forall t'\ [\ t' \subseteq \tau(e) \rightarrow t' \leq t]]$

Where there exists an event (e) and time (t) in which the terminal point (TP) of the event is equal to the time defined by the atomic time lattice (T_a) and time which includes the run time of the event and a proper time interval (t'). Where t' is included in the run time of the event in which the proper time is less than or equal to t

32b) *Telic Event Predicates*

$\forall P\ [STP(P) \leftrightarrow \forall e[P(e) \rightarrow \forall e'[P(e') \wedge e' \subseteq e \rightarrow TP(e) = TP(e')]]]$

Where there exists a predicate (P) defined by a set terminal point (STP) of (P). The set terminal point of P is equivalent to an event (e) which is defined by P(e). P(e) is defined by a proper event (e') that includes P(e') and a proper event included in an event (e) which is defined by the terminal point (TP) of the event (e) equaling the terminal point of the proper event (e').

Thus (32a&b) formalize the semantics of telic events where STP generates all sub-events of an event to have the same terminal point.

In turn atelic predicates possess a *cumulative* property defined in (33), which mirrors the properties of mass nouns defined above in (26) and (27).

33) $\forall P\ [CUM(P) \leftrightarrow \forall x,y\ [P(x) \wedge P(y) \rightarrow P\ (x \sqcup y)]]$

A nominal (*x*) or verbal (*e*) predicate *P* is cumulative (CUM) iff some *x* and *y*, or *e* and *e'* falls under *P* as a proper part. Cumulative predicates, whether nominal or verbal, reflect the mereological sum ⊕ of their respective proper parts.

It follows then if the denotation of cumulative NPs and atelic verbs establish no unique individual or endpoint respectively (e.g. run, wine) then atelic events are simply the negated counterpart of telic events.

34) *Atelic Event Predicates*

$\forall P\ [CUM(P) \wedge NEG(P) \rightarrow \neg STP(P)]$

There exists a predicate (P) which is defined by the cumulative predicate P and the negated form of P which is defined by the negation of the set terminal point of P

To summarize, Krifka (1992) illustrates that semantic distinction between telic and atelic events is that the former is quantized (i.e. an event that is not reflected by any of its proper parts) and the latter is cumulative (i.e. an event is reflected by its proper parts or by the sum of its proper parts). This therefore establishes that telic events are to be understood as those with an interval which reaches an endpoint and atelic events to be understood as those whose interval do not reach an endpoint.

The *part-of* relation presented by Krifka (1992) creates a homomorphic mapping between the event and their intervals, which is ultimately a set. As mentioned, maximalization is an operation which compares a set of alternative meanings to return the most maximal (Landman, 1998) or largest unique event (Filip, 2008). Thus the mapping distinctions Krifka (1992) makes can arguably be viewed as sets of event intervals (i.e. various progressions of the event) that are partially ordered. In turn, this partial ordering can arguably be viewed as a scale, which is what makes Krifka's (1992) proposal of events capable of undergoing the process of maximalization.

Essentially, so far Krifka's (1992) account illustrates to some degree, *what* the alternatives are that Filip's (2008) MAX_E operator compares. However, as mentioned in Landman (1998) these alternatives reside on a scale, a scale which is lexically induced. How then, is a scale lexically induced under Filip's (2008) proposal? The answer to this is presented in the next two sections, where I first establish the characterization of a scale Filip (2008) assumes and then go on to illustrate how a scale is induced using Krifka's (1992) *Strictly Incremental Relations*.

1.5 Scales as characterized by Kennedy & McNally (2007)

Kennedy and McNally (2007) provide a semantic typology of gradable predicates, namely deverbal adjectives. In their account they argue that the derivation and interpretation of degree modifiers is dependent to two parameters: the gradable predicate's association with an open or closed scale and standard of comparison applied to the predicate. The former parameter is relevant for the current paper as it is Kennedy and McNally's (2007) characterization of a scale which Filip (2008) assumes.

Kennedy and McNally (2007) assert that gradable adjectives map their arguments onto abstract representations of measurement. This representation is comprised of the following parameters which ultimately devise a scale:

35) A scale reflects:

i) *a set of degrees* (measurement values) *totally ordered* with respect to some

ii) *dimension*, which indicates the property being measured (volume, temperature, length, width, loudness, intensity, etc.); and

iii) *an ordering relation* on the set of degrees, which distinguishes between predicates that describe increasing properties (e.g. tall) and those that describe decreasing properties (e.g. short)

For example, the gradable adjective *expensive* denotes a relation between the degrees of cost of an object *x* that is considered to be expensive. Therefore the adjective denotes a set of degrees that are ordered (e.g. increases of cost) with respect to a dimension (e.g. dollars). Where the default ordering relation of the scale is '>' or greater than and is antisymmetric and transitive. Thus the ordering relation of costs associated with *expensive* progress along the scale in increasing increments.

Filip's (2008) proposal for telicity assumes the characterization of scales provided by Kennedy and McNally (2007). However, Filip's (2008) maximalization operator cannot apply directly to a scale of objects but a scale which measures quantities. Therefore the operator cannot simply apply to a set of ordered event intervals but a set which contains an ordering of the progression of intervals with respect to a dimension (i.e. events that evolve into more developed versions). How is a scale with this criterion semantically determined?

The answer comes from an observation made by Krifka (1992). Example (36) illustrates how the event of eating an apple denotes a quantized or telic event as the event's telos is not able to be modified by the durative *for x time*. Likewise the event of eating apples in (37) does not denote such an endpoint as it is only compatible with the durative construction.

36) John ate an apple in five minutes/ * for five minutes
37) John ate apples for five minutes/ *in five minutes

Interestingly, this observation does not hold for all verbs, as is born out in (38) and (39).

38) Peter saw a cardinal for a few minutes/ *in a few minutes
39) Peter saw cardinals for a few minutes/ *in a few minutes

This suggests the lexical semantics of the verb as well as its object (henceforth called the theme) play a crucial role in the event's temporal composition. The event structures under Krifka's (1992) mereological approach are sensitive to the lexical semantics of the verb *and* its correspondence with its thematic predicate. This relation is formalized by Krifka (1992; 1998) as *Strictly Incremental Relations*; a structure preserving mapping between the denoted semilattices associated with an eventuality and its theme.

The theme is critical for Filip's (2008) proposal in several respects. Firstly, it provides a *dimension* (i.e. a measurable quantity) that the event and its run time apply to. In turn, the combination of the event and theme induce a set in which the event intervals progress in accordance with the measurable property denoted by the theme. For example, the verb *drink* on its own denotes a drinking event, however with the addition of its predicate *wine*, the event of drinking progresses in accordance to a quantity of wine (compare *drink wine* vs. *drink a glass of wine*). Based on this union, a homomorphic relation is created which captures the overall progression of the event (i.e. as the drinking event progresses so too does the amount of wine being consumed).

Secondly, the extension of the thematic predicate after joining with the verb causes the initial partial ordering of events to become a total ordering, or a *scale*. Therefore the sematic properties of the verb and its theme together provide a scale, which ultimately lends itself to be maximalized as predicted by Filip (2008). The next section expands on the semantic framework of the homomorphic mappings relations in greater detail.

1.6 Krifka (1992): Strictly Incremental Relations

As argued by Krifka (1992) the composition of telicity is mediated by the interaction between nominal and verbal predicates, namely, the verb and its theme. Together, both of these lexical components create a homomorphic mapping relation which Krifka calls *Strictly Incremental Relations* (SINC). Specifically, SINC relations reflect the progression of an event with respect to its theme in a non-recursive fashion with one stage of the event progressing into another. For example *drinking a glass of wine* can consist of multiple drinking events where e_1 represents the first sip and e_2 would represent the first sip (e_1) plus the sequential sip. The formal definition of such a progression is provided in (40):

40) *θ is Strictly Incremental (SINC)*
a) MSO (θ) ^ UO (θ) ^ MSE (θ) ^ UE (θ)
Where SINC relations are comprised of four mapping relations: Mapping of Subobjects (MSO), and Unique Objects (UO), and Mapping of Subevents (MSE), and Unique Events (UE)

b) $\exists x,y \in U_P \exists e,e' \in U_E[y < x \wedge e' < e \wedge \theta(x, e) \wedge \theta(y, e')]$
There exists a nominal (*x*) and theme (*y*) which are included in a set of participants (U_P) and there exists and event (*e*) and proper event (*e'*) which are included in set of events (U_E). Where the relation between sets is defined by: the ordering of *y* and *x* corresponding with *e'* and *e* respectively and which the theta role for *x* maps onto the event *e* and the theta role for *y* maps onto the proper parts *e'* of the event *e*

Therefore (40) essentially accounts for the fact that as the drinking event progresses so too does the consumption of wine within the glass. Specifically that the entire drinking event (e) corresponds with the nominal predicate (x) and the proper parts of the event correspond with the theme (y). Each of the four mappings which comprise the overall definition of SINC relations will now be discussed in turn.

The first mapping between events and predicates distinguishes incremental from non-incremental relations. Specifically, mapping to subevents (MSE) ensures no proper part of an object (*x*) maps to the whole event (*e*). For example, *drinking two glasses of wine* which denotes two events of drinking incrementally *(e' and e'')* within the entire event of drinking wine (*e*). Therefore the first glass of wine, which is a proper part of the entire drinking event, must correspond with a proper part of the event (*e'*) and not the event itself (*e*). This is formally defined in (41):

41) *Mapping to subevents, MSE(θ)*
$$\forall x,y \in U_P \forall e \in U_E[\theta(x, e) \wedge y <_P x \rightarrow \exists e'[e' <_E e \wedge \theta(y, e')]]$$
There exists a nominal (*x*) and theme (*y*) which are included in a set of participants (U_P) and there exists an event (*e*) is included in set of events (U_E). This relation is defined by: the theta role mapping *x* with *e* and the ordering of participants where *x* precedes *y*. This ordering corresponds with at least one proper event (*e'*) defined as *e'* being a subpart of *e* and the theta role maping *y* and *e'*

Strengthening this relation further, we can claim that subevents and their corresponding subobjects are unique. The uniqueness of events (UE) states that for events which involve an object, the object is only subjected at the most to one event instance of a given type. In the case of *drinking two glasses of wine* where the glasses are drank incrementally, there is a unique part of *e'* within *e* where the first glass of wine is drank. This is formalized in (42) below.

42) *Uniqueness of events, UE (θ)*

$$\forall x,y \in U_P \forall e \in U_E[\theta(x, e) \wedge y \leq_P x \rightarrow \exists! e'[e' \leq_E e \wedge \theta(y, e')]]$$

There exists a nominal (*x*) and theme (*y*) which are included in a set of participants (U_P) and there exists an event (*e*) is included in set of events (U_E). This relation is defined by: the theta role mapping *x* with *e* and the ordering of participants where *y* is less than or equal to *x*. This relation corresponds with there being exactly one proper event (*e'*) defined as *e'* being equal to or less than *e* and theta role mapping *y* and *e'*

Likewise, the same incremental relations hold for objects. The mapping to subobjects (MSO) prohibits proper parts of an event (*e', e''*) from mapping to a whole object (*x*). Thus a proper event of the drinking event of the first glass of wine cannot correspond to the consumption of the entire two glasses of wine. In addition, the uniqueness of objects (UO) states that for an event corresponding with an object, the event is subjected to one object instance of a given type. Thus the event of drinking two glasses of wine (*e*) forms a relation strictly with the unique object (two glasses of wine). The definitions of these two are presented in (43) and (44) respectively.

