Linking Form to Meaning: Reevaluating the Evidence for the Unaccusative Hypothesis

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citable link</th>
<th><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:40049976">http://nrs.harvard.edu/urn-3:HUL.InstRepos:40049976</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Linking form to meaning: reevaluating the evidence for the

Unaccusative Hypothesis

A dissertation presented

by

Yujing Huang

to

The Department of Linguistics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Linguistics

Harvard University

Cambridge, Massachusetts

January 2018
© 2018 - Yujing Huang

All rights reserved.
Linking form to meaning: reevaluating the evidence for the Unaccusative Hypothesis

Abstract

This dissertation investigates the linking between meaning and form of intransitive verbs by examining a popular linguistic hypothesis that postulates a one-to-one correspondence between the syntactic position of the argument and the semantic role it takes (the Unaccusative Hypothesis). It shows that the empirical data do not support the hypothesis that there is a binary syntactic distinction. In three eye-tracking experiments, I demonstrate that the processing patterns that were previously argued to support the Unaccusative Hypothesis are not replicable or reliable. Next, I conduct a close investigation of three unaccusative diagnostics and find that these diagnostics do not clearly distinguish unaccusative verbs from unergative verbs. This dissertation argues that the movement account of unaccusativity is neither compatible with the empirical data nor theoretically meaningful in contemporary syntactic theories. It further questions the notion that there is a dichotomy in intransitive verbs and suggests that multiple categories of verbs may better explain the data pattern. This claim needs further investigation which is beyond the scope of this dissertation.
# Table of Contents

Chapter 1: Introduction ........................................................................................................ 1

Chapter 2: Eye-tracking study on English unaccusativity ................................................. 14

Chapter 3: Eye-tracking study on Dutch Unaccusativity ................................................... 48

Chapter 4: Unaccusative Diagnostics in English ............................................................... 73

Chapter 5: Conclusion ........................................................................................................... 115
Acknowledgements

I am grateful to many people who made this dissertation happen. I would like to thank Jesse Snedeker for all the support and encouragement and much more throughout my graduate career. The entire dissertation started with a directed reading course with Jesse, where I learned how to find a research topic, how to design experiments, how to conduct experiments and how to be a researcher. This dissertation would never be possible without her devotion to the research and her dedication to my intellectual growth. All the studies presented in this dissertation are in collaboration with Jesse and Chapter 2 is adopted from a manuscript that we coauthored. Jesse was amazingly supportive in the most difficult times through my years at Harvard. My meeting time with her has been the most cheerful time of the week for a long time. It is hard to express my gratitude to her with words. I would also like to thank Jim Huang. He is the one who introduced me to the field of linguistics and who inspires me with his great passion for research. Jim has always been very respectful and understanding even in the moments where we disagree. He is generous with his support, always brainstorming with me about new arguments and challenging me intellectually. He also gives me all possible respect and the intellectual freedom to work on my arguments. His support gives me the courage to become an independent researcher. His caring personality helped me adjust to the new environment in my first few years of graduate school. I would also like to thank Isabelle Charnavel who discussed the research, provided the feedback and challenged me despite her time constraints. She also broadened my horizons by collaborating with me on topics beyond the scope of this dissertation.

I would like to thank my collaborators Annelot de Rechteren van Hemert and Angeliek van Hout. The Dutch experiment would not have been possible without their help and contribution. I would also like to thank many people who provided comments, critiques and help.
with this work, including Ev Fedorenko, Ted Gibson, Roger Levy, Alec Marantz, Janet Randall
and many others. I would like to thank my colleagues and friends in the Snedeker lab and the
Linguistics Department: Laurence B-Violette, Jenny Lee, Lena Borise, Yimei Xiang, Laine
Stranahan, Chrissy Zlogar, Tanya Levari, Liz Chalmers, Sarah Raulston, Jayden Ziegler, Melissa
Kline, Josh Hartshorne, among others. They offered me tremendous help and support with
research and more. I appreciate their intellectual challenges, help and friendship. I am also
grateful to my research assistants who helped me with the studies.

Finally, I would like to thank my family who provides me with all the support and love I
could ask for and never complained about my absence over the years. I would like to thank my
mom who encourages me to chase my wildest dreams. All my curiosity and interest in learning
and research originate from her. I would also like to thank my aunt and uncle who have provided
tremendous help over the years when I was in graduate school. I would like to thank my
grandma, who could not see me finish this Ph.D. Nevertheless, all the qualities that helped me to
finish, such as optimism, persistence and bravery, come from her.
INTRODUCTION

1. Verb meanings and verb structures

The meanings of verbs have many different facets. Some of them are structurally relevant and others are not. For example, a jumping event requires only one type of participant - the entity that jumps - and therefore the verb *jump* in (1a) takes only one argument. A dropping event requires two types of participants - the entity that does the dropping and the entity that is dropped - and therefore the verb *drop* in (1b) takes two arguments. Some aspects of meaning are irrelevant to the argument structure of the verb. For example, *kill* in (1c) and *die* in (1d) both imply that Mufasa ceases to live; however, this facet of the meaning does not determine how many participants the verb takes or how they are syntactically realized.

(1) a. Chandler jumped.

    b. Chandler dropped the ball.

    c. Scar killed Mufasa.

    d. Mufasa died.

Linguists have worked for many decades to understand what aspects of meaning are relevant to the argument structure of verbs (Fillmore 1968, 1969; Carter, 1976; Jackendoff, 1990; Levin & Rappaport, 2005; Pinker, 2013; a.o.). In addition to the number of arguments a verb can
take, they have discovered that there are some mapping regularities between the syntactic position of the arguments and the roles they play. One such proposal can be traced back to Fillmore (1968), Gruber (1965) and Jackendoff (1972) who each used a small set of thematic roles such as “agent” and “patient” to denote the entity’s relation to the event. For example, the “agent” is the doer of the event and the “patient” is the undergoer. They also discussed some regularity between the roles an argument can take and the positions of these arguments. For example, the subject (external argument) of a verb tends to be the agent (the actor) and the object (internal argument) tends to be the patient (the undergoer). As in (1b) Chandler is the dropper and the ball is the droppee, and in (1c) Scar is the killer and Mufasa is being killed. Developmental research seems to support this observation. Previous studies have found that toddlers as young as 21-months old tend to identify the subject as the agent and the object as the patient (Gertner, Fisher, & Eisengart, 2006).

At the early age of the Government and Binding Theory (Chomsky, 1981), it was assumed that the subject of a sentence is generated at the specifier of TP/IP and the object is generated as the complement of the verb. Therefore, the subject and the object occupy different structural positions in syntax. Within this framework many researchers proposed that there was a one-to-one correspondence between structural positions and thematic roles. One influential example of such a proposal is Baker’s (1988) Uniformity of Theta-Assignment Hypothesis (UTAH), which states that identical thematic relationships between items are represented by identical structural relationships between those items at the level of deep structure. Under this proposal, there is a consistent mapping such that the subject of the sentence is always assigned an agent role and the object is always assigned a theme role (used as an umbrella term for “the entity affected by the action or state expressed by the predicate” including patients, see
Haegeman, 1994). This mapping rule could be very useful in language comprehension. To correctly understand sentences like those in (1), a listener needs to know who did what to whom. If the subject is always the agent and the object is always the theme, figuring out the roles of each argument is trivial. The consistent mapping will make language acquisition easier for children as well. Suppose that the children can understand the event of dropping correctly, they already know that the dropping event involves kinds of two participants (the dropper and the droppee), the only task they have to use the verb drop correctly is to find where to put the dropper and the droppee in their sentence. Once they have this consistent mapping rule, whether acquired or innate, they should be able to produce their sentences with the right structure.

There are some counterexamples to the consistent mapping rule proposed by Baker (1988). For example, for psych-verbs, the subject may not be agent. As can be seen in (2), while the subject John in (2a) is the agent, the subject of (2b) is the experiencer.

(2) a. John frightened Mary.
   b. John feared Mary.

Another counterexample to the consistent mapping rule as stated in Baker (1988) comes from intransitive verbs. Intransitive verbs only take one argument. However, they differ in what kind of thematic roles their sole arguments take. For example, the subject of the verb jump takes an agent role whereas the subject of the verb fall takes a theme role. This can be shown by adding an -er morpheme to the verb to nominalize it as in (3). The subject the boy in (3a) can be a jumper while it is awkward to call the subject for (3b) a faller. This is because the boy in (3a) is agentive and is interpreted as the initiator of the jumping event, whereas the boy in (3b) is interpreted as a theme that undergoes the falling event rather than initiating it.
In this dissertation, I will use intransitive verbs as an example to investigate the mapping regularities between the structure and meaning of verb arguments. Intransitive verbs are a particularly interesting and challenging case for theories about the mapping between form and meaning. Unlike transitive and ditransitive verbs, intransitives have only one argument to be realized. Because many languages (like English) require all sentences to have a subject (the Extended Projection Principle, Chomsky, 1981), this argument must surface as the subject, and we cannot use surface argument position to disambiguate different roles.