43) *Subobjects, MSO (θ)*

$$\forall x \in U_P \forall e,e' \in U_E[\theta(x, e) \wedge e' <_E e \rightarrow \exists y[y <_P x \wedge \theta(y, e')]]$$

There exists a nominal (*x*) which is included in a set of participants (U_P) and there exists an event (*e*) and proper event (*e'*) which are included in set of events (U_E). This relation is defined by: theta role mapping *x* and *e* and the ordering of the proper event *e'* being a subpart of *e* which is defined as there being at least one *y*. Where *y* is defined as the ordering relation of *x* preceding *y* and theta role mapping of *y* and *e'*

44) *Uniqueness* of objects, UO (θ)

$$\forall x \in U_P \forall e,e' \in U_E[\theta(x, e) \wedge e' \leq_E e \rightarrow \exists! y[y \leq_P x \wedge \theta(y, e')]]$$

There exists a nominal (*x*) which is included in a set of participants (U_P) and there exists an event (*e*) and proper event (*e'*) which are included in set of events (U_E). This relation is defined by: theta role mapping *x* with *e* and the ordering of the proper event *e'* being less than or equal to *e* which is defined as there being exactly one *y*. Where *y* is defined as the ordering relation of *x* preceding or being equal to *y* and theta role mapping of *y* and *e'*

Overall, these mappings present an interesting consequence in the sense that they create two new verb classes based on the mereological mapping between verb and theme: strictly incremental and incremental verbs. As has been illustrated thus far, strictly incremental verbs follow the SINC relations and are verbs of consumption (eat, drink), destruction (demolish, burn) and creation (build, draw, write). Likewise the second group of verbs, which function similarly to strictly incremental verbs, are incremental verbs such as *read* or *re-write, re-learn*. These verbs describe events where the corresponding object can be subjected to the same event more than once or reflect iterativity. For example, one can read chapters 1 through 5 of a textbook but go back and re-read chapter 3. In this case there are two subevents of reading which are mapped to the same part of the object (chapter 3). Therefore a mereological relation of this nature is defined formally by Krifka (1992) in (45):

45) *θ is incremental (INC)*[3]

a. there is a strictly incremental relation θ' such that SINC (θ') and,
b. θ is the smallest relation that contains θ' and is closed under the sum formation: $\theta' \subseteq \theta$ and $\forall x,y \forall e,e' \in U_E[\theta(x, e) \wedge \theta(y, e') \rightarrow \theta(x \oplus_P y, e \oplus_E e')]$

There exists a nominal (*x*) which is included in a set of participants (U_P) and there exists and event (*e*) and proper event (*e'*) which are included in set of events (U_E). this relation is defined by: theta role mapping *x* and *e* and theta role mapping *y* and *e'* which is defined by the theta role mapping the sum of participants ($\oplus_P$) *x* and *y* to the sum of events ($\oplus_E$) *e* and *e'*

Overall, the following has provided the mapping relations which comprise Strictly Incremental and regular Incremental relations. Krifka (1992) illustrates that these mappings are lexically elicited by the verb *and* its theme and create a partial ordering of events with respect to a measurable property. Filip (2008) assumes Krifka's (1992) mapping relations and argues that the ultimately create the scale which MAX_E applies to.

In addition to SINC and INC relations providing Filip's (2008) account with a scale, these relations also provide a crucial element used for the construction of the stimuli for the experiment. As was illustrated in examples (37) and (38) above, repeated here as (46) and (47), some verbs along with their theme are able to render quantized and cumulative readings with both *in/for-x time* adverbials:

46) Peter saw a cardinal for a few minutes/ in a few minutes
47) Peter saw cardinals for a few minutes/ in a few minutes

Interestingly, the same flexibility in quantized (telic) and cumulative (atelic) readings can be observed using SINC and INC verbs:

48) Peter wrote the exam for an hour/ in an hour SINC
49) Mary read the textbook for an hour/ in an hour INC

Crucially, (46)-(49) illustrate that SINC and INC verbs are aspectually neutral in the sense that they are lexically underspecified for telicity. It is only with the addition of the adverbial phrase that the mereologial mappings become modified to reach an endpoint or not. Based

[3] An important note must be made here, namely, that the framework presented so far is compatible with verbs whose events progress in a stage like fashion. The incremental notion does not account for verbs such as achievements or states, which arguably possess no proper parts (consider the achievement *reaching the top of the mountain* or state *Mary loves Peter*). Although Krifkta (1992, 1998) addresses this issue in his frameworks, I will not go into these details here as the current experiment only uses strictly incremental, incremental and scalar verbs to test the working hypothesis.

on this, SINC and INC verbs prove to be perfect verbs to use for the stimuli as they do not bias the participants perception of the event's completedness until the adverbial phrase is read. In addition to SINC and INC verbs, scalar verbs used also used in the stimuli. Scalar verbs such as *freeze*, *melt*, *warm*, *cleaned* differ from SINC and INC verbs as they lexically denote a scale but still freely shift between atelic and telic readings as illustrated in (50).

50) The maid cleaned the kitchen in and hour/ for an hour

Interim Summary

The above sections have set the stage nicely for the foundation of the semantic composition of event structures assumed by Filip (2008) and the present paper. Before moving on to the hypothesis, let us summarize what has been established thus far:

- Filip (2008) proposes telicity is derived via a maximalization operator MAX_E
- This proposal is based on several theoretical tools which include: maximalization, event semantics and scalar semantics
- Landman (1998) argues maximalization involves an operator which compares a set of alternatives and returns the maximal value within the set. This set is elicited by a scale that is lexically induced
- Filip's (2008) proposal for telicity therefore requires: i) a lexically induced scale and ii) a set of alternatives which are compared by the operator
- Krifka (1992) illustrates that the scale is lexically induced by the verb and its theme. This scale is a (strictly) incremental relation which meets the criteria of scales defined by Kennedy (2005)
- Krifka (1992) also illustrates that the set of alternatives are event intervals comprised by the verb and its theme, which MAX_E compares to return the largest unique event

However, one outstanding matter remains: *where* in the derivation does a set of eventualities rendered by the verb and its theme differentiate between telic and atelic using the MAX_E operator? As noted by Landman (1992) maximalization is an operation that occurs in place of existential closure, thus entering the derivation at a scope domain. In a similar fashion, Filip (2008) assumes maximalization enters into the derivation at the adverbial scope domain. It is here that the operator takes the set of event stages built by the verb and its theme and selects the largest unique event (exclusively in the case of telic constructions) by comparing all possible event stages. In contrast, Filip (2008) argues no such operator enters the derivation by means of the *for-x time* adverbial phrase, as there is no largest unique event present within the event stage sets of atelic constructions. Thus, the derivation of telic events consists of a comparison of set alternatives, whereas no such comparison process takes place in the derivation of atelic events. If Filip's (2008) predictions for telicity are on the right track, arguably the derivation involving a comparison of set of alternatives should require more cognitive effort to compute compared a derivation lacking such a process. The final section of the first chapter illustrates how a cognitive cost is associated with the derivational complexities between telic and atelic events. These costs are illustrated using strictly, incremental and scalar verbs. Overall, building this association provides the grounds for a theoretical hypothesis to be put forth.

1.7 Hypothesis

The current experiment seeks to investigate the cognitive nature of aspect, in terms of telicity. Based on the predictions made by Filip (2008) for the derivational differences between telic and atelic events, it follows that these derivations can be associated with distinct computational costs. Specifically, the current experiment argues derivations which involve maximalization require more cognitive effort or resources to be used in order for successful computation to take place. In other words, telic events are hypothesized to involve a higher computation cost due to the processing complexity associated with its derivation. The reason why telic events involve a higher computation cost is because the process of maximalization involves a comparison of set alternatives. Thus the operator proposed by Filip (2008) takes the set and compares all the individual event intervals within the set in order to select the largest unique event. Based on this, I argue that maximalization is a type of r*eference-set computation* (Reinhart, 1995).

Reference- set computation is a type of processing carried out in a competitive manner as it involves considering a set of alternatives (Reinhart, 1995). Several linguistic processes are argued by Reinhart (1995 and subsequent work) to rely on the comparisons of alternative derivations. Reinhart (1995 and subsequent work) theoretically, and empirically illustrates how phenomena such as quantifier scope, focus, anaphoric interpretation and scalar implicatures are all cases which are derived using reference set computation. Similarly, case agreement systems such as those in Finnish, Basque, French and Georgian (Rezac, 2006) as well as the principle of *Maximize Presupposition* (Schlenker, 2006; Kucerova, 2007; Katzir, 2007; Sauerland, 2008; among others) are accounts which employ reference-set computation. However, although reference-set computation is deemed an efficient operation that makes derivations as economical as possible, the computation itself is quite costly in terms of the amount of cognitive resources it utilizes (Reinhart, 1995; Fox, 2000). Reference-set computations consume a high degree of cognitive resources, such as attention used by memory systems, as the computation itself is a global operation which compares entire derivations rather than local portions (Reinhart, 1995).

Based on this, the following definition can be put forth stating maximalization as a type of reference-set computation:

51) Reference Set for Filip's (2008) MAX_E operator

The reference set that the MAX_E operator applies to is

a) A set of event intervals which reflect various stages of an event, and
b) Within this set there exists a largest unique event

To illustrate how (51) differentiates between telic and atelic events consider the following examples which contain an INC verb:

52) Mary read the textbook for an hour
53) Mary read the textbook in an hour

Both examples elicit reference sets comprised of event intervals of various stages of Mary reading the textbook (e.g. e_1 of Mary reading chapter 1, e_2 of Mary reading chapters 1 and 2 etc.). However, the set modified by the adverbial phrase in (53) will contain a unique largest event which is not present within the set modified by the adverbial phrase in (52). This is because the *in-x time* adverbial provides a quantized event interval of Mary reading the textbook (i.e. the adverbial closes the scale elicited by the INC relation). Therefore, as (53) contains a unique largest event as well as a set of intervals which reflect various stages of the reading event, the MAX_E operator is able to apply to this reference set.

However, although the definition provided in (51) demonstrates how the application of MAX_E differs with respect to reference-sets, it does not address the question regarding how the derivations differ in terms of cognitive cost. To illustrate how the derivations differ on this front, I establish a basic metric which relates the linguistic components to a computational cost.

I will simply refer to the units representative of a computational cost as CC-U (Computational Cost-Units). Specifically, these units represent the amount of attention systems such as working memory use when processing and manipulating linguistic information during sentence comprehension. Although there are numerous linguistic processes that occur throughout sentence comprehension, I will assign CC-units to the crucial processes present in the realization of telicity: establishing the reference-set and maximalization.

Reference sets are fundamentally built by either the union of the verb and theme (as in the case of SINC and INC verbs) or are lexically induced by the verb (in the case of scalar verbs). Thus the CC-Us corresponding to the computation of reference-sets is as follows:

54) CC-U corresponding to reference-set computation
a) 1 CC-U for the integration of theme and verb
b) 2 CC-U for the mapping of verb and theme to create a scale (set of alternatives)

Therefore, in constructions possessing SINC and INC verbs the overall cost is 3 CC-U. However, as scalar verbs lexically induce a scale or set of alternatives rather than through the integration of a verb and theme, their cost is 1 CC-U. In addition, the integration of an adverbial phrase that either modifies the reference-set with an additional quantized event interval or not is associated with a cost of 1 CC-U. Lastly, the MAX_E operator induced by the adverbial phrase corresponds with the following CC-U cost:

55) CC-U corresponding to MAX_E evaluation of reference-set
a) 3 CC-U for the comparison of all event stages within the reference set
b) 1 CC-U for returning the largest unique event

Therefore the entire process of maximalization corresponds with a cost of 4 CC-U.

With these metrics in place, let us now see how telic constructions differ from atelic constructions in terms of cognitive cost. These differences will be illustrated using constructions possessing SINC, INC and scalar verbs as all three of these verbs are used in the stimuli for the experiment.

CC-U of constructions using SINC verbs

a) *Peter wrote the exam for an hour* *ATELIC*

- 1 CC-U for the integration of theme and verb
- 2 CC-U for the mapping of verb and theme to create a scale
- 1 CC-U for integration of adverbial phrase
- 0 CC-U for maximalization the criterion for (46) is not met

TOTAL = 4 CC-U

b) *Peter wrote the exam in an hour* *TELIC*

- 1 CC-U for the integration of theme and verb
- 2 CC-U for the mapping of verb and theme to create a scale
- 1 CC-U for integration of adverbial phrase
- 3 CC-U for the comparison of all event stages within the reference set
- 1 CC-U for returning the largest unique event

TOTAL: 8 CC-U

CC-U of constructions using INC verbs

a) *Mary read the textbook for an hour* *ATELIC*

- 1 CC-U for the integration of theme and verb
- 2 CC-U for the mapping of verb and theme to create a scale
- 1 CC-U for integration of adverbial phrase
- 0 CC-U for maximalization the criterion for (46) is not met

TOTAL = 4 CC-U

b) *Mary read the textbook in an hour* *TELIC*

- 1 CC-U for the integration of theme and verb
- 2 CC-U for the mapping of verb and theme to create a scale
- 1 CC-U for integration of adverbial phrase
- 3 CC-U for the comparison of all event stages within the reference set
- 1 CC-U for returning the largest unique event

TOTAL: 8 CC-U

CC-U of constructions using Scalar verbs

a) *The maid cleaned the kitchen for an hour* *ATELIC*

- 0 CC-U for the integration of theme and verb
- 2 CC-U for the mapping of verb and theme to create a scale
- 1 CC-U for integration of adverbial phrase
- 0 CC-U for maximalization the criterion for (46) is not met

TOTAL = 3 CC-U

b) *The maid cleaned the kitchen in an hour* *TELIC*

- 0 CC-U for the integration of theme and verb
- 2 CC-U for the mapping of verb and theme to create a scale
- 1 CC-U for integration of adverbial phrase
- 3 CC-U for the comparison of all event stages within the reference set
- 1 CC-U for returning the largest unique event

TOTAL: 7 CC-U

Based on these examples, it is evident that the computation cost units associated with telic constructions are greater than those associated with atelic constructions, regardless of verb type: SINC and INC = 8 vs. 4 and Scalar = 7 vs. 3 Therefore, the following hypothesis can be put forward:

Hypothesis

a) As the derivation of telic events (i.e. those denoting an endpoint) involve a maximalization process, their overall computation should predictably be more costly compared to atelic events (i.e. those not denoting endpoints) as they involve a comparison of set alternatives
b) The derivational complexities established in (a) are further hypothesized to correspond with discrete computational costs. Specifically, as telic events involve the MAX_E operator they are predicted to be more cognitively costly than atelic events which do not use MAX_E

Crucially, cognitive costs associated with complex derivations is a potentially measurable entity and so lends itself to experimental testing. The next chapter of the thesis discusses how the above hypothesis can be tested using behavioural methods and concludes by putting forth the experimental predictions.

Chapter 2: Testing the hypothesis

Overview

With the hypothesis of the current paper established, we now move on to the second chapter of the paper which establishes *how* this hypothesis can be experimentally tested. To recapitulate, it is hypothesized that the derivation of telic structures involves a greater level of complexity and in turn, computational cost, compared to atelic structures. The reason is that the former involves a maximalization process whereas the latter does not. Therefore a method which is sensitive to the costs associated with complex processing is required in order to test for this distinction.

The section below establishes that a class of memory tasks termed complex span paradigms, are sensitive to complex processing and therefore able to be employed to test the hypothesized difference in complexity and computational cost. However, complex span tasks are based on several assumptions stemming from the concept of working memory.

Thus, in order to explain how this paradigm can be used to test for the computational costs associated with semantic complexity, section 2.1 provides an overview of how working memory operates in order to maintain information in memory. Within this overview a crucial feature associated with working memory is revealed, namely, the system's capacity limitation. Specifically, capacity refers to the system's ability to retain a finite number of memory traces (Cowan et al., 1995), or reliance on a finite activation resource (Just & Carpenter, 1980) that can be divided over information units, or ability to inhibit only a limited amount of distracting information (Oberauer et al., 2012). The assumption that working memory has a limited capacity is vital. The reason is that capacity is a measurable property and therefore lends itself to empirical testing.

Complex span paradigms measure capacity limitations of working memory by straining the system to execute concurrent tasks. Although numerous complex span paradigms exist, the paradigm of interest to this paper is the sentence span task. Section 2.2 discusses this paradigm in detail as it is adapted to test the current hypothesis.

Lastly, sections 2.3 and 2.4 discusses in detail two lines of reasoning as to why working memory has a limited capacity: decay and interference. Following this, section 2.5 illustrates that working memory plays a role in sentence comprehension. Although the current study is testing for effects related to a specific semantic process, namely telicity, this process is nonetheless involved in overall sentence comprehension. The last section illustrates how decay and interference-based approaches have been used to interpret results of psycholinguistic experiments investigating how working memory interacts with sentence comprehension. This portion of the thesis concludes by putting forth the predictions for the current experiment.

2.1 Working Memory

Working memory is defined as an executive system with a set capacity, which actively maintains and manipulates information related to current goals. One of the first conceptualized models of working memory is the multi-component model proposed by Baddeley and Hitch (1974; Baddeley, 2000). The original version of the multi-component model presents an architecture of working memory as comprised of three units: the central executive and two 'slave' systems which are the articulatory (phonological) loop and visuospatial sketchpad[4]. The central executive is argued to be the core component of the system as it is primarily responsible for attentional control and the coordination of information between the two slave systems. The articulatory loop and visuospatial sketchpad are proposed to encode and actively maintain memory traces for verbal and visual information respectively. The multi-component model further specifies how the system retains information. Baddeley and Hitch (1974; Baddeley, 2000) stress that the capacity of working memory depends on memory maintenance and resource sharing. Processing information involves the use of cognitive resources such as attention. Likewise, mechanisms within working memory responsible for actively maintaining (i.e. attentionally

[4] In addition, Baddeley (2000) argues another component of working memory exists, namely the episodic buffer.

refreshing or rehearsing in a modality-specific manner) memory traces also use cognitive resources. When working memory becomes strained under a heavy memory load, the resources used for maintenance are assumed to no longer be available for processing information. Thus the capacity of working memory is limited by the amount of available resources that can be used for processing and maintenance (storage) of information.

2.2 The Sentence Span Task

The notion of capacity is empirically supported by a vast amount of evidence from complex span paradigms which interleave an information processing task with encoding of items to memory for subsequent recall. Although several variations of the paradigm exist (reading span; Daneman & Carpenter, 1980, counting span; Case, Kurland & Goldberg, 192, operation span; Turner & Engle, 1989 and spatial span; Shah & Miyake, 1996) all paradigms are designed to jointly tap into the processing and storage functions of working memory.

Of specific interest to the current study is the reading span task (hereby referred to as the sentence span task) developed by Daneman and Carpenter (1980) in which sentences are read and words are committed to memory. This task was derived from the notion of resource sharing to serve as a tool for measuring working memory capacity. The sentence span task is based on the following assumptions: i) working memory is involved in the processing and maintenance (storage) of incoming information ii) the system has a limited capacity as cognitive resources are shared between processing and storage components. Thus, if a cognitive task requires resources exceeding the available capacity of the system, a trade-off between storage and processing occurs. This compromise is known as the *processing storage trade-off*; the phenomenon where more demanding or complex processes consume resources and thus decrease capacity available for additional information to be stored or processed.

The sentence span task consists of a primary and secondary task. The primary task requires a participant to remember a list of memory words for later recall. The purpose of this task is to tap into the storage aspect of working memory. Remembering words for later recall forces the participant to encode and actively maintain words in memory throughout the secondary task. The secondary task requires a participant to read a sentence which is visually presented to them. Therefore the secondary task forces resources used for sentence processing to compete with those used for maintaining memory words, and consequently elicit a trade-off.

In the original experiment 1 of Daneman and Carpenter (1980) participants read aloud a series of sentences while simultaneously remembering the final word of each sentence for later recall. The final words thus served as the memory words. Sentences were presented visually, one at a time, in various set sizes. Sets ranged from 2 to 6 sentences per set, which respectively contained 2-6 memory words. After all sentences within a set had been read, the participant was asked to recall the memory words in the order they had appeared. Each set size was repeated three times. Participants proceeded through increasing set sizes until they failed to recall 66% or more of the memory words within an entire set phase (3 sets in a row of the same size), at this point the experiment ended. In a second experiment a verification task was incorporated. Experiment 2 followed the same procedure as experiment 1, only this time participants were asked to verify if each sentence was true or false after reading it. The addition of the verification task was instilled to guarantee compliance with the sentence processing instructions. Importantly, most adaptations of complex span tasks include probing for performance in the secondary processing (also called distractor) tasks by, for instance, comprehension or verification questions, or judgements relating to the content of the secondary task.

The purpose of the experiments conducted by Daneman and Carpenter (1980) was to investigate whether an individual's working memory capacity predicts their reading comprehension level. Reading comprehension levels were determined by the researchers using participants' scores in reading comprehension task (SAT). The results indicated a substantial correlation between participants average recall spans with SAT scores. Specifically, participants who correctly recalled memory words from greater set sizes also possessed higher SAT scores. Thus the evidence supports the notion that working memory capacity is the source of individual differences in reading comprehension abilities. Variations in working memory capacity in both individuals and groups have been shown to correlate with various other complex cognitive tasks (see Conway et al., 2007 for a review).

However, the sentence span task, as well as other complex span tasks do not merely reveal that working memory is a predictor of cognitive performance. More crucially, complex span tasks have furthered the overall understanding of working memory's capacity limitation by providing insights into why working memory has a limited capacity (Oberauer et al., 2012). Essentially, capacity relies on two factors: incoming information into an active memory space and the subsequent loss of this information (i.e. forgetting). The dynamics of this relationship has been subject to much debate, namely, whether the loss of information from memory is due to spontaneous decay with time or interference from other contents entering working memory.

2.3 Decay

One of the earliest explanations for forgetting information from short-term memory proposed by the multi-component model is *decay:* the process of a memory trace fading over time unless it is actively maintained. Classic studies demonstrating this phenomenon are those of Brown (1959) and Peterson & Peterson (1959) where participants were asked to remember three consonant strings (e.g. TWF) while actively engaging in a distractor task, such as counting back in steps of 3. Memory recall declined as the length of distraction increased from three to eighteen seconds. This observation, therefore, suggests that the capacity of memory to store traces is dictated by how actively these traces can be maintained. The decay explanation of forgetting from short-term and working memory has been incorporated into Baddeley's multicomponent model.