To preserve the consistent mapping rule proposed by Baker (1988), researchers hypothesize that there is a difference in the underlying syntactic structure between agent-subject verbs and theme-subject verbs (Unaccusative Hypothesis, Perlmutter, 1978; Burzio, 1981, 1986). For the verbs with an agent subject (the unergative verbs), the subject originates in the subject position and remains there, for example, the boy in (4a). For the verbs with theme subjects (unaccusative verbs), the subject originates from the internal argument position and moves to the subject position in surface structure, leaving a trace in the object position. For example, the boy in (3b) repeated as (4b) originates from the object position and moves to the subject position in surface structure, leaving a trace in the object position.

(4) a. The boy jumped.
   b. The boy, fell t_i.

2. The history of Unaccusative Hypothesis

The above proposal was put forward at the early age of Government and Binding (henceforth GB) Theory in which the subject of a sentence was proposed to originate from the
specifier of TP/IP and the object was proposed to originate as the complement of the verb (Chomsky, 1981). Combining these claims with the Uniformity of Theta-Assignment Hypothesis (Baker, 1988) leads to proposal that the verb assigns the theme role to its complement (the object, which is the internal argument) and the verb and the complement together assign the agent role to its specifier (i.e. the subject, which is the external argument).

Since the introduction of the VP-shell hypothesis (Larson, 1988), elaborated functional heads (Ouhalla, 2003; Pollock, 1989) and the proposal of the VP-internal Subject Hypothesis (Kitagawa, 1986; Koopman & Sportiche, 1991), the proposal above is no longer relevant. In contemporary theories, the subject is no longer assumed to originate from the spec TP/IP. Instead, it originates vP-internally. This diminishes the distinction between internal and external arguments. Also, in the later development of the theory, the object is not the complement of the verb anymore. Therefore, the theme role is not assigned to the complement of the verb (V), but to the specifier of the verb. As the transitive example in (5) shows, the direct object the book (the theme) is the specifier of the verb read and so is the subject John (the agent). In these later developments of the theory, both the subject and the object are base generated VP-internally and take a specifier position. Therefore, the structural distinction between these two arguments is not as clear as in the early GB theory.

One consequence of this development of the theory is that, for intransitive verbs, the sole argument is the specifier of the VP regardless of whether it takes the agent or theme role. One possibility is to assume that there is only one argument position, in which case both the agent and the theme need to be mapped to this position. This means that we have to give up the one-to-one correspondence proposed in UTAH. Another possibility is to assume that the argument position for agent and theme are different. This can be realized by using different functional heads (such
as $v/\text{voice}$) to introduce additional structure. In this account, it is assumed that the subject of an unaccusative verb is introduced by the verb whereas the subject of an unergative verb is introduced by the functional head $v/\text{voice}$ (Chomsky, 1995; Kratzer, 1994) as shown in (6). As can be noticed, (5) is different from (6) in that one object seems to be the specifier and the other the complement. One way to reconcile the discrepancy is to assume that the argument of an unaccusative verb originates from the specifier of the VP and the argument of an unergative verb originates from the specifier of the vP. This slightly different structure bears the same consequence as the one in (6) assuming that vP constitutes a phase and VP does not. As the phase theory is not relevant to the dissertation, I will not discuss it further.

(5)
Under the above account, the arguments of both the unaccusative and unergative verbs still need to move to the specifier of the TP for feature checking. However, this movement is not relevant to the thematic role assignments of the verbs. What causes a thematic difference between unaccusative and unergative verb is the difference in their functional heads. One way of interpreting this account is that each argument is introduced by a head that assigns a unique theta role. Therefore, there is still a one-to-one correspondence between the structure (head position) and the thematic role. However, this one-to-one mapping theory is different from Baker (1988) in two interesting ways. First, it opens up the potential to introduce further functional heads which can assign different thematic roles. This might solve some of the problems that UTAH faces assuming that there are only two positions (the specifier of TP and complement of the verb) to map thematic roles onto. For example, these theoretical resources could be used to explain the syntactic realization of experiencers for psych-verbs. Second, on this account, both unergative and unaccusative subjects undergo movement while in the original Unaccusative Hypothesis only unaccusative subjects undergo movement. These two accounts make different processing predictions as we will see in later chapters.
3. Difference between unaccusative and unergative verbs

As the structural distinction between unaccusative and unergative sentences cannot be seen in the surface structure, how do we know if this structural distinction is real? While we may have theory-internal reasons for positing such a distinction (e.g., to preserve UTAH), we will need independent evidence to support such a proposal. There are two major types of evidence for the original Unacusative Hypothesis. First, linguists have found that unaccusative and unergative verbs appear to differ in the kinds of linguistic constructions that they can appear in. These differences are believed to reflect the structural difference between unaccusative and unergative verbs. For example, these two types of verbs choose different auxiliaries to form the past participles in some languages. Example (7) shows an example from French.

(7)  

a. Il est tombé  
   He is fallen  
   ‘He fell/he has fallen.’

b. Il a souri  
   He has smiled  
   ‘He smiled/he has smiled.’

This pattern in auxiliary selection has been found in many Romance and Germanic languages (Rosen 1984 for Italian, Haider & Rindler-Schjerve 1987 for Italian and German, Zaenen 1988 for Dutch, Legendre 1989 for French, a.o.). It has been argued to distinguish unaccusative and unergative verbs since Perlmutter (1978). Although Levin & Rappaport (1995) pointed out that it is unclear why auxiliary selection should be linked to argument structure of the verbs. It has been claimed that this auxiliary split also happens in other unrelated languages such as Basque (Ortiz de Urbina 1989; Arregi 2004) and old Japanese (Washio 2004). It is hard,
however, to judge whether these phenomena are actually comparable to the phenomena in the Romance and Germanic languages.

Many other phenomena have been argued to be linked to this unaccusative-unergative dichotomy including: whether a prenominal perfect participle is possible when it is used to modify the subject (Zaenen 1993, see (8)); the availability in there-insertion (Burzio 1986, Hoekstra & Mulder 1990, see (9a, (10a)); the causative alternation (Burzio 1986, see (9b,c); (10b,c)); impersonal passivization (Abraham 1986, see (11)); and whether a reflexive is needed after the verb in the resultative constructions (Simpson 1983, see (12)). As we will see, many of these phenomena (such as causative alternation) have been argued to only apply to a subset of the verbs and thus cannot provide clear evidence of an exhaustive two-way split. Thus, while these constructions have been used as arguments to show the distinction between unaccusative and unergative verbs, they cannot be used as straightforward diagnostics for the unaccusative verbs (in each case unaccusativity it is argued to be a sufficient but not necessary condition).

(8) a. the fallen leaf
    b. *the smiled boy

(9) a. There arrived an old man.
    b. An old man arrived.

(10) a. A vase broke.
    b. The boy broke a vase.
(11) a. Er werd gesprongen

There was jumping

b. *Er werd gevallen

There was falling (Dutch, Koring et al. 2012)

(12) a. The river froze solid.

b. The boy sang himself hoarse.

Construction that differentiate unaccusative and unergative verbs have been argued to be present in language families other than Romance and Germanic languages. For example, researchers have argued that this dichotomy exits in Chinese (Huang, 2007). In the following examples, *come* in (13) is believed to be an unaccusative verb, so subject-verb inversion is possible (with a complication that in the subject-verb order, Mandarin needs an existential expression *you* “have” for the sentence to be well-formed). *Cry* in (14) is believed to be an unergative verb, and subject-verb inversion is impossible.

(13) a. lai le san ge keren

come ASP three CL guests

‘There came three guests.’

b. ?san ge keren lai le

three CL guests come ASP

‘Three guests came.’
Although many phenomena have been linked to unaccusativity, it is not clear: (1) why, in some cases, the phenomenon is linked to the argument structure (e.g., auxiliary selection); and (2) whether these phenomena can actually classify all the intransitive verbs into two categories which are consistent across the phenomena and consistent with our intuitions about the thematic role of the subject. This is because most research in this area presents just a few examples and the judgements made by the researchers. There is also theoretical controversy about many of these phenomena and their underlying causes. For example, Zaenen (1993) argues that the auxiliary in Dutch reflects a semantic distinction rather than a syntactic one.

A second type of evidence that has been used to support the Unaccusative Hypothesis comes from studies of language processing. Psycholinguists have found apparent differences in the processing of unaccusative and unergative verbs which could be due to differences in their syntactic structures (Friedmann, Taranto, Shapiro, & Swinney, 2008; Koring et al., 2012; a.o.). I will discuss this evidence in detail in Chapter 2.