A recent model of complex span performance: the Time-Based Resource Sharing Model put forth by Barrouillet and Camos (2004), relies on the general phenomenon of decay as an explanation for forgetting. The model is based on the following main assumptions. Firstly, Barrouillet and Camos (2004) follow Baddelely and Hitch (1974) in assuming that processing and maintenance require attention, and that attention is a limited resource pool which both these memory functions share. Secondly, in line with Cowan (1995), they postulate that once attention is diverted from maintaining a memory trace, the activation strength of that memory trace immediately suffers from time-based decay. Lastly, Barrouillet and Camos (2004) assume that a central bottle neck constrains memory retrieval by limiting the system to retrieve one memory at a time (i.e. memory retrieval occurs in a serial fashion). The notion of the bottle neck constraint is based on Pashler (1998) who showed that the retrieval of two memory traces cannot be carried out simultaneously. Based on the constraint imposed on the bottle neck, Barrouillet and Camos (2004) further assume that attention sharing between maintenance and processing is achieved via switching attentional focus between the two.

Therefore, under the Time-Based Resource Sharing model, working memory capacity is constrained by the time attentional resources are diverted from memory items to the processing task when these tasks are executed concurrently. For example, when an attention switch occurs during a complex span task, item maintenance is argued to be impeded upon. Consequently, the traces for to-be recalled memory words begin to weaken and consequently become harder to retrieve.

The rationale behind Barrouillet and Camos's (2004) Time-Based Resource Sharing Model can arguably be used to account for the findings in Daneman and Carpenter (1980). For instance, because participants with higher reading comprehension levels possess better reading skills, this allows them to direct less attention to reading and more attention to retaining memory words compared to participants with lower reading levels (Daneman & Carpenter, 1980). Due to this ability, attentional resources are able to be focused on maintenance more often and thus lead to increased memory recall. However, the theory of decay is not the only explanation for why working memory has a set capacity. Alternatively, memory capacity can be accounted for by interference.

2.4 Interference

While decay is primarily concerned with the consequences on memory traces of the length of time attention is diverted from them, memory item loss due to *interference* is not. Essentially, the focus shifts from the quantity of information which can be maintained in memory, to the quality of the information retrieved from memory. Interference therefore accounts for capacity in respect to the system's tolerance of intervening material that may preclude information from being retrieved.

One type of approach used to account for how interference constrains capacity is cue-based (Nairne, 2002; Van Dyke and McElree, 2006 and subsequent work). Under this approach, successful memory access is determined by the reliability of retrieval cues used to discriminate between relevant and distracting information in memory. Retrieval cues are often associated with a single memory trace. Consequently, the one-to-one correspondence creates a strong link which minimizes the risk of interference from distracting information. For example, consider the memory items A, B and C where each letter possesses a set of features indicated by the following numbers:

A {1,2,3}, B{4,5,6}, C {7,8,9}

After viewing this list, a retrieval probe, α is presented where α associates with feature {4}. The retrieval probe is assumed to retrieve the memory item B as both the probe and B associate with the feature {4}. However, cues that associate with several memory items are said to be "overloaded" (Watkins & Watkins, 1975) and thus the probability of interference increases as one cue is used to recall multiple items. For example, consider again the memory items A, B and C only this time with the following set of features:

A {1,2,**4**}, B{**4**,5,6}, C {7,8,9}

If the same α retrieval probe is used to recall B it is susceptible to interference because A also now possesses the feature {4}. This type of interference is proactive interference as the similar feature {4} precedes the target to be recalled. Likewise, a similar feature following the target creates retroactive interference (e.g. if C possessed feature 4). Therefore, unlike decay, interference accounts for the inability to retrieve memory items due to information which weakens the association between cue and trace. Cue-based models of interference such as those in line with Nairne (2002), Van Dyke and McElree (2006) among others, can account for phenomena observed in complex span tasks. Specifically, in sentence span tasks, lexical interference follows from the concurrent memory and sentence processing tasks.

A recent interference model for complex span proposed by Oberauer et al. (2012) postulates memory capacity can be explained by susceptibility to interference. The assumptions made by this model deviate from those made by decay-based approaches such as the Time-Based Resource Sharing model (Barrouillett & Camos, 2004). Firstly, the interference model for complex span assumes all items which enter working memory are encoded the same way. However, the strength to which these items are encoded depends on the voluntary effort of an individual. This includes both memory words and distractors (i.e. information other than memory words). The additional coding of distractors alongside memory words causes interference because distracting information is thought to distort the associations between memory words and the ordering in which they appear in the complex span task. Thus the distortion ultimately intrudes on memory trace sequences and makes them harder to reconstruct during retrieval.

Secondly, novelty-gated encoding applies to both types of information. That is, novelty is assessed based on various similarity properties between incoming information and information that is already stored. For example, if similar memory words are encountered, they become harder and harder to encode in a distinctive manner. Likewise, if similar distractors are repeatedly encountered the ability of later items to interfere is attenuated. In contrast, the encoding of dissimilar distractors or memory words increases the amount of interference within working memory. The general notion is supported by Farrell (2002). Other studies have shown negative effects on recall of different type of similarity, i.e. categorical similiarity when, e.g., both distractors and memory items are words compared to words vs. numbers (Turner & Engle, 1989) and domain similarities (Bayless et al., 2003). In this case the whole set of distractors may or may not be categorically different from the whole set of memory items. Based on this, interference arises from the fact that many dissimilar memory words exhaust the pool of distinctive elements. Consequently then, all subsequent words are seen as more and more similar to the already encoded set of words and thus creating interference effects.

Thirdly, working memory engages in active restoration of traces to compensate for incomplete or lost information. However, the interference model by Oberauer et al. (2012) does not assume restoration is due to active maintenance (i.e. refreshing/rehearsal) as decay-based models do. Instead, restoration is postulated to involve removing interfering material from memory. Based on this point, removal in decay-based models occurs by default; memory traces automatically deteriorate unless actively maintained. However, removal of material under the interference model assumes all information is retained by default (but may be hard to retrieve). It is only through active removal that irrelevant information can be eliminated. This removal of irrelevant information serves to preserve relevant information. Oberauer et al. (2012) propose the mechanism responsible for such removal is Hebbian anti-learning. This operation essentially ‘unbinds’ the association between the distractor and its location in relation to other encoded material. Farrell and Lewandowsky (2004) provide evidence supporting the existence of this suppression-type of mechanism. Based on this and the above assumptions, the individual differences in working memory capacity can be accounted for by the ability to control what information is retained and what is removed from memory.

Overall, the importance of decay and interference is that they provide two lines of reasoning as to why working memory has a limited capacity. To summarize, decay-based theories argue working memory has the ability to retain a finite amount of information due to attentional resources. The processes executed by working memory and those used for other cognitive functions are fueled by the same resource pool. Therefore tasks which pose higher resource demands leave little or no resources left for other tasks to be successfully completed. Based on this, poor performance on the sentence span task is accounted for by the two concurrent tasks competing for resources. In contrast, interference-based theories argue working memory’s ability to retain information is determined by the amount of interference caused by distractor tasks. The addition of tasks such as processing sentences and answering verification questions impede on the memory traces established for memory words. This interference in turn weakens the memory traces making them harder to retrieve for later recall. The next section illustrates how these approaches have been used to formulate predictions and interpret results of psycholinguistic experiments investigating how capacity plays a role in sentence comprehension.

2.5 Capacity and Sentence Comprehension

Language processes such as sentence comprehension, involve working memory. This is evident from numerous psycholinguistic studies, although there is a debate as to whether working memory is a domain general (King & Just, 1991) or domain specific (Caplan & Waters, 1999) entity and what role it plays in sentence comprehension. The following sections illustrate i) how complexity associated with linguistic derivations is prone to interference and ii) how linguistic complexity can be associated with a cognitive cost. Although these associations are based on insights from research focusing on syntactic complexity, it nonetheless serves as a foundation that offers a rationale for investigating the effects of semantic processing load hypothesized by the current study, which is recapitulated in (51) below:

Hypothesis

a) As the derivation of telic events (i.e. those denoting an endpoint) involve a maximalization process, their overall computation should predictably be more complex compared to atelic events (i.e. those not denoting endpoints) as they involve a comparison of set alternatives
b) The derivational complexities established in (a) are further hypothesized to correspond with discrete computational costs. Specifically, as telic events involve the MAX_E operator they are predicted to be more cognitively costly than atelic events which do not use MAX_E

2.5.1 Syntactic Complexity and Decay

As hypothesized in (51b), the derivation of telic events are assumed by the current study to be semantically more complex than the derivation of atelic events and, based on this, hypothesized to be associated with discrete computational costs. Although section 1.7 presented a basic metric relating the various linguistic components involved in telicity with computational cost, similar approaches have been used to associate syntactic components with computational costs. This section illustrates how cognitive costs associated with linguistic complexities can be dissociated under a decay based account.

Early psycholinguistic theories of sentence comprehension were based on the concept of capacity in the sense of decay, as they assumed resources used for active maintenance and processing are limited and share the same pool of resources (Just & Carpenter, 1992; Lewis, 1996; Gibson, 1998 among others). Based on these assumptions, Gibson (1998) proposed a theory which argues that a reported difference in complexity associated with comprehending subject vs. object extracted relative clauses is due to the computational resource demands they place on a parser.

Specifically, the Syntactic Prediction Locality Theory hypothesizes two components account for why subject extracted relative clauses are easier to comprehend than object extracted relative clauses. The first is a *memory cost* component, which manages the quantity of computational resources required for maintaining incoming incremental pieces of information. The second component accounting for the differences in syntactic complexity is *integration cost*. This component dictates the quantity of resources apportioned to the integration of new words into the structure the parser builds. Although the theory proposes various factors can be used to quantify memory and integration cost, it places specific focus on locality and the number of discourse items which intervene between discourse referents. Therefore the complexity difference between subject and object extracted relative clauses is influenced by locality in the following respects: i) the longer a syntactic category is held in memory before satisfying a prediction of the sentence, the greater the memory cost is in order to maintain that category ii) the greater the distance between a word and the head which it attaches to, the higher its integration cost will be.

Gibson (1998) calculated memory and integration costs with a metric system of memory units (M) and integration units (I) respectively. Memory costs increase when the minimal number of syntactic head categories exceed 2 (head noun phrase for the subject and head verb phrase for the predicate). When additional heads are encountered, the parser needs to make further predictions regarding the syntactic structure and these additional predictions are responsible for increased memory loads. Likewise, factors which substantially increase integration cost are lexical items indicating new discourse referents. Discourse referents are entities which have been established previously in the discourse or through common-ground that are referred to using anaphoric expressions (e.g. pronouns).

The complexity difference between subject and object extracted relatives in terms of computational cost can be illustrated as follows. Consider examples (56) and (57) from Gibson (1998):

56) The reporter who the senator attacked admitted the error. (Object extracted relative clause)
57) The reporter who attacked the senator admitted the error. (Subject extracted relative clause)

According to Gibson (1998) (57) is predicted to be easier to process than (56) as the memory and integration costs associated with (57) are 'less expensive' than (56).

Using the metric established by Gibson (1998) the object extracted relative clause in (1) has four integration processes with the following integration costs. The first integration cost occurs when the verb 'attacked' is encountered by the parser. This cost is due to the verb 'attacked' assigning an agentive thematic role to the NP 'the senator', and thus establishing a new discourse referent (I(1)). The second integration cost comes co-indexing the relative pronoun 'who', which is an additional discourse referent (I(2)). These costs are illustrated below in (56a).

56a) The reporter	who	the senator	attacked	admitted	the error.
I(0)	I(0)	I(0)	I(1) +I(2)	-	-

Next, the verb 'admitted' adds the integration cost to I(3) as the verb is integrated back as the main verb of the subject NP 'the reporter'. Lastly, 'the error' is integrated for a cost of I(1) as it is recognized as a new discourse item since the verb 'admitted'. The complete schemata of the integration costs for example (56) above is illustrated in (56b).

56b) The reporter	who	**the senator**	**attacked**	admitted	the error.
I(0)	I(0)	**I(0)**	**I(1) +I(2)**	I(3)	I(0) + I(1)

Essentially similar integration costs are associated with the incremental portions of the subject extracted relative clause in (57) except for the verb 'attacked' and 'the senator'. The integration of the verb 'attacked' is associated with an integration of I(1) as no new discourse referent is established with the verb other than ‘the reporter’. However, once ‘the senator’ is a new discourse referent, and so comes with an integration cost of I(1). The integration costs associated with the subject extracted relative clause in (57) is illustrated below in (57a). The integration cost differences between (56) and (57) is indicated by the bolded sections within (56b) and (57b).

57a) The reporter	who	**attacked**	**the senator**	admitted	the error.
I(0)	I(0)	**I(0) + I(1)**	**I(0) + I(1)**	I(3)	I(0) + I(1)

In addition a similar difference can be noted between memory costs associated with examples (56) and (57). Memory costs are associated with the number of predictions a parser makes based on the syntactic head categories encountered in a sentence. Several memory costs are therefore affiliated with the object extracted relative clause in example (1) such as: the NP-pronoun gap position that is coindexed with the wh-pronoun, and the integration of new discourse referents (e.g. the senator). The costs associated with these components are illustrated in (56c).

56c)	…	who	**the**	**senator**	attacked	admitted	the error
Total		2M(0)	**3M(0)**	**2M(1)**	-	-	-
Wh-gap		M	M	M			
Embedded VP		M	M	M			
Embedded sub-NP		-	-	M			

Similar memory costs are observed for the subject extracted relative clause. However, the overall cost of this construction is less than that of the object relative due to the absence of an embedded subject. The costs associated with the subject extracted relative clause is illustrated in (57b). Again, the differences in memory costs between examples (56) and (57) are indicated via the bold text.

57b)	…	who	**attacked**	**the**	senator	admitted	the error
Total		2M(0)	**1M(0)**	**1M(0)**	-	-	-
Wh-gap		M	-	-			
Embedded VP		M	-	-			
Embedded obj-NP		-	M	M			

To summarize, the theory presented by Gibson (1998) argues the reason why subject extracted relative clauses are more easily comprehended than object extracted relative clauses is because of their respective computational resource costs. Together, the computational costs associated with object relative constructions are greater than those associated with subject relative constructions. Based on this, the following predictions can be put forth if these types of constructions were used in a complex span task. As object relative constructions are associated with higher computation costs it follows little to no resources would be left to use for memory maintenance throughout the complex span task. Thus memory words would become susceptible to decay due to minimal available resources[5]. Alternatively, if resources were allocated to maintaining memory words then little to none would be left for processing sentences. Either way, a processing-storage trade off would predictably take place.

Based on this, one would expect to see a lower number of recalled memory words for trials containing object relative constructions. In contrast, as subject relative constructions do not consume as many computational resources as object relatives, more resources would remain available for memory maintenance (or alternatively processing). Thus, one would expect to observe a higher number of recall memory words in trials with subject relatives compared to trials with object relatives.

2.5.2 Syntactic Complexity and Interference

As hypothesized in (51a), the derivation of telic events is assumed by the current study to be semantically more complex than the derivation of atelic events as the former involves a comparison of set alternatives. Interestingly, assumed differences in syntactic complexity present in sentence comprehension can be accounted for by interference effects. This section illustrates how linguistic complexity is susceptible to interference using a variation of the sentence span task.

Gordon et al. (2002) suggests thatt capacity is not strained by the quantized value such as the number of recall words or factors such as those used in Gibson (1998; 2000). Instead, memory is strained by the amount of interference caused by lexical items, constructions or memory words that need to be kept active in working memory. To test this, Gordon et al. (2002) conducted a memory load experiment. Participants read a list of three to-be-recalled words which consisted of either names (e.g. Peter) or occupations (e.g. carpenter). Following this, participants were then presented a sentence which was either a subject or object cleft construction which was read at self-pace. Sentences either contained category words which were names or occupations. These category words either matched or were dissimilar to the memory words as illustrated in (58) and (59) below. Crucially, this manipulation was done to elicit interference when the recall and sentence category words matched. Immediately following each sentence was a true or false comprehension question. Once participants indicated their answer, they were then asked to recall the three memory words.

Subject cleft sentences with matching (a) and dissimilar (b) category and memory words
58a) It was the dancer that liked the fireman ~ poet
b) It was the dancer that liked the fireman ~ Andy

[5] For the purposes of this section, I am focusing on how Gibson's (1998) theory of the comprehension differences between relative clauses can be accounted for by decay-based accounts. However, putting aside this emphasis on decay, it is important to recognize that the predictions made by Gibson's (1998) can, in certain respects, be accounted for by interference. Specifically, in the case where the cost associated with intervening items between discourse referents. Here no decay need be assumed, as intervening words or nodes may create interference.

Object cleft sentences with matching (a) and dissimilar (b) category and memory words
59a) It was the dancer that the fireman liked before the argument began ~ poet
b) It was the dancer that the fireman liked before the argument began ~ Andy

The results illustrated two significant findings. Firstly, when recall words matched those of the category words within the sentence, an increase in comprehension errors were observed. Secondly, an interaction was observed between syntactic complexity and interference caused by matching recall and sentence category words. Specifically, interference had a greater detrimental effect on memory words in trials with object cleft sentences.

As discussed by Gordon et al. (2002) these results have two implications: syntactic processing appears to share the same pool of resources used to maintain recall words, and interference can play a role during sentence comprehension. The former implication is supported by the fact that a significant interaction between comprehension accuracy and type of cleft sentence, when sentence category words matched recall words. Gordon et al. (2002) assert interference is possible during the retrieval stage of sentence comprehension. This becomes apparent as cleft sentences require a word trace to be retrieved during processing from a specific sentential position. Thus when the recall word and the NP within the sentence match, multiple traces for a similar cue are created. In turn, the risk of interference increases as the retrieval of the relevant piece of information becomes harder to identify due to the multiple traces associated with it.

A follow up study by Fedorenko et al. (2006) was conducted as the comprehension results from Gordon et al. (2002) were argued to be an offline measure. The purpose of the Fedorenko et al. (2006) study was to determine whether the findings of Gordon et al. (2002) could be replicated with an online measure.

Although the experiment of Fedorenko et al. (2006) was similar to Gordon et al. (2002) a new type of memory load manipulation was implemented into the experimental design. Participants were first presented with either one *or* three recall words which again were either names or occupations. This manipulation was done to provide a quantitative and qualitative measure of interference caused by memory words. That is, whereas Gordon et al. (2002) manipulated similarity interference by matching recall and sentential category words, Fedorenko et al. (2006) also manipulated number of interfering units.

After the presentation of the memory word list, a sentence was read at self-pace. The category NPs within sentences did, or did not, match for occupation only. In addition, syntactic complexity was manipulated by presenting either object or subject relative clause constructions rather than cleft constructions.

The results contained three significant findings: better recall performance was seen in trials where memory lists contained one recall word compared to those with three, participants recalled memory words that were names more accurately than occupations and finally, an interaction was observed between syntactic complexity and matching recall words in conditions with lists of three memory words. Specifically, reading times were slower for object relative clauses compared to subject relative clauses only in conditions where the recall and category words matched and a list of three recall words were used.

The observation of slower reading times between conditions for object relative clauses was not found to be significant between conditions for subject relative clauses.

The results observed in Fedorenko et al. (2006) extend those found in Gordon et al. (2002) as not only does the type of memory word (i.e. matching) cause interference but so does memory load (i.e. number of words to be recalled). Therefore, effects of interference on sentence comprehension (reflected by participants reading times) can be caused by similarity in the memory traces which are encoded or by the number of items kept active in working memory.

Crucially, these interference effects were observed exclusively for conditions with object relative sentences. This finding suggests constructions which are syntactically more complex are more susceptible to interference effects during processing. One explanation accounting for why such constructions are prone to interference is the proposal by Gibson (1998), which was discussed in section 2.5.1. As mentioned, Gibson (1998) argues the integration of new discourse items can be local or non-local. Subject relative constructions involve local integration as the integration of the embedded verb with its embedded subject is immediate due to no potential attachment sites being present. Object relative constructions involve non-local integration as an embedded subject intervenes between the subject of the main clause and embedded verb. Thus the integration of the main clause subject with the main clause verb is postponed, causing a trace to be retrieved once the main clause verb is encountered. Based on this, Fedorenko et al. (2006) argue this difference between integration processes accounts for why objective relative constructions are more vulnerable to interference effects. As subject relatives involve no search for an attachment site, a heavy memory load or similarity effects cause minimal interference on the integration process. However, as object relatives require attachment sites to be retrieved from memory and have intervening attachment sites, these integration processes are more complex and therefore more susceptible to interference.

Capacity and Sentence Comprehension Summary

Crucially, the above section has presented two ways in which syntactic complexity in sentence processing can be accounted for by capacity limitations of working memory: decay or interference based accounts. As proposed by Gibson (1998), syntactic complexity can be associated with computational costs. The more complex the syntactic derivation is, the greater amount of computational resources will be required in order to successfully process it. On the other hand, as proposed by Fedorenko et al. (2006) the more complex a syntactic structure is, the more vulnerable that structure is to interference caused by other items held active in memory. As the current study is investigating hypothesized complexity and associated computational costs at the semantic level, the insights presented above are directly relevant as they offer a rationale to testing the effects of aspectual realization in terms of telicity on working memory. Lastly, the decay-based account of Gibson (1998) and interference-based account of Fedorenko et al. (2006) set the stage for what predictions should be made for the current hypothesis to be tested by using the sentence span task.

2.6 Predictions

The current experiment takes as its starting point the proposal that aspect, in terms of telicity, is realized via maximalization and tests a hypothesis of this causing increased cognitive costs for processing telic event descriptions. Under Filip's (2008) account maximalization occurs via the operator MAX_E which compares all event intervals within a set of events and returns a maximal candidate. This process can be argued to be a type of reference set computation which entails complex processing related to increased computational resources. With this said, as MAX_E is exclusive to telic sentences (Filip, 2008), these structures are hypothesized in the current study to involve a higher level of complexity and in turn, utilize more cognitive resources compared to atelic sentences which do not recruit the operator.