To preview the data pattern, these researchers found evidence of reactivation of the unaccusative subject at a post-verbal position. They generally interpret this as evidence for a trace in the unaccusative sentences as proposed in the original Unaccusative Hypothesis. Under the original account, we would only expect the reactivation of the unaccusative subject. This is
different from the predictions of more recent syntactic theories. In the more recent account for unaccusativity, both unaccusative and unergative sentences have movement and trace positions. The predictions for this account could differ depending on the details of the theories. If we assume that both unaccusative and unergative subjects move once, then unaccusative and unergative verbs are probably not distinguishable in processing. We would expect both subjects to be reactivated around the same time. If we assume that the unaccusative subjects undergo a two-step movement (perhaps due to the minimality requirement which leads them to first move to Spec, vP) then the unaccusative sentences are more complex in structure. If differences in structural complexity result in differences in the timing of argument activation (as discussed in Koring et al. 2012), then we might expect the reactivation of unaccusative subject to happen later (assuming that the experimental method has enough temporal resolution to capture this difference). This means that the unergative subjects should be reactivated earlier than the unaccusative subjects. However, as has been shown in Levy (2008), processing time does not appear to closely reflect structural complexity but instead depends largely on the predictability of a word or structure and hence its frequency. I will return to Koring’s data in Chapters 2 and 3.

In the following chapters of this dissertation, I will examine (1) whether the empirical evidence supports the claim that there is a dichotomous distinction in intransitive verbs; (2) and if so, which version of the syntactic theories are the data compatible with. This dissertation aims to shed light on argument realization in intransitive verbs by bringing in new empirical data from my own research. In Chapters 2 and 3, I will present data from eye-tracking studies to investigate whether there is a processing difference between unaccusative and unergative verbs. In Chapter 4, I will present judgment data from the unaccusative diagnostics to examine whether these phenomena can truly distinguish between unaccusative and unergative verbs. In Chapter 5, I will
summarize the findings and reevaluate the two versions of the Unaccusative Hypothesis and the alternative semantic explanation, in light of this new data.
Chapter 2

EYE-TRACKING STUDY ON ENGLISH UNACCUSATIVITY

1. Introduction

Researchers have been using experimental methods to examine whether there is a processing difference between unaccusative and unergative sentences. These studies have been constructed under the assumption that there is a movement in the unaccusative but not unergative sentences (the early GB account) (e.g. Bever & Sanz, 1997; Burkhardt, Piñango, & Wong, 2003; Friedmann, Taranto, Shapiro, & Swinney, 2008; Koring, Mak, & Reuland, 2012; a.o.). If such a processing difference is found, it would be strong support for the structural difference between unaccusative and unergative verbs under the movement account.

In one of the early studies on Unaccusativity Hypothesis, Bever and colleagues tested the trace facilitation effect with reaction time method (Bever & Sanz, 1997). They hypothesized that because there is a trace position in the unaccusative and not the unergative sentences, the recognition of the subject of an unaccusative verb at a post-verbal position should be faster than that of an unergative verb because the trace of the unaccusative subject facilitates the processing. To test whether there is truly a trace facilitation effect, they asked the participants to scan a linguistic sequence on the screen. After the sequence disappeared, a probe word was presented to the participants who decided whether they had seen the word in the sequence or not. They found that, for the participants who are sensitive to the syntactic structure of the sentence, if the probe was the subject of an unaccusative sentence, the reaction time for the judgement was shorter than
if it was the subject of an unergative sentence. They argued that the shorter reaction time is the evidence for the trace facilitation. However, the link between the trace position and shorter reaction time is not clearly driven by a syntactic difference. One can argue that unaccusative verbs and unergative verbs differ in many ways such as taking difference semantic roles. The meanings of the verbs in these two categories also differs systematically. For example, many unaccusative verbs are telic (break, arrive, etc.) and many unergative verbs are not (laugh, cry, etc.). We cannot tell whether the difference in reaction time is caused by a structural difference or a semantic difference that may not even be relevant to the argument structure. Like Bever & Sanz, the offline studies generally cannot build a necessary link between their results with a trace position. Therefore, they cannot serve as evidence for the Unaccusative Hypothesis.

Studies with better temporal resolution and continuous measurements may be more promising methods to justify a structural difference between unaccusative and unergative sentences by finding a processing difference at the trace position. That is, if there is a trace of the subject at the object position, we expect the subject to be reactivated at that position. Therefore, by testing the reactivation of the subject after the verb, we should be able to test whether there is syntactic movement involved in the unaccusative sentences.

In a cross modal priming study, Friedmann and colleagues found a reactivation of the subject after unaccusative verbs but not after unergative verbs (Friedmann et al., 2008). In this study, participants were asked to attend to the aurally presented sentence and to make a lexical decision (word/nonword) on a letter sequence via a button press. The measurement was the reaction time for making the decision. If the subject is a semantic associate to the word in the lexical decision task, thinking of the subject will facilitate the decision, and therefore, the reaction time should be shorter. For example, if a participant hears the sentence (15) and the
probe is the word *map*, they should be faster to respond compare to a probe word *rod*. Participants are asked to make this decision in different position of the sentence (① ② and ③). If at any position, the reaction time is shorter, it suggests that the subject is reactivated at that position.

(15) The geographer ① from East Orange, New Jersey, mysteriously disappeared ② when it was ③ time to adjust the tuxedos and dresses for the participants in the wedding party.

They found that the subject was reactivated 750 ms after the verb onset (position ③) in unaccusative sentences but not in unergative sentences. They interpreted this as reactivation of the subject at the trace position. Their finding is consistent with the hypothesis that the subjects of the unaccusative verbs undergo movement and those of the unergative verbs do not. In contrast, Burkhardt et al. (2003) used a similar cross-modal lexical priming paradigm and found a reactivation of unaccusative subjects 650 ms after the offset of the verb, and a reactivation of unergative subjects 100 ms after verb offset.

In an eye-tracking study, Koring et al. (2012) used the Visual World Paradigm and found a different reactivation pattern. In their study, participants viewed a visual display while listening to an auditory sentence. Based on prior research, they assumed that the participants eye movements would be affected by the concepts they entertain. If they hear the word “geographer” and there is a picture of geographer on the screen among other pictures, then they should look to the picture of the geographer. This is also true when the picture is not the word but a semantic associate (e.g. geographer and map). To link this to reactivation: if a concept is reactivated, participants will also fix their eye gaze to the relevant picture. Therefore, we can use it as a clue
for reactivation (i.e. there is a trace). Like Friedmann et al. (2008) and Burkhardt et al. (2003), Koring and colleagues found a late reactivation at 750 ms after verb offset for unaccusative verbs. In contrast with Freidmann et al (2008) but similar to Burkhardt et al. (2003), they found an early activation near verb onset for unergative verbs. However, this effect appeared much earlier than the one in Burkhardt study, emerging shortly after verb onset.

These apparent differences in processing are theoretically critical. If there is a systematic delay in the reactivation or integration of unaccusative subjects, it would lend support to the early GB theories, like the Unaccusative Hypothesis, which propose fundamental differences in the structure and complexity of these verbs. There are, however, reasons to hesitate before accepting these conclusions. First, across these three studies there are subtle differences in the time course of reactivation that are difficult to reconcile. If unergative subjects are activated shortly after verb onset (Koring) and 100 ms after verb offset (Burkhardt) then why aren't they activated at verb offset (Friedmann)? How do we account for studies which fail to find processing differences between the verb classes (Flett, 2006)? In this chapter, I present two experiments, where we sought to replicate the Koring findings in English to verify the processing pattern and ensure that findings are stable enough to build a theory on.

In replicating the Koring study, we wanted to rule out a range of potential confounds which were not controlled in the previous experiments. Unergative and unaccusative sentences necessarily differ in their verbs. In the prior studies, these sentences also differed in the subject nouns that were used and thus in the primed pictures or probe words (Burkhardt et al., 2003; Friedmann et al., 2008; Koring et al., 2012). To conclusively link the processing differences between the two sets of stimuli to the two verb class, we must rule out other potential confounds. The authors of all three studies were aware of this problem and carefully matched their stimuli
on several relevant features. However, other factors which are known to play a role in language processing were not matched, opening up the door to other potential interpretations of their findings. For example, unergative verbs tend to be more imageable than unaccusative verbs. To explore the role that this might play in these experiments, we asked participants to rate the imageability of the verbs in the published stimulus sets for Friedmann and Koring. We found a reliable difference in imageability consistent with our hypothesis (Friedmann N = 28, unaccusatives = 4.10, unergatives = 5.65, p<0.01; Koring N = 28, unaccusatives = 4.37, unergatives = 5.97, p<0.01). This factor alone might account for the prior findings: more imageable words are recognized more quickly (see Paivio, 1991 for review) and thus we should expect faster processing for the unergative verbs in the previous experiments, potentially resulting in more rapid reactivation of the subject.

Other confounds could also play a role. For example, the unergative sentences in the previous studies were more likely to have animate subjects (100% vs. 67% in Friedmann and 100% vs. 40% in Koring, Fisher’s exact test p’s < 0.05). This could make these nouns more salient and easier to retrieve. We also worried about the role of uncontrolled factors that are more difficult to quantify, such as the nature of the picture or probe word that indexes subject reactivation or the complexity of the critical sentence after the verb. In our study, we eliminated all of these confounds in two ways: First, we carefully normed features of the verb and its fit with the subject noun with ratings studies. Second, we eliminated confounds related to the subject noun, the sentence continuation, and the pictures by using the same stimuli (counterbalanced) across verb classes.

Our final goal in this study was to explore the stability of the growth curve analyses used by Koring. Growth curve analyses model changes over time. Thus, they seem like a promising
method for linking our cognitive theories to the rich data provided by eye-tracking. Perhaps for this reason, the use of growth curve analyses in visual world studies seems to be spreading quickly (see Mirman & Magnuson, 2009; Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Brown, Salverda, Dilley, & Tanenhaus, 2011; Lee, Middleton, Mirman, Kalénine, & Buxbaum, 2013; Hadar, Skrzypek, Wingfield & Ben-David, 2016; Pozzan, Gleitman & Trueswell, 2016; Cane, Ferguson, & Apperly, 2017; a.o.). In looking closely at the analysis in Koring, we noted three potential problems. First, the analysis uses a linear linking function, thus it assumes that the error is normally distributed on a linear scale. This idealization is false in the limit with eye-tracking data because the behavior itself is binary (i.e. at a given time point, the participant either looks at the target picture or not). Whether this results in a substantial violation of the assumption of normalcy is an empirical question. Second, the Koring analysis averages across different items rather than treating them as a random effect (see Clark, 1973 for discussion of item effects). Third, Koring's analysis provides multiple places where a difference between the verb classes could emerge. Since no correction is made for multiple comparisons this procedure could increase the probability of false positive. Looking closely at these particular models may help us understand the advantages and perils of growth curve modeling for visual world data more generally.

2. The current study

The current design is similar to Koring but with a more careful control of potential confounds discussed in the previous section. During the study, the participants viewed a visual display while listening to an auditory sentence. Previous research has shown that the participants’ eye movements are affected by the concepts they entertain (Altmann & Kamide, 2004; a.o.). If they hear the word “geographer” and there is a picture of geographer on the screen among other
pictures, they should look at the picture of the geographer more than other pictures. This is also true when the picture is not the word but a semantic associate (e.g. geographer and map) (Yee & Sedivy, 2006). Koring found that if a concept is reactivated, the participants will also fix their eye gaze to the semantically associated picture. Therefore, we can use it as a clue for reactivation. To be more specific, based on Koring, when the participants hear the verb, if they reactivate the subject of the sentence, they will look at the picture that is related to the subject. The timing of this fixation is influenced by the structure of the verb, meaning that it should differ between unaccusative and unergative verbs.

2.1. Experiment 1

2.1.1. Method

Subjects

In this study, 40 monolingual native English speakers are recruited from the Harvard community for either course credit or $5 payment. They all reported that they have normal or corrected to normal vision and normal hearing.

Materials

Selection of verbs

All unaccusative verbs in this study are non-alternating verbs, thus we excluded verbs (for example, break in (16) which can be both unaccusative and causative. This is because there is theoretical controversy about whether alternating and non-alternating unaccusative verbs involve the same syntactic mechanism (Chierchia, 2004) and previous research suggests that the alternating category may behave differently from non-alternative unaccusative in online processing (Friedmann et al., 2008).
(16)  
a. The vase broke.

b. The boy broke the vase.

13 of the 20 unaccusative verbs and 15 of the 20 unergative verbs from our study were taken from previous studies where a difference of these two categories were claimed to be detected (Friedmann et al., 2008; Koring et al., 2012; Agnew, van de Koot, McGettigan, & Scott, 2014). The rest of the verbs are classified in the same subcategory as at least one of above verbs in VerbNet 3.2 (Kipper, Korhonen, Ryant, & Palmer, 2008) or Levin (1993). The new unaccusative verbs in our study meet the criteria for unaccusative verbs in Friedmann et al. (2008), i.e. (1) ability to be used in there-insertion construction, (2) ungrammaticality with a direct object, and (3) inability to undergo passivization. The new unergative verbs are all unambiguously intransitive and meet the criteria for unergative verbs in Friedmann et al. (2008), i.e. (1) ungrammaticality in the there-insertion construction, (2) ungrammaticality in the resultative construction, and (3) inability to occur with a reflexive pronoun unless the reflexive pronoun is followed by a resultative.

The lemma frequencies of unaccusative verbs and unergative verbs were determined by Corpus of Contemporary American English (COCA; Davies, 2008). The Log transformed mean frequencies of unaccusative and unergative verbs (3.53 and 4.15 respectively) do not differ significantly ($t=-1.45$, $p=0.16$). The imageability of the verbs was determined by an Amazon Mechanical Turk (AMT) norming study. The mean imageability of unaccusative and unergative verbs (4.13 and 3.84 respectively) do not differ significantly (Mann-Whitney-Wilcoxon, $p=0.29$).
All of our test and control sentences have the same structure:

<table>
<thead>
<tr>
<th>Subject</th>
<th>Prepositional Phrase</th>
<th>Modifier</th>
<th>Temporal Clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cookie monster said that</td>
<td>the geographer</td>
<td>with a loud voice and quick temper</td>
<td>suddenly</td>
</tr>
</tbody>
</table>

With the pairing of the subjects and the pairing of the verbs, we manage to keep the rest of the stimuli invariant and therefore have a better control of unaccusative and unergative conditions. The rationale to add the modifier to the argument is to assure that the activation will decay after the participants hear the argument (c.f. Koring et al., 2012). We also included an adverb before each verb, which is used on only some of Koring et al. stimuli. The reasons are twofold. First of all, some of our sentences sounded more natural with an adverb. More importantly, in Koring et al. (2012), an early reactivation of the argument is detected immediately after the verb onset. According to the integration story, this effect could be due to the fact that many of their sentences have adverbs. That is, when the participants hear the adverbs, they know there is an upcoming verb and therefore prepare to integrate the argument with the verb by activating the argument. Since our goal was to find out whether their data pattern could be replicated (in English and with greater control over extraneous factors), we decide to include adverbs in our sentences as well. To be consistent, we added an adverb in every sentence frame. The modifiers and the adverbs together were 11-20 syllables long (mean=15.12). Neither the modifier nor the adverb was semantically related to the critical argument, the target image, the verb or the distractors. As reactivation in unaccusative sentences was found about 750 ms after the verb onset in the previous studies (Friendmann et al., 2008; Koring et al., 2012) and researchers found that reactivation may occur at the end of sentence for reasons independent of
syntax (Balogh, Zurif, Prather, Swinney, & Finkel, 1998), we added a temporal clause in order to
detect the reactivation and to separate it from sentence wrap up effects. We added a prepositional
phrase after the verbs in those sentences where the verb required it due to selectional restrictions.
The temporal clauses and the post-verbal prepositional phrases together were 10-19 syllables
long (mean=13.3). Neither of the temporal clause nor the prepositional phrase was semantically
associated with the critical argument, the target image or the distractors. We also made sure that
the verbs were not semantically related to either the critical argument or the target image.

All the subject nouns were terms for occupations and hence animate. We paired our verbs
and placed them in the same sentence frame so that in each pair, the only difference between the
sentences was the verb. For each pair of verbs there were two frames and two different subject
nouns so that the same pictures could be used as semantically related and unrelated items.

In each trial, there are four images on the screen. The images are all black and white line
drawings from Bank of Standardized Stimuli (BOSS, Brodeur, Guérard & Bouras, 2014), Bonin,
Peereman, Malardier, Méot, and Chalard (2003) and Szekely et al. (2004).

We conducted an AMT norming study to control for the relatedness between the critical
images and critical arguments. This is a relatedness judgment test. According to the results of this
test, the mean relatedness between the argument and the related image is 4.72 on a scale from 0
to 5 and the relatedness between the argument and the paired non-related image is 0.36 on a scale
from 0 to 5.

In the test condition, all the critical arguments have an image that has a schematic relation
to it (for example, geographer and map in Table 1). When the participants hear sentences with
geographer, they see a map as in Figure 1. In the control condition, there is no image that is
related to the critical argument. For example, a rod substitutes the map and the other three
images stay unaltered as in Figure 1. The image that is not related to the argument is related to the alternate argument for that pair of verbs (for example, the rod is related to the conductor). Therefore, when the participants hear sentences with *conductor*, the left image in Figure 1 becomes the control condition and the right image in Figure 1 becomes the test condition.