Based on sections 2.5.1 and 2.5.2, two sets of predictions can be put forth: predictions coinciding with the effects of decay-based and interference-based accounts of working memory capacity. Under the assumptions of decay-based models such as the Time Based Resource-Sharing Model (Barrouillet & Camos, 2004) working memory capacity is assumed to be limited due to the following:

a) Memory traces fade unless actively maintained or refreshed
b) The resource used for active maintenance is assumed to be attention. Attentional resources come from a pool which maintenance and processing functions share

Based on these two principles, the ability to remember and recall memory items during a sentence span task is constrained by the amount of time attentional resources are diverted from memory items to the processing task. Therefore if the processing task is assumed to be cognitively costly then the processing task will impede on the resources left available for any subsequent tasks, such as memory item maintenance. On the other hand, if the processing task is associated with lower cognitive costs this task will free-up resources, making more available for memory item maintenance. Therefore the current study predicts that if telic events are more complex and thus more costly than atelic events, the following results will be seen in the sentence span task:

i) If telic events involve more computational resources to be successfully computed, their processing will impede on the resources available for memory item maintenance and result in a lower number of words recalled across telic trials
ii) If atelic events involve a lesser degree of computational resources this will allow for more resources to be available for memory item maintenance and result in a higher number of words recalled across atelic trials

Contrary to the decay-based account, interference-based accounts of memory capacity such as Oberauer et al. (2012) assume capacity limitation is due to the following:

a) Distractors and memory words distort or interfere with the associations made for items already encoded into memory

b) Active removal is required in order to filter out irrelevant information from being stored in working memory

Based on these principles, the ability to remember and recall memory items during a sentence span task is constrained by the amount of interference with memory processes and the ability to remove unnecessary information. As observed in the study by Fedorenko et al. (2006) processing tasks which involve a greater level of linguistic complexity are more susceptible to interference. On the other hand, processing tasks involving lesser levels of complexity are more tolerant to withstanding the effects of interference. A crucial manipulation implemented in the experimental design for the current study is the position of the to-be-recalled memory words. Memory words were placed either at the end of the sentence (following the classic Daneman & Carpenter 1980 design) or placed before the adverbial phrase. By placing the memory word within the sentence, the parse becomes interrupted and hence, the level of interference increases. Therefore the current study predicts that if telic events are more complex as they involve a comparison of set alternatives whereas atelic events do not, the following results will be seen in the sentence span task:

i) If telic events involve a greater level of semantic complexity in their derivation (due to the comparison of set alternatives) then it is predicted that their processing will be more prone to interference and result in a lower number of words recalled or lesser accuracy in correct responses to comprehension questions across telic trials.

ii) Trials of telic events are predicted to be more prone to interference in conditions where the memory word is placed before the adverbial phrase compared to conditions where the memory word is placed sentence finally. This is because the position of the memory interrupts the parse of the sentence.

iii) If atelic events involve a lesser degree of semantic complexity in their derivation (due to the absence of MAX_E) then it is predicted that their processing will be more tolerant to the effects of interference and result in a higher number of words recalled or greater accuracy in correct responses to comprehension questions across telic trials

iv) Trials of atelic events are predicted to be more prone to interference in conditions where the memory word is placed before the adverbial phrase compared to conditions where the memory word is placed sentence finally. This is because the position of the memory interrupts the parse of the sentence.

This concludes the second chapter of the thesis. The final chapter presents the details of experiment (sections 3.1 to 3.3), the results in section 3.4, and discusses to what extent the results support the hypothesis in section 3.5. Lastly, section 3.6 concludes.

Chapter 3: The Experiment

3.1 Participants

Twenty confirmed native speakers of English participated in the experiment after giving informed consent. Two participants' data were excluded from the analysis as they scored less than 65% on the comprehension questions component. The final analysis included data from eighteen participants (13 females and 5 males) between the ages of 18 and 22 (M=19.05, SD =1.03). All participants were undergraduate students from McMaster University who received course credit for their participation.

3.2 Stimuli

Overall 100 critical sentences were created, consisting of 50 telic and 50 atelic sentences. In addition, 100 filler sentences were used from another experiment. The filler sentences contained manipulations unrelated to telicity but also used similar memory word location manipulations.

3.2.1 Telicity & Minimal Pairs

Sentences contained either SINC (31), INC (8) or Scalar (11) verbs and were an average of 8.61 words in length (SD= 1.60). Verb frequencies were extracted from the Corpus of Contemproary American English (COCA; Davies, 2008): SINC verbs (COCA mean frequency= 10487) INC (COCA mean frequency= 18060) and Scalar (COCA mean frequency= 4513).

As mentioned in section 1.0, Giorgi and Pianesi (2000) illustrate the *in/for-x time* adverbial test provides explicit telic and atelic readings. Therefore, incorporating these adverbial phrases into the stimuli is crucial, as it ensures the aspectual interpretations made by participants are definitively telic or atelic.

The sentences were thus constructed as minimal pairs with the following syntactic structures: SINC and INC sentences contained a simple noun phrase followed by a past tense verb with a theme and an adverbial phrase. In the case of Scalar verbs no theme was provided. As mentioned in section 1.3.6, these verbs lexically induce a scale whereas SINC and INC verbs rely on a theme to do so. Examples (60a-c) illustrate the minimal pair constructions for all verb types below.

<u>SINC</u>
(60a) The oven warmed the pizza for a few minutes (*atelic*)
The oven warmed the pizza until the cheese melted (*telic*)

<u>INC</u>
(60b) The student read the book for an hour (*atelic*)
The student read the book in an hour (*telic*)

<u>SCALAR</u>
(60c) The laundry dried outside for the afternoon (*atelic*)

The laundry dried outside before it started to rain (*telic*)

In addition, all sentences were in past tense as opposed to being conjugated in the present or future tense due to the following semantic complications. Firstly, the present progressive denotes that the event time and speech time are equal. This characteristic allows for an atelic reading, as at the time of uttering *The pizza is warming for 10 minutes*, the initial boundary of the warming event is established but the right boundary remains unspecified. However, in the case of **The pizza is warming in 10 minutes* the progressive tense is incompatible with the durative adverb. The semantic nature of the progressive requires the event time be left unspecified, and so the *in 10 minutes* AdvP is ungrammatical as it closes the right boundary of the event.

Secondly, the future is also problematic as this tense stipulates sets of possible worlds or outcomes of an event. In turn, the model of possible worlds allows for the progression of an event and time to not necessarily proceed in a linear fashion (Thomason, 1970; Dowty, 1977). As a scale is strictly linear is a specified direction, the use of the future would not render an explicit temporal boundary nor denote a maximal event, only a possible one.

The past tense denotes that an event occurs before the speech time. However, an event's run time can be perceived as terminated at the speech time (*The oven warmed the pizza in 10 minutes)* or can continue beyond the time of speech as well (*The oven warmed the pizza for 10 minutes as is still warming it*). However, not all verbs in the past tense are possess the ability to remain aspectually neutral and combine with either adverbial phrases. For example, verbs such as *recovered* and *moved* in the past tense are interpreted by default as telic. Therefore judgements were collected for all verbs used in the experiment to control for this effect. Individuals (who did not participate in the actual experiment) provided judgement for sentences, indicating whether they thought the verb denoted a completed action regardless of the addition of the adverbial phrase. Verbs which were judged to be completed by default were not used. For these reasons, using verbs in the past tense is optimal as it gives way to both grammatical telic and atelic readings.

Lastly, the minimal pair design is crucial as it provides the reader with an explicit telic or atelic reading. In addition, each verb (regardless if it is SINC, INC or scalar) is not inherently specified in terms of telicity. Thus, when the reader encounters the verb no priming or expectations are created and the parse remains open for a telic or atelic interpretation to be made. Together, these manipulations ensure participants would make clear telic or atelic realizations and to eliminate possible semantic confounds.

One note should be made here. Although the minimal pair design provides the experiment with a direct means of comparison, it is also creates an experimental concern. By using minimal pairs, the participants are ultimately exposed to the same sentence twice and are essentially primed. As has been established in the literature, priming facilitates word recall as the strain of the secondary task is somewhat alleviated. Based on this, using minimal pairs confounds the taxing nature of Sentence-Span Task to a certain degree. However, priming confounds were counter-acted by presenting half of the participants with the minimal pairs in one order and half with the opposite order. As this section is dedicated to the stimuli, a more detailed discussion regarding this point will be provided in the discussion section.

3.2.2 Memory Word Location

Traditional Sentence-Span tasks include the memory word sentence-finally (SF) (Conway, Kane et al. 2005). However, as no experiments testing for this semantic component of the grammar exist within the literature, it was unclear to the researchers if the Sentence-Span Task would be able to successfully demarcate the predicted aspectual processing loads. The additional manipulation instilled by the current experiment was done to increase the task's sensitivity of measuring cognitive costs. By placing the memory word before the adverbial phrase (B.Adv) the parse of the participant is interrupted. This interruption increases the overall difficulty of the secondary task as it forces the initial parsed sequence to be stored when the memory word is encountered. Once the memory word is read it, too, must be stored and the previous information recalled, as it now needs to be integrated with the remaining portion of the sentence (the adverbial phrase). Examples (61) and (62) illustrate the corresponding memory word location manipulations.

<u>Atelic</u>
(61a) The oven warmed the pizza for a few minutes **HAND** (SF position)
(61b) The oven warmed the pizza **HAND** for a few minutes (B.Adv position)

<u>Telic</u>
(62a) The oven warmed the pizza until the cheese melted **HAND** (SF position)
(62b) The oven warmed the pizza **HAND** until the cheese melted (B.Adv position)

3.2.3 Recall Words

One hundred memory words were devised for the experiment. All recall words were nouns of 3 to 4 letter lengths and had a mean frequency rating above 1000 using the Corpus of Contemporary American English (COCA; Davis, 2008). Two lists of memory words were created by randomizing the recall words twice offline. Each list was assigned to two lists of stimulus sentences, therefore ensuring the same recall words were associated with the same trials in 1A &1B list pair and the 2A & 2B. No words were repeated within or across lists.

3.2.4 Trial Groupings

Each experimental session consisted of 40 trials (20 critical and 20 filler) with the total number of stimuli used being 200 sentence-word pairings. Critical trials were divided into 4 groups, with each group containing 5 trials and each trial comprised of 5 sentences. The four groupings are as follows:

i. Five trials containing Atelic and Sentence Final sentence-word pairings
ii. Five trials containing Atelic and Before Adverbial sentence-word pairings
iii. Five trials containing Telic and Sentence Final sentence-word pairings
iv. Five trials containing Telic and Before Adverbial sentence-word pairings

Overall four lists of stimuli were created for the experiment. The trials within the lists were pseudo-randomized to ensure the following: no more than two consecutive trials containing the same manipulations for telicity or word position were present within the overall sequence. Lists 1B and 2B were lists with the opposite trial ordering of lists 1A and 2A respectively. As there is a high degree of practice effects observed within sentence span tasks (Conway et al., 2005) the reverse ordering was created as a means to control for this effect. Lastly, lists were distributed in a pseudo-randomized fashion across participants to ensure that each list was conducted on an equal number of participants.

Lastly, a hundred yes/no comprehensions questions corresponding with the hundred sentences were created. Each trial contained 2 questions pertaining to the aspectual content of the sentence while the remaining 3 questions served as filler. Filler questions pertained to other components of sentences such as the subject or theme. The same questions were used across all 4 stimulus lists.

3.3 Experimental Procedure

The experiment is a within-subjects 2x2 design with the following independent variables: telicity and memory word location. Sentences contained either telic or atelic interpretations and the memory words were presented either sentence finally or before the adverbial phrase.

Each experiment began with an instructions screen on an iMAC computer, followed by 5 practice trials to familiarize the participant with the procedure and expectations of the experiment. Following the completion of the practice trials, a screen appeared signalling to the participant the actual experiment would begin after pressing any key. All information and stimuli was presented visually with only instruction, answer and recall screens being self-paced.

Before each trial a fixation cross appeared for 500 msec followed by the first sentence. Each sentence was presented phrase by phrase. The visual duration of each phrase was calculated using the following standards: 200 msec for function words and 400 msec for content words. Memory words presented either in the sentence final or before adverbial phrase position were presented in isolation for 1000 msec. All phrases were presented in black font whereas memory words were presented in red bold font and in capital letters. This was done to ensure that the memory words were visually discrete from the rest of the sentence.