### Table 1 illustration of sample stimulus

<table>
<thead>
<tr>
<th>Verb</th>
<th>Sentence</th>
<th>Match</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>unaccusative</td>
<td>Cookie Monster said that the <strong>geographer</strong> with a loud voice and quick temper suddenly <em>fell</em> when the boat lurched violently because of the storm.</td>
<td>map</td>
<td>rod</td>
</tr>
<tr>
<td>unergative</td>
<td>Cookie Monster said that the <strong>geographer</strong> with a loud voice and quick temper suddenly <em>screamed</em> when the boat lurched violently because of the storm.</td>
<td>map</td>
<td>rod</td>
</tr>
<tr>
<td>unaccusative</td>
<td>Cookie Monster said that the <strong>conductor</strong> with a loud voice and quick temper suddenly <em>fell</em> when the boat lurched violently because of the storm.</td>
<td>rod</td>
<td>map</td>
</tr>
<tr>
<td>unergative</td>
<td>Cookie Monster said that the <strong>conductor</strong> with a loud voice and quick temper suddenly <em>screamed</em> when the boat lurched violently because of the storm.</td>
<td>rod</td>
<td>map</td>
</tr>
</tbody>
</table>

![Figure 1](image)

With this design, we ensure that the sentence frames and the visual stimuli are exactly the same between the unaccusative and the unergative conditions and across the test and control
conditions. Therefore, any differences we find between unaccusative and unergative conditions will be due to the differences between these two verb categories rather than uncontrolled differences between the pictures, arguments or sentence frames.

We normed the average naturalness of the sentences with an AMT norming study and found no significant difference (Mann-Whitney-Wilcoxon Test, p=0.54) between sentences with unaccusative verbs and sentences with unergative verbs (5.28 and 5.19 respectively). We also controlled the plausibility of the verb given the argument as the subject for that verb (Mann-Whitney-Wilcoxon Test, p=0.12).

We created four lists so that each participant only hears the same sentence frame and sees the same visual stimuli once. That results in twenty sentences in each list. In order to have the same number of sentences as Koring et al. (2012), we created two different sentence frames and two sets of visual stimuli for each pair of the verbs. If the unaccusative verb in a pair is selected for sentence frame 1, then we will use the unergative verb in the same pair for sentence frame 2 in the same list so that the participant does not see the same verb twice. Our filler items, like those in Koring et al. (2012), had transitive verbs and were paired with displays in which one image depicted the subject of the sentence. There are 40 fillers which remained the same across all the four lists, resulting in a total of eighty sentences in each list.

The audio stimuli were recorded at a normal speaking rate by a female native speaker of English, sampled at 44.1 kHz.

2.1.2. Procedure

Our procedure was closely based on Koring et al. (2012). Participants were seated comfortably in front of a monitor. Their eye movements were measured by a Tobii T60 sampling at 60 Hz. Each session started with a calibration procedure with seven fixation points.
Participants were told that they would hear some sentences and look at some pictures. They were told to listen to the sentences carefully in order to answer some questions at the end of the study. To reduce noise in the data, each trial started with a centrally-located fixation dot. Participants were instructed to look at the dot briefly. There was a 1 second preview of the display before the onset of the spoken sentence. After the end of the spoken sentence, there was another 1 second of silence before the fixation dot appeared. The entire experiment lasted about 20 min. At the end, participants were given a memory test with 32 sentences. Some of the sentences were the audio stimuli and some were not. Participants need to identify the sentences they heard in the study.

2.1.3. Results

In this section, we first analyze the data with a growth curve analysis as Koring et al. (2012) did. Then, we compare the results from the Growth Curve models with a logistic mixed effect model which is widely used in analyzing eye-tracking data. This will enable us to compare the stability and informativeness of the models as well as the consistency of the results between the growth curve analysis and the traditional regression model. To preview our results: with the growth curve analysis, several terms were significant in our two replications of Koring et al. (2012), however, several terms and the directions of the effects were different; in the logistic regression, the two replications both gave negative results, meaning that we did not find an effect between our unaccusative and unergative conditions. Therefore, in both analyses, we did not replicate the results of Koring et al. (2012).

Growth curve model

Following Koring et al (2012), we analyzed our data with growth curve models to make the results comparable to their results. During the study, we record the eye-fixation of the participants. At each time point, we can calculate the proportion of looks to the target image. As
we continuously record eye fixation, we will have a curve that resembles the change of the proportion over time such as in Figure 2. In the growth curve analysis, the shape of the curves is modeled and therefore, we can compare two curves in different experimental conditions. Therefore, Mirman, Dixon, and Magnuson (2008) have argued that the growth curve analysis can give us information on the change of eye fixation proportions in a given time window.

Like Koring et al. (2012), instead of using the raw score, we used the differential score (the proportion of looks to the target picture in the test condition minus the proportion of looks to the target picture in the control condition). We first had only time polynomial factors as predictors and gradually added the condition and interaction on the different terms. Orthogonal power polynomials are included as predictors for higher-order time terms. The dependent variable is the difference score.

We analyzed our data in two time windows defined by Koring et al. (2012). The first window, i.e. the verb frame, starts 600 ms before verb offset and ends 1000 ms after verb offset, taking the verb offset plus 200 ms as the mid-point (as it takes about 200 ms to program eye movements; Altmann & Kamide, 2004; a.o.). The second window, i.e., the post-verb frame, starts 200 ms after verb and ends 1700 ms after verb.

Verb frame

In this window, the difference score is modeled by condition, linear term, quadratic term and their interactions with condition at individual participant level (i.e. participants as random effect). Unergative is taken as the baseline condition in the analysis. The results of model comparison are summarized in Table 2. As can be seen, gradually adding the condition and its interaction with the higher order time terms improves the model.
Figure 2 Fitted lines for the differential looks to the target in the two conditions

Table 2 model comparison (random slopes in all models includes all time terms in the fixed effect)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Chisq</th>
<th>p&lt;</th>
<th>AIC</th>
<th>Chisq</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>-1122.30</td>
<td></td>
<td></td>
<td>-2761.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>linear+quadratic</td>
<td>-1253.70</td>
<td>139.36</td>
<td>.001***</td>
<td>-3194.00</td>
<td>440.38</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+condition</td>
<td>-1254.00</td>
<td>2.28</td>
<td>.1</td>
<td>-3193.70</td>
<td>1.74</td>
<td>.19</td>
</tr>
<tr>
<td>linear+quadratic+condition+linear*condition</td>
<td>-1264.80</td>
<td>12.78</td>
<td>.001***</td>
<td>-3192.30</td>
<td>0.56</td>
<td>.45</td>
</tr>
<tr>
<td>linear+quadratic+condition+linear<em>condition+quadratic</em>condition</td>
<td>-1266.40</td>
<td>3.60</td>
<td>.06</td>
<td>-3209.60</td>
<td>19.26</td>
<td>.001***</td>
</tr>
</tbody>
</table>

We choose to report the results of quadratic model, i.e. the same model used in Koring et al. (2012). The results are summarized as Table 3. However, this model does not turn out to be significantly different from the linear model as can be seen in Table 2. Different from Koring et al. (2012), we have a significant interaction of condition and linear term but no significant
interaction of condition and quadratic term. In Koring et al. (2012), the unergative has a significant negative quadratic component, which means that there is a rise followed by a fall. They take it as the evidence for early reactivation. The unaccusative in their study has a positive quadratic component, i.e. a fall followed by a rise. As we can see in Figure 2, this is not the pattern we have in our data. In our experiment, both unergative and unaccusative follows a rise-and-fall pattern, the only difference is the shape of the curve which is not significant in our model.

Table 3

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Experiment 1</th>
<th>Koring et al (UE-UA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.26</td>
<td>7.15</td>
</tr>
<tr>
<td>Linear*condition</td>
<td>-0.17</td>
<td>0.05</td>
</tr>
<tr>
<td>Quadratic*condition</td>
<td>-0.09</td>
<td>-1.90</td>
</tr>
</tbody>
</table>

Post-verb frame

![Post-verb Frame](image)

Figure 3 Fitted lines for the differential looks to the target in the two conditions
In this window, the difference score is modeled by condition, linear term, quadratic term, cubic term, quartic term and their interaction with condition at individual participant level. Unergative is taken as the baseline condition in the analysis. As can be seen in Table 4, gradually adding the condition and its interaction with the higher order time terms does not consistently improve the model. To compare the results with those of Koring et al., we choose to report the results from the quartic model with all interactions. However, as we can see, this model is not significantly better than the model without interaction of the quartic term and condition.

Table 4 model comparison (random slopes in all models includes all time terms in the fixed effect)

<table>
<thead>
<tr>
<th>Model</th>
<th>Experiment 1</th>
<th></th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC</td>
<td>ChiSq</td>
<td>p</td>
<td>AIC</td>
<td>ChiSq</td>
<td>p</td>
</tr>
<tr>
<td>linear+quadratic+cubic</td>
<td>-760.70</td>
<td></td>
<td></td>
<td>-2026.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic</td>
<td>-829.40</td>
<td>80.64</td>
<td>.001***</td>
<td>-2168.60</td>
<td>154.55</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition</td>
<td>-830.57</td>
<td>3.21</td>
<td>.07</td>
<td>-2171.50</td>
<td>4.92</td>
<td>0.05*</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear*condition</td>
<td>-833.04</td>
<td>4.47</td>
<td>.05*</td>
<td>-2170.50</td>
<td>0.96</td>
<td>.33</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear<em>condition+quadratic</em>condition</td>
<td>-837.43</td>
<td>3.63</td>
<td>.05*</td>
<td>-2185.30</td>
<td>16.82</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear<em>condition+quadratic</em>condition+quartic*condition</td>
<td>-838.31</td>
<td>2.89</td>
<td>.09</td>
<td>-2186.20</td>
<td>2.87</td>
<td>.1</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear<em>condition+quadratic</em>condition+quartic<em>condition+quartic</em>condition</td>
<td>-836.41</td>
<td>0.09</td>
<td>.76</td>
<td>-2184.60</td>
<td>0.40</td>
<td>.53</td>
</tr>
</tbody>
</table>
Similar to Koring et al (2012), we found a significant interaction between the linear component and the condition \((p<0.05)\). Different from Koring et al., we found a significant interaction of quadratic term and condition, but the interaction between the quartic term and condition is not significant. All other factors are not significant in both studies (all \(p\)'s>0.05), as can be seen in Table 5.

### Table 5

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Experiment 1</th>
<th>Koring et al (UE-UA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>(t)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.24</td>
<td>6.94</td>
</tr>
<tr>
<td>Linear*condition</td>
<td>-0.05</td>
<td>2.11</td>
</tr>
<tr>
<td>Quadratic*condition</td>
<td>-0.09</td>
<td>2.53</td>
</tr>
<tr>
<td>Cubic*condition</td>
<td>0.08</td>
<td>-1.70</td>
</tr>
<tr>
<td>Quartic*condition</td>
<td>0.17</td>
<td>0.31</td>
</tr>
</tbody>
</table>

**Logistic mixed effect model**

The last section showed that we did not replicate the results of Koring et al. (2012) in either window. Rather, the results differ in many terms. In this paper, we analyze the data with Logistic mixed effect models which is a more standard analysis for visual world paradigm data. In this analysis, an average proportion of looks to the target is calculated in a given window (therefore, instead of having multiple data points in a window like the growth curve analysis, we have a single data point in each experimental condition per subject per item). We compare the proportion with a predetermined threshold. If the number is larger than the proportion, we code the dependent variable as 1 which means that the participant is looking at the target image.
Otherwise, we code it as 0. As our dependent variable is binary, we use logistic regression to analyze the data.

To clean up the data, we removed trials with high track-loss, i.e. where the validity of more than half of the samples is higher than 3 (higher numbers indicate lower validity). The two analysis windows are the same as those of the growth curve models.

Prior to the noun onset, there is no difference among the four conditions. The proportion of looks to the target is around chance (25%) for all four conditions. After the noun onset, there is an increase to the target picture which shows the initial activation of the noun.

Figure 4 shows the average looks to the target picture in the verb and post-verb windows. Within each window of each trial, an average look to the target is calculated. If in that time window, the participant looked at target more than 30% of the time, it is determined that they looked at the target. Otherwise, they are categorized as not looking at the target (1 vs. 0). The average look in each region is the mean of this binary variable for all the trials.

![Figure 4](image)
As can be seen in Figure 4, during the verb and post-verb regions, the looks to the target picture are higher in the match condition (e.g. the subject is the librarian and the target is the book cart) than the mismatch condition (e.g. librarian vs. telescope). This verifies that participants looked at the target picture more when they heard a related subject compare to the condition where there is no matching noun.

We analyzed the data with a logistic mixed effect model, whether the participants look at the target in each time window in each trial is the dependent variable and the verb type (unaccusative vs. unergative coded as 1 and -1), sentence region (baseline, noun, verb, early post-verb and late post-verb) and match condition (match vs. mismatch coded as 1 and -1) and all interaction terms are included as fixed effects and participant and item as random intercepts. The results are summarized as Table 6 below.

Table 6

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Verb frame</th>
<th></th>
<th></th>
<th>Post-verb frame</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>z</td>
<td>p&lt;</td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.23</td>
<td>0.10</td>
<td>-2.21</td>
<td>.05*</td>
<td>-0.25</td>
<td>0.10</td>
</tr>
<tr>
<td>Match</td>
<td>0.59</td>
<td>0.09</td>
<td>6.36</td>
<td>.001***</td>
<td>0.52</td>
<td>0.09</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.01</td>
<td>0.05</td>
<td>-0.27</td>
<td>.79</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>Match*Condition</td>
<td>0.02</td>
<td>0.04</td>
<td>0.44</td>
<td>.66</td>
<td>-0.01</td>
<td>0.04</td>
</tr>
</tbody>
</table>

For the verb region, the model with verb type and match condition and the interaction term as the fixed effects, verb type and match condition as the random slopes and participant and item the random intercepts has a better fit than the simpler models (p<0.001). Therefore, we chose to report the result from this model. There is a reliable effect of match condition on the
looks to the target (p<0.001). The positive coefficient shows that the looks to the target picture is higher in the match condition than the mismatch condition during the verb region. There is no effect of verb type (p=0.79), and critically there is no interaction between verb type and match (p=0.66) and thus no evidence for a difference in the activation of the subject across the two verb types.

The same model was run in the post-verb region. There is a significant effect of match, (p<0.001), and of verb type (p<0.05), and no interaction (p=0.81).

In summary, no difference was found between unaccusative and unergative condition in the analysis windows when match is controlled for. As can be seen in Figure 4, in the post-verb frame, the average of looks to the target is higher in the unaccusative condition than that of the unergative condition. However, critically, there is no interaction between the verb type and match, which means that this difference is regardless of the presence of a target picture but when there is a target picture, no difference is found between the unaccusative and unergative conditions.

2.2. Experiment 2

Experiment 2 is similar to Experiment 1 with minor changes specified below.

2.2.1. Method

Subjects

Sixty monolingual native English speakers were recruited from the Harvard community for either course credit or $5 payment. They all reported that they have normal or corrected to normal vision and normal hearing.
Materials and procedure

The procedure and stimuli are the same as Experiment 1 except that 16 questions about the sentences or the pictures are interspersed among the test sentences. For each question, two choices were provided. Participants need to use the touch screen to select an answer. Once a choice is made, the study will proceed. The questions are interspersed to distract the participants from the target picture. This is because in the debriefing procedure of Experiment 1, several participants mentioned that they think the study is about the target picture and deliberately fixed their eyes on those pictures.

2.2.2. Results

Growth curve

The analysis follows the same procedure as that of Experiment 1.

Verb Frame

![Verb Frame](image)

Figure 5 Fitted lines for the differential looks to the target in the two conditions
In the verb frame, the difference score is modeled by condition, linear term, quadratic term and their interaction with condition at individual participant level. Unergative is taken as the baseline condition in the analysis.

As can be seen in Table 7, adding the interaction of the quadratic term and the condition improves the model and it is the model that Koring et al. (2012) used. Therefore, we choose to report the results from this model.

Table 7

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>T</th>
<th>p&lt;</th>
<th>Estimate</th>
<th>t</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.11</td>
<td>5.25</td>
<td>.001***</td>
<td>-0.08</td>
<td>-2.21</td>
<td>.05*</td>
</tr>
<tr>
<td>Linear*condition</td>
<td>0.03</td>
<td>0.74</td>
<td>.46</td>
<td>-0.02</td>
<td>-0.10</td>
<td>n.s.</td>
</tr>
<tr>
<td>Quadratic*condition</td>
<td>-0.16</td>
<td>-4.39</td>
<td>.001***</td>
<td>0.20</td>
<td>7.89</td>
<td>.001***</td>
</tr>
</tbody>
</table>

There is a no main effect of condition (p=0.19) or interaction of condition and linear time term (p=0.46). There is a significant interaction of quadratic term and condition (p<0.001).

In this window, the interaction of quadratic term and condition in Koring et al. (2012) is replicated. However, different from the result of Koring et al. (2012), the sign of this term is different (see Table 7).
Post-verb frame

![Post-verb Frame](image)

Figure 6 Fitted lines for the differential looks to the target in the two conditions

In the post-verb frame, the different score is modeled by condition, linear term, quadratic term, cubic term, quartic term and their interaction with condition at individual participant level. Unergative is taken as the baseline condition in the analysis.