Following each sentence, a simple yes/no comprehension question would then appear. The inclusion of comprehension questions within the experiment primarily served as a probe ensuring attention to the sentence processing task. Questions either pertained to the aspectual nature of the sentence or other subject matter such as the subject, verb or object. Participants were instructed to answer questions as accurately and quickly as they could as their reaction times were being recorded. Answers were indicated by pressing the corresponding Y and N keys on the keyboard for yes or no respectively. Once the answer had been provided the next sentence in the trial began.

After the fifth question in each trial was answered, a screen appeared prompting the participant to recall the memory words from the trial in the order they appeared to the best of their abilities. Responses were manually documented using a corresponding answer key. The following trial began by the participant clicking any key. No feedback regarding the participant's progress was provided at any time during or following the experiment. The overall length of the experiment ranged between 40 and 50 minutes.

3.4 Results

Recall across the four conditions is shown in Figure 1. The results of this experiment were analyzed using a generalized multiple linear logistic regression. Several variables were included as predictors of word recall. Telicity and memory word location were of primary interest, however memory word frequency and trial number were control variables in the regression.

Binary values were assigned to the dependent variable of participants' word recall of individual words: 1 for correct responses and 0 for incorrect responses. This scoring procedure is described as all-or-nothing load scoring by Conway et al. (2005) and is the most commonly used scoring procedure. The all-or-nothing procedure thus returns a proportion of correctly recalled words for all trials. Specifically, words must be recalled in their correct serial order to receive a score of 1. If participants did not recall a word in its correct position (e.g. if *cat* was the second word but recalled as the third) this was recorded as 0.

The results of the regression indicate that frequency of memory words was a significant predictor of correct word recall (β= 0.06, SE= 0.14, p <0.01). There was also an interaction of telicity and memory word position (β= -0.38, SE= 0.19, p <0.5). All other variables were not found to be significant. A chi-square test of the model including all the predictors against a constant only model was statistically significant (p <0.1, df= 5), indicating the predictors reliably distinguish between incorrect and correct word recall. Table 3 below provides a summary of the regression results.

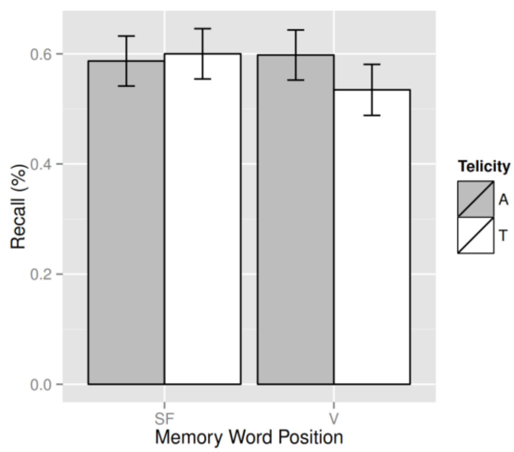
Table 1: Summary of regression analysis for variables predicting correct memory word recall (N= 1800 before trimming, and N= 1798 after trimming). Recall accuracy reported as the regression coefficient, standard errors, z-values and p-values

Variable	**Coefficient (β)**	**Std. Error**	**z-value**	**p-value**
Constant	0.95	0.26	3.63	< 0.001
Telicity (T)	0.09	0.14	0.62	0.534
Memory Word Position	0.06	0.14	0.47	0.640
Trial Number	0.01	0.01	0.96	0.339
Frequency	-0.09	0.03	-3.13	<0.01*

Telicity (T) x Memory Word Position	-0.38	0.19	-1.97	<0.05*

The interaction of telicity and memory word position being significant predictors of word recall supports the notion that telicity involves a higher complexity. This finding is further illustrated in Figure 1 and Table 4 below. The proportion of correctly recalled memory words decrease from 60% in the sentence final condition to 53% in the before adverbial condition.

Figure 1: Proportion of correctly recalled memory words in Sentence Final (SF) and Before Adverbial Phrase (V) Memory Word Positions in both Atelic and Telic Trials



Significant finding observed in Telic (white) trials between SF (60%) and V (53%) Memory Word Positions

Table 2: Proportion of words recalled correctly (1) and incorrectly (0)

Sentence Final Word Position (SF)		**Adverbial Phrase Word Position (B.Adv)**	
Atelic	Telic	Atelic	Telic
0= 0.10 (41%)	0= 0.01 (40%)	0= 0.10 (40%)	0= 0.12 (47%)
1= 0.15 (59%)	1= 0.15 (60%)	1= 0.15 (60%)	1= 0.13 (53%)

Interaction indicated in Table 3 above for Telicity x Memory Word position ($p<0.05$) is reflected in decrease in proportion value of 60% correct in Telic (1) SF to 53% correct in Telic (1) B.Adv

3.5 Discussion

The current experiment investigates whether aspectual realization, in terms of telicity, is realized by means of competitive based processing (i.e. maximalization). Evidence supporting the hypothesis would confirm the following prediction: a lower number of recalled memory words observed for atelic conditions (both SF and B.Adv) compared to telic conditions.

The results support the general prediction to some degree: a significantly lower number of words were recalled between telic conditions. Specifically, a lower proportion of words were recalled in the telic condition with the memory word situated before the adverb (~53%) compared to being placed sentence finally (~60%). Interestingly, this observation was not mirrored in the atelic conditions nor was a significant difference found between telic and atelic conditions. With this said, although the evidence presented here does not strongly support the original hypothesis it indeed suggests that a difference in complexity between the two aspectual readings does exist.

The significant difference of words recalled between telic conditions, which is absent between atelic conditions suggests the former possesses a higher degree of processing complexity than the latter. Although the evidence provided here does not strongly support the notion that maximalization is the cause of such complexity, we should not be quick to reject the original hypothesis completely.

3.5.1 Methodological Considerations

As this experiment is the first of its kind there are several methodological factors which need to be considered before accepting the null hypothesis. The sentence-span task is an established and reliable behavioural paradigm for testing processing-storage trade-offs associated with complex cognitive tasks. Originally the task was designed as a means to measure the capacity of working memory which stressed the functional importance of the system itself. With this said, such working memory span tasks are indeed sensitive to depicting cognitive behaviours across domains such as those used for reading comprehension and problem solving, however span tasks' sensitivity to depicting domain specific behaviours is debatable (Conway et al., 2005). Thus span tasks capture general executive demands such as attention which can be indicative of specific linguistic complexities.

For instance, several reports have been made of offline interactions between syntactically complex structures (such as object vs. subject relative clauses) and memory word recall (Waters et al., 1987; Wanner & Maratsos, 1978). As argued here, telicity is an intricate semantic component of the grammar which is predicted to be associated with a cognitive processing cost. However it may be that the method adopted is too crude a tool to pick up on the linguistic phenomenon predicted by the current experiment.

However, what other method could be used to conduct the current experiment? An alternative method which could potentially test for increases in cognitive processing is electroencephalography (EEG). This method would instead reflect online neurological responses associated with competition based processing in terms of when it occurs in the derivation. Such an experiment would shed light on whether telicity is a local vs. global phenomenon more so than confirm if the derivation is computed via competition. This brief proposal, if carried out, would ultimately serve to further the investigation being started here.

In addition to the considerations of the methodology itself, two components of the stimuli pose as confounding. Firstly, a significant effect ($p= <0.01$) for memory word frequency was observed. Although the analysis regresses out this effect, it is still a concern that would need to be better controlled for if the experiment were to be re-administered. Specifically, the memory words in the present study were only controlled for content frequency, not for familiarity and imaginability. In addition, the range of frequency ratings should be further restricted to less than 500. Implementing such changes would further strengthen the validity of the present results and possibly strengthen the significance of telicity and memory word location as predictors of word recall.

Secondly, the use of a minimal pair design for the stimuli raises concern of priming effects amongst participants. The purpose of using a minimal pair design ensures sentences possess explicit telic and atelic readings in English. Unlike Slavic languages, English is impoverished in terms of aspectual morphology. Thus the *in/for* adverbial phrases provide the reader with the necessary information in order to perceive an event as bound or unbounded. Although designing the stimuli in such as manner is beneficial in the sense that it guarantees distinct aspectual readings it gives way to priming effects.

Essentially participants read the same sentence twice, with the only difference being the concluding adverbial phrase. With this said, the stimuli may make the participants susceptible to the following: increased word recall or correctly answered comprehension questions. By being exposed to the same sentence twice, the first exposure unfortunately familiarizes the participant with the sentences' content. Thus upon second exposure, the sentences' content is activated again and in turn strengthens its associated memory trace. This provides participants with an advantage: more working memory resources and space become available for other tasks, such as memory words rehearsal. In addition, multiple exposure to sentence content arguably facilitates correct responses of comprehension questions.

Overall the use of minimal pairs may be partially responsible for the subtle variances observed in proportions of word recall between conditions. However, a question to put forward here is: what other design could be used to create stimuli which tests for telicity? As mentioned prior, this experiment is the first of its kind. Although the current stimuli presents challenges, it does provide a foundation and direction for future research to be carried out.

After taking into consideration the above methodological factors, the results of the experiment appear as weak evidence in support of telicity via maximalization. However, the results clearly indicate that telic interpretations do involve a higher level of complexity than atelic interpretations. Alternatively, if telicity is not realized through competition-based processing, what then is responsible for the increased complexity observed between telic conditions?

3.5.2 Atelicity as Default

A common ground exists between scalar implicatures and maximalization: both possess a scalar mechanism used by the grammar. In the case of scalar implicatures, quantifiers such as *some*, *all*, *many* etc. elicit inferences with varying degrees of informativeness. For example, the following sentence is stated by a speaker:

63) *Some turtles have shells* (Noveck, 2001)

Most likely, we would take (1) to mean *some but not all* turtles have shells. In this case the meaning of *some* is perceived as relatively weak in terms of informativeness, as the stronger claim *All turtles have shells* was not instead stated by the speaker. However, semantic accounts define *some* as *some and possibly all*. With this said, (1) then has a semantic meaning of *some and possibly all turtles have shells*. This meaning is derived in a similar fashion as proposed by Landman (1998) as discussed in section 1.3. The implicature *some but not all* is therefore not lexically derived but inferentially by the listener. This meaning of *some* arguably arises due to a violation of the Gricean Maxim of Quantity: the speaker should make their contribution as informative as possible.

Recent studies have investigated the psychological nature of scalar implicatures, testing if we truly take them into account when processing utterances. To test this, the meanings of *some* predictably differ in terms of the cognitive efforts involved in processing them. The stronger *some and possibly all* is lexically derived and thus serves as the default interpretation. In terms of cost, it is hypothesised that those associated with default interpretations are minimal. In contrast, the weaker *some but not all* is derived via an inference made by the speaker which overrides the semantic meaning. Thus in comparison to the stronger meaning, the weaker *some but not all* has been tested to be cognitively more costly. This distinction should sound quite familiar, as it closely parallels the predictions made by the current experiment.