Table 8

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Experiment 2</th>
<th>Koring et al (UE-UA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.09</td>
<td>-0.05</td>
</tr>
<tr>
<td>Linear*condition</td>
<td>-0.04</td>
<td>0.42</td>
</tr>
<tr>
<td>Quadratic*condition</td>
<td>-0.15</td>
<td>-0.03</td>
</tr>
<tr>
<td>Cubic*condition</td>
<td>0.06</td>
<td>-0.01</td>
</tr>
<tr>
<td>Quartic*condition</td>
<td>0.02</td>
<td>0.10</td>
</tr>
</tbody>
</table>

37
As can be seen in Table 8, adding the interaction of cubic term and condition and the interaction of the quartic term and condition does not improve the model. However, we choose to report the results from this model so as to compare them to those of Koring et al. (2012).

There is a significant main effect of condition (p<0.05) and a significant interaction of condition and the quadratic term (p<0.001). All the other interactions between the condition and time terms are not significant (all p’s>0.05).

**Logistic mixed effect model**

The analysis is the same as Experiment 1. We modeled the proportion of looks to the target in two time windows by condition, verb type, and their interaction at the individual participant and individual item level.

Figure 7

In the verb region, there is a significant match effect (p<0.0005), no verb effect (p=0.12) and no interaction between these two terms (p=0.89). In the post-verb region, there is a main
effect of match (p<0.005) but no effect of verb (p=0.38) or the interaction of verb and matching (p=0.60).

Table 9

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Verb frame</th>
<th>Post-verb frame</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.60</td>
<td>0.07</td>
</tr>
<tr>
<td>Match</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td>Condition</td>
<td>-0.06</td>
<td>0.04</td>
</tr>
<tr>
<td>Match*Condition</td>
<td>0.00</td>
<td>0.03</td>
</tr>
</tbody>
</table>

2.3. Summary

In summary, we did not replicate the results of Koring et al. (2012) with either statistical method. With the Growth Curve analysis, if we were to replicate Koring et al (2012), we would expect to find the same effect for the same terms on the same direction (the sign of the coefficients). However, as Table 10 shows, many effects showed up in different terms and conditions in both Experiment 1 and Experiment 2, comparing to Koring et al. (2012).

With the logistic mixed effect model, were we to replicate the results of Koring et al (2012), we would expect the interaction of verb type and match to be significant at the verb region and post-post verb region in the way that during the verb region the average looks to the target is higher in the unergative match condition and in the late post-verb region the average looks to the target is higher in the unaccusative match condition. However, we did not find an interaction of verb type and match in either region in the two experiments.
Table 10

<table>
<thead>
<tr>
<th>window</th>
<th>Term</th>
<th>Koring et al. (2012)</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>verb</td>
<td>intercept</td>
<td>p&lt;0.05 (-)</td>
<td>p&lt;0.001 (+)</td>
<td>p&lt;0.001 (+)</td>
</tr>
<tr>
<td></td>
<td>linear*condition</td>
<td>n.s.</td>
<td>p&lt;0.005 (-)</td>
<td>p&lt;0.005 (+)</td>
</tr>
<tr>
<td></td>
<td>quadratic*condition</td>
<td>p&lt;0.001 (-)</td>
<td>n.s.</td>
<td>p&lt;0.001 (-)</td>
</tr>
<tr>
<td>Post-verb</td>
<td>intercept</td>
<td>n.s.</td>
<td>p&lt;0.001 (+)</td>
<td>p&lt;0.001 (+)</td>
</tr>
<tr>
<td></td>
<td>linear*condition</td>
<td>p&lt;0.005</td>
<td>p&lt;0.05 (-)</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>quadratic*condition</td>
<td>p&lt;0.001 (-)</td>
<td>p&lt;0.05 (-)</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>cubic*condition</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>quartic*condition</td>
<td>p&lt;0.0001 (+)</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

3. Resampling analysis

The data analysis has shown that we did not replicate the results of Koring et al (2012), i.e. we found no difference in subject activation for the unergative and unaccusative verbs in the logistic mixed effect models and found different effects in the Growth Curve analyses. Given that the mixed effect model does not find a difference between the unaccusative and unergative verbs and the Growth Curve model finds effects on several different terms but do not fully replicate the results of Koring et al. (2012), one may argue that the Growth Curve models can detect a difference in the shape of the curves that a logistic model cannot find if the average of the windows are the same. However, if we believe that the shape and the direction of the curves reflect the property of language processing (e.g. when the antecedent is reactivated, how long the reactivation/integration lasts, etc.), we would expect the curve to have the same shape across the three experiments. Clearly, we see in the previous section, our two studies not only cannot fully replicate the specific terms Koring et al (2012) found in their study, the results of these two studies differ in many terms. This is dubious given that these two studies have exactly the same
stimuli and very similar set-up. Therefore, we are left with two possibilities. The first possibility is that, given the lack of stability of the effects, it is likely that the Growth Curve models found the effects when there is none, probably due to fitting multiple models on the data set, or finding a random difference in the curve to be significant due to the number of the time points. The second possibility is that the results in Koring et al. (2012) are false positives because growth curve analysis is not appropriate for the data. There are three major problems with using growth curve models to analyze visual world paradigm data. First of all, it assumes that the data follows a normal distribution which is false because eye-tracking data are binary (i.e. at a given time point, the participant either looks at the target picture or not). Secondly, this analysis averages across different items rather than treating them as a random effect (see Clark, 1973 for discussion). Thirdly, the procedure in Koring et al. involves fitting multiple models without error correction and thus increases the possibility of false positive.

To test the stability of the growth curve analysis, we conducted a resampling analysis. In each resampling procedure, we created a null distribution by randomly assigning the condition labels. Then we perform statistical test on the resampling sample and recorded result for each term in the model. The analysis followed the following procedure:

1. Loop through all the 40 items. For each item $i$ (a set of sentences like in Table 1), randomly decide whether to switch the condition label (unaccusative and unergative) or not.

2. Process the data as described in the previous section and use the new condition label as the condition predictor in the growth curve analysis and record the relevant parameters

3. Repeat 1 and 2 1000 times
Table 11

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th></th>
<th>Experiment 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>term</td>
<td>false positives</td>
<td>p</td>
<td>false positives</td>
</tr>
<tr>
<td>Verb</td>
<td>linear*Cond</td>
<td>69%</td>
<td>&lt;0.001***</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>quadratic*Cond</td>
<td>50%</td>
<td>&lt;0.001***</td>
<td>57%</td>
</tr>
<tr>
<td>post-verb</td>
<td>linear*Cond</td>
<td>64%</td>
<td>&lt;0.001***</td>
<td>68%</td>
</tr>
<tr>
<td></td>
<td>quadratic*Cond</td>
<td>52%</td>
<td>&lt;0.001***</td>
<td>53%</td>
</tr>
<tr>
<td></td>
<td>cubic*Cond</td>
<td>40%</td>
<td>&lt;0.001***</td>
<td>44%</td>
</tr>
<tr>
<td></td>
<td>quartic*Cond</td>
<td>18%</td>
<td>&lt;0.01**</td>
<td>33%</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>match</td>
<td>4%</td>
<td>1.00</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>match*Cond</td>
<td>9%</td>
<td>0.41</td>
<td>8%</td>
</tr>
<tr>
<td>post-verb</td>
<td>match</td>
<td>4%</td>
<td>1.00</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>match*Cond</td>
<td>15%</td>
<td>0.03*</td>
<td>6%</td>
</tr>
</tbody>
</table>

Step 1 created a null distribution as the condition label is randomly assigned. Step 2 calculated the p-value of each parameter from the growth curve models. After Step 3, we created a distribution of p-values. We calculated the percentage of the p-values that are below the threshold 0.05. Because the 1000 samples are null distributions, we would expect the percentage to be around 5%. As can be seen in Table 11, all the parameters from the growth curve analysis are below 0.05 about 20-70% of the time which are significantly different from 5% (Fisher’s exact test all p’s<0.01). To compare this result with logistic regression models, we calculated the p-values from the logistic models following the same resampling procedure. Because condition is not significant in the logistic regression, we chose to relabel Matching which is significant and used it as the predictor instead. The p-values are below 0.05 3-15% of the time. Only once was the result significant (p=0.03, all other p’s>0.05). The results show that the growth curve analysis is highly anticonservative, i.e., the ratio of false positives of any single parameter is 20%-70%
when the threshold is set to 5%, whereas for the logistic regression analysis, it fluctuates around 5%.

Given the high ratio of significant results on a null distribution, the results confirmed our concern that growth curve analysis finds effects when there is none.

4. Cluster analysis

The strength of eye-tracking compared to paradigms such as cross-modal lexical priming is that it has a better temporal resolution. However, part of this strength is lost in the analysis with averaging across a relatively large window. The growth curve analysis is an attempt to model the differences between experimental conditions in a fine-grained manner and therefore increases the temporal resolution in the analysis (Mirman et al., 2008). However, in the last section, we see that the growth curve analyses result in a high rate of false positives with the current data. In this section, we suggest an alternative to handle the data without losing the temporal resolution.