To test this distinction the sentence-span task was used. Specifically, participants were asked to provide judgments as to what degree sentences such as *Some turtles have shells* is true in lieu of a secondary task. Overall, results illustrate an increase in cognitive resources is associated with the implicatures' pragmatic *some but not all* interpretation but not the logical *some in fact all* interpretation (Noveck, 2001; De Neys & Schaeken, 2007; among others). This evidence argues that the pragmatic *some but not all* interpretation is not default or automatic, as its processing does not survive when cognitive loads increased (Bott & Noveck 2004; De Neys & Schaeken, 2007). Additionally, evidence from developmental behavioural studies with children support the notion that the logical interpretation is defult. Typically developing children ranging from 5 to 10 years old appear to compute scalar implicatures in terms of their associated logical meaning rather than their pragmatic one (Guasti et al., 2005; Papafragou & Musolino, 2003; Noveck, 2001).

Thus the results from the present experiment can be interpreted in another light as the sentence-span paradigm also predicts default processing. Processes which are complex require more cognitive resources to compute, and therefore do not survive when additional strain is put on working memory. When this happens, in the case of scalar implicatures, participants reverted to more logical and less costly computation (Noveck, 2001; De Neys & Schaeken, 2007). Processes which are default or automatic require minimal resources and in turn makes working memory space less susceptible to processing-storage trade-offs. This is exactly what is reflected by the results here: telic sentences were more susceptible to interference in the B.Adv condition as opposed to the SF condition. However, no such interference was observed between conditions for atelic sentences. Thus, regardless of whether maximalization is used for telic sentences, the psychological realization of telic sentences involve a higher level of complexity.

Although insights within the scalar implicature research provide the current results with an alternative interpretation in respect to processing complexity, it does not offer explanation as to what could be responsible for such complexity.

3.5.3 Semantic-Pragmatic Interface

The primary purpose of the thesis investigates the cognitive nature of one component of aspect, namely, telicity in English. In accordance with the predictions made by mereological and scalar theories for the semantic composition of telicity, I hypothesize that telic constructions are semantically more complex than atelic constructions. This complexity specifically refers to maximalization; a process which Filip (2008) predicts takes place exclusively in the derivation of telic events. This prediction is made under the assumption that a correlation exists between the number of derivational steps and cognitive complexity. Specifically, constructions involving a greater number of steps in their derivation also involve a greater amount of cognitive resources in order to be successfully computed. This correlation is supported by the results of the conducted sentence-span task experiment: telic events are computed at a greater cognitive cost than atelic events. However, the results only provide evidence suggesting that a complexity difference exists, specifically that telicity is a grammatically more complex process than atelicity. Crucially, the results from the sentence-span task do not speak to whether the observed difference is due to the process of maximalization.

To test the hypothesis of whether the complexity difference between telic and atelic events is due to maximalization, an experiment using electroencephalography (EEG) could be conducted. EEG is a neuroimaging technique which captures exceptional online neurological responses to stimuli. Specifically, EEG records event-related potentials (ERPs): wavelength signatures reflecting electrical fluctuations in brain activity during various cognitive processes such as language. In the case of language experiments, various lexical items are *time locked* i.e. a procedure where manipulated portions of the stimuli are specifically marked to be associated with the ERP at that time interval (Rodelen & Stemmer, 2008). Based on the excellent temporal resolution EEG provides for processing, this methodology would be ideal for testing *when* the complexity difference in telicity occurs. Predictably, using this methodology would give insight as to whether the complexity associated with telic events primarily unfolds at a local (i.e. semantic) or global (i.e. pragmatic) point in the derivation of the sentence.

The ERP predicted to indicate when the complexity occurs in telic events is the P600. The P600 is a positive wavelength occurring approximately 500 to 1000msec post-stimulus (Osterhout & Holcomb, 1992) which has been indicated to be a marker for structural integration of congruent information (Kaan et al., 2000; Lamb et al., in preparation). Based on this, if more cognitive resources are used to compute and integrate telic events, a P600 will predictably arise reflecting the greater complexity of the derivation. It follows that if the complexity associated with telic events takes places locally within the derivation the P600 is predicted to occur for words within the VP and AdvP portion of the sentence. Likewise, if the complexity is associated with telic events takes place globally, the P600 is predicted to occur after the AdvP.

Following Filip (2008) MAX_E is predicted in English to operate at the AvdP. Based on this, as well as the proposal by Landman (1998) for maximalization, the reference-set computation assumed for telic events is predicted to take place locally. Therefore it is hypothesized that if the complexity associated with telicity is due to maximalization, this will be supported by a P600 occurring at the AdvP. In contrast if a delayed P600 is observed sometime after the sentence has been parsed, this can be interpreted as a *wrap-up effect* and in turn be argued as evidence for maximalization being a global phenomenon.

3.6 Conclusion

Overall, this thesis presents a novel psycholinguistic experiment in which the findings and work established, take the first step in unveiling semantic derivational complexities associated with telicity. Specifically, the experiment and results provide insight that the complexities stem from mereological and scalar theories of event composition.

Based on the derivational predictions made by Filip (2008) for telic and atelic events, it is hypothesized telic events are cognitively more costly to compute compared to atelic events due to the former involving a maximalization process. This hypothesis is made under the assumption that a correlation exists between the number of derivational steps and cognitive complexity. Specifically, constructions which involve a greater number of steps in their derivation also involve a greater amount of cognitive resources in order to be successfully computed. Recent studies using the sentence span task to investigate the role of working memory in sentence comprehension have found constructions with greater derivational complexity are susceptible to interference and processing trade-offs when processed under heavy memory loads. The significant interaction between telicity and memory word recall ultimately supports the hypothesis, as telic trials were more susceptible to decay and interference effects. However, the results do not directly shed light onto whether this complexity difference is due to the maximalization process predicted by Filip (2008). Based on this, and furthered by recent studies on scalar implicatures, the significant interaction of the experiment arguably substantiates atelicity as the default derivation of events in English. In order to investigate whether the complexity associated with telic events is due to maximalization, a follow-up experiment using EEG is proposed. The findings from this experiment would further our understanding of telicity in respect to whether it is primarily a local or global phenomenon in English. In addition, the results from the EEG, would also attest to the assumptions which Filip (2008) bases her account of telicity on.

References

Bach, E. (1981). On time, tense, and aspect: An essay in English metaphysics. In P. Cole (ed.), *Radical pragmatics* (pp. 63-81). New York: Academic Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47-89.

Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417-423.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71.

Bennett, M., & Partee, B. H. (1978). *Toward the logic of tense and aspect in English*. Santa Monica, CA: System Development Corporation. Reprinted 1978, with an afterword, Bloomington: Indiana University Linguistics Club. Also reprinted 2004, in P. Portner (ed.), *Compositionality in formal semantics: Selected papers by Barbara H. Partee* (pp. 59-109). Oxford: Blackwell.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437-457.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*(1), 12-21.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and brain Sciences*, *22*(01), 77-94.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of experimental child psychology*, *33*(3), 386-404.

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford: Oxford University Press.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163-183.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450-466.

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *54*(2), 128-133.

Dowty, D. R. (1979). *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ* (Vol. 7). Springer.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic bulletin & review, 9*(1), 59-79.

Fauconnier, G. (1975). Pragmatic scales and logical structure. *Linguistic inquiry*, 353-375.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, *54*(4), 541-553.

Filip, H., & Rothstein, S. (2006). Telicity as a semantic parameter. In *Formal approaches to Slavic linguistics* (Vol. 14, pp. 139-156).

Filip, H. (2008). Events and maximalization: The case of telicity and perfectivity. *Crosslinguistic and Theoretical Approaches to the Semantics of Aspect*, , 217-256.

Filip, H. (2012). Lexical aspect. *The Oxford handbook of tense and aspect*, 721-751.

Fox, D. (2000). *Economy and semantic interpretation* (No. 35). MIT Press.

Garey, H. B. (1957). Verbal aspect in French. *Language*, 91-110.

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

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, 95-126.

Giorgi, A., & Pianesi, F. (2001). Ways of terminating. *Sematic interfaces. references, anaphora and aspect*. Ed. Carlo Cechetto, Gennaro Cherchia, and Maria Teresa Guasti. Standford, California: SCLI Publications.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425-430.

Guasti, M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes*, *20*(5), 667-696.

Horn, L. R. (1976). *On the semantic properties of logical operators in English*. Bloomington, IN: Indiana University Linguistics Club.

Hovav, M. R. (2008). Lexicalized meaning and the internal. *Theoretical and crosslinguistic approaches to the semantics of aspect*, *110*, 13.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122.

Katzir, R. (2008). *Structural competition in grammar* (Doctoral dissertation, Massachusetts Institute of Technology).

Kennedy, C., & Levin, B. (2008). Measure of change: The adjectival core of degree achievements. *Adjectives and adverbs: Syntax, semantics and discourse*, 156-182.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, 345-381.

Klein, W. (1994). *Time in language*. Routledge.

Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. *Semantics and contextual expression*, *75115*.

Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. *Lexical matters*, *2953*. Tenny, C. 1994. *Aspectual roles and the syntax-semantics interface.* Dordrecht: Kluwer.

Krifka, M. (1998). The origins of telicity. In *Events and grammar* (pp. 197-235). Springer

Netherlands.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of memory and language*, *30*(5), 580-602.

Kučerová, I. (2007). *The syntax of givenness* (Doctoral dissertation, Massachusetts Institute of Technology).

Landman, F. (1992). The progressive. *Natural language semantics*, *1*(1), 1-32.

Landman, F. (1998). Plurals and maximalization. In *Events and grammar* (pp. 237-271). Springer Netherlands.

Lewis, R. (1996). A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research*, *25*, 93-116.

Link, G. (1983). The Logical Analysis of Plurals and Mass Terms: A Lattice-Theoretic Approach.

Link, G. (1987). Algebraic semantics of event structures. In *Proceedings of the sixth Amsterdam colloquium* (Vol. 243, p. 262). ITLI, Amsterdam.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual review of psychology*, *53*(1), 53-81.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165-188.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: an interference model of complex span. *Psychonomic bulletin & review*, *19*(5), 779-819.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253-282.

Parsons, T. (1990). *Events in the Semantics of English* (Vol. 5). Cambridge, Ma: MIT Press.

Pashler, H. E. (Ed.). (1998). *Attention*. Psychology Press.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of experimental psychology*, *58*(3), 193.

Rooth, M. (1985). Association with Focus. University of Massachusetts PhD. dissertation.

Reinhart, T. (1995). *Interface strategies*. OST, Universiteit Utrecht.

Rezac, M. (2006). Escaping the Person Case Constraint: Reference-set computation in the-system. *Linguistic variation yearbook*, *6*(1), 97-138.

Sauerland, U. (2008). Implicated presuppositions. *Sentence and Context. Language, Context and Cognition,–. Berlin, Germany: Mouton de Gruyter*.

Scha, R. (1984). Distributive, collective and cumulative quantification. *Truth, Interpretation, and Information, GRASS*, *2*, 131-158.

Schlenker, P. (2006). Maximize presupposition and Gricean reasoning. *Manuscript, UCLA and Institute Jean-Nicod*.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4.

Sharvy, R. (1980). A more general theory of definite descriptions. *The philosophical review*, 607-624.

Tenny, C. T. (1987). Grammaticalizing aspect and affectedness. *RLE Progress report*, (130), 129-130.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of memory and language*, *28*(2), 127-154.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157-166.

Vendler, Z. (1957). Verbs and times. *The philosophical review*, 143-160.

Verkuyl, H. J. (1972). *On the compositional nature of the aspects* (Vol. 15). Dordrecht: Reidel.

Wanner, E. (1978). Maratsos. M.(1978). An ATN approach to comprehension. *Linguistic theory and psychological reality*, 119-61.

Waters, G., Caplan, D., & Hildebrandt, N. (1987). Working memory and written sentence comprehension.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442.