One concern of logistic regression is the averaging. Because the dependent variable is the average of the looks to the target in the window, and the windows are about 1.5 seconds long, one may argue that there may be a verb effect in a shorter time window that is lost with the averaging.

One may also argue that the windows chosen are somewhat arbitrary and may have missed the critical time points. In our approach, we address these two concerns with a permutation cluster analysis following Maris and Oostenveld (2007) (c.f. Hahn, Snedeker, & Rabagliati, 2015 for the application to eye-tracking data). Instead of modeling the trajectory of the curves, we tested a series of consecutive time bins. The rationale is that, if there is a series of consecutive time bins that show a significant difference between unaccusative and unergative
verbs, and if the cluster is larger than what we would observe in a null distribution, we can claim that unaccusative sentences are different from the unergative sentences in that series of time bins. This method is helpful when we do not have a priori prediction on where the effect will show up. Below, we take our data from Experiment 1 as an example.

The data we chose ranged from 600 ms prior to the verb offset to 2000 ms after the verb offset. Six hundred millisecond is the average duration of the verb. We would not expect the effect to show up before the verb onset because the sentences in the same item group were the same until that point. We also had no reason to believe that the effect would show up two seconds after the verb. Because the stimuli sentences differed in length, to make sure that we did not get any random difference due to the difference in durations, and we decided to stop at 2000 ms after the verb. This was already beyond the windows used in the previous literature (c.f. Friedmann et al., 2008; Koring et al., 2012). To find the clusters, we first grouped our data into consecutive time bins with the size of 100 ms. Then, we performed the logistic mixed effect model described in Section 2 on each time bin. All the p-values for the time bins were larger than 0.05. Therefore, we did not find any cluster at the level of 0.05.

One benefit of the resampling analysis is that we can do it at any p-value without increasing the chance of false positives. To test whether there is a smaller but long-lasting effect, we set our threshold at a higher level, i.e., 0.2. With this standard, there is one cluster of size four in our result for Experiment 1 that occurs at -200 before verb offset to 100 ms after the verb offset. Then, the question is how likely it is for us to find a cluster of this size with a random sample.

To figure out the likelihood of finding a cluster of the same size in a null distribution, we performed a permutation test. The permutation followed the same steps as the resampling
analysis above. To recap here, in each iteration, for each item, we randomly switched the condition label at chance level. This means that the verb label was randomly assigned to each pair of the verbs. There was a 50% chance that a verb received an unaccusative or unergative label. This helped us to maintain the counterbalanced design and the temporal relation within an item.

We resampled 1000 times. In each iteration, we counted the number of clusters when the threshold was set to 0.2. This created an empirical distribution of the clusters. Comparing the likelihood of getting a cluster in a null distribution that was the same or larger than size four, we could tell whether the cluster we had in the data was a true effect.

We performed the resampling procedure on the data from Experiment 1 and created 1000 permuted samples of null distributions. For each sample, we ran the same test on each time bin. Among the 1000 samples, there were 467 clusters of size four or larger where the p-value was smaller than 0.2. We concluded that the likelihood that we would find a cluster of the same size in a random sample was about 46.7%. Therefore, the cluster we had was probably random, and we had no reasons to believe that there was a true difference between unaccusative and unergative conditions.

We conducted the same cluster analysis for Experiment 2. For the experimental data, there was no cluster at the 0.05 level and a cluster of size three at the end of the window (1600-1800 ms after the verb offset) when the threshold was set to 0.2. In the 1000 permuted samples, there were 648 clusters of size three or larger where the p-value was smaller than 0.2. Given the results and given that the clusters in the two studies did not appear around the same time, the clusters we found were most likely random. To conclude, no processing difference was found between the unaccusative and unergative verbs in finer-grained analyses even with a more generous threshold.
5. Discussion

In these experiments we tested the claim that there is a fundamental difference in how unaccusative and unergative verbs are processed due to the greater syntactic complexity of unaccusatives. We found no evidence to support this claim. When we conducted simple analyses over large time windows, using logistic models, we found no differences between the verb classes. When we conducted a more fine-grained cluster analyses, we again found no differences. Curiously, however, when we conducted growth curve analyses, closely modeled on Koring, we found effects of verb class on the temporal parameters, in both of our time windows and in both of our studies, which appeared to be highly significant. But critically, these apparent effects were not consistent with those that Koring reported, nor were they consistent across our two experiments. To check the validity of these analyses, we conducted a resampling analysis, using our experimental data, with verb condition labels randomly assigned, which revealed that, for our data sets at least, the growth curve analyses produced p-values that were wildly anti-conservative.

One clear difference between the studies is that ours was conducted in English while Koring’s was conducted in Dutch. Unlike English, Dutch is a language that marks the distinction between unergative and unaccusative verbs in the auxiliary system. Auxiliaries were not used in the Koring sentences, but it is possible that speaking a language which marks this distinction could change the way in which one processes unaccusative verbs. If true, this explanation would radically change how we think about this distinction. Generative theories have assumed that the purported syntactic distinction is present in all languages due to deep properties of the syntax-semantics interface. Shifting the explanation to cross-linguistically variable aspects of the
morpho-syntax would radically alter these theories. We cannot rule out this explanation with the current chapter, but I will come back to it in Chapter 3.
Chapter 3

EYE-TRACKING STUDY ON DUTCH UNACCUSATIVITY

1. Introduction

In the previous chapter, I reevaluated the processing evidence for the Unaccusative Hypothesis for English. We have seen that the results of Koring et al. (2012) cannot be replicated in English. One open question is whether we can find a processing difference between unaccusative and unergative sentences in Dutch.

Unlike English, where different unaccusative diagnostics often pick different subsets of verbs (see Chapter 4 for a detailed discussion), Dutch has more explicit linguistic clues for the unaccusative/unergative distinction. It has been argued that impersonal passivization (17), past participles prenominal modifiers (2) and HAVE/BE auxiliary selection (19) are all associated with unaccusativity in Dutch (Perlmutter 1978; Hoekstra & Mulder 1990 a.o.; c.f. Van Valin 1990 and Zaenen 1988 for counterarguments). In addition, different from English where a single diagnostic only picks out a distinct subset of verbs, the Dutch diagnostics seem to be able apply to the same verb (as shown in Koring et al. 2012 for their verb selection). This suggests that Dutch may consistently mark the verbs with their syntactic constructions across different diagnostics, although further work needs to be done to test the distribution of the verbs.
impartial passivization:

a. In de zomer wordt er hier vaak gezwommen.
   In the summer become it here often swim.
   ‘In the summer, it is swum here frequently.’

b. *In de zomer wordt er hier vaak verdrongen.
   In the summer become it here often drowned.
   ‘In the summer, it is drowned here frequently’

(Perlmutter 1978)

past participles prenominal modifiers

a. Susan lachte.
   ‘Susan laughed.’

b. *de gelachen Susan
   the laughed Susan

c. De vaas viel.
   ‘The vase fell.’

d. de gevallen vaas
    the fallen vase

(van Hout, 1996)
(19) HAVE/BE auxiliary selection

a. Susan heeft gelachen.
   Susan has laughed
   ‘Susan laughed.’

b. De vaas is gevallen.
   the vase is fallen
   ‘The vase fell.’

(van Hout, 1996)

In the above examples, *swim* is an unergative verb in Dutch, therefore in (17a), it can be used in impersonal passivization. *Drown*, on the other hand is an unaccusative in Dutch, and therefore the impersonal passivization sentence (17b) is ungrammatical. Similarly, *laugh* is an unergative. Therefore, its past participle is bad as a pronominal modifier and it selects auxiliary *have*. *Fall* is an unaccusative and its past participle can be a pronominal modifier and it selects auxiliary *be*.

In addition, van Hout (1996) has found that the particle or prefix variant seems to be associated with the unergative-unaccusative distinction. For example, adding the prefix “*in*” can change the unergative *slapen* “sleep” to an unaccusative *inslapen* “fall asleep”. This also gives us some morphological clues to distinguish unaccusatives from unergatives. In addition, van Hout (1996) has argued that the semantic factor, telicity, which is also associated with the unaccusative-unergative distinction in Dutch (see Grimshaw, 1990; Borer, 1994; a.o. for general discussions), may provide a clue for the acquisition of the unergative-unaccusative distinction in terms of semantic bootstrapping (Pinker 1989).
If we believe that the syntactic and morphological distinctions are reflexes of a difference in the syntactic structure, we have stronger reasons to believe that Dutch has two distinct intransitive classes. From an acquisition point of view, Dutch has more an explicit and consistent marking for the verbs than English (auxiliary and morphology). It may be easier for a Dutch-speaking child to learn the distinction through syntactic and semantic bootstrapping (Pinker, 1989; Gleitman, 1990) than an English-speaking child.

Therefore, in the next section, we test whether a processing difference can be found between unaccusative and unergative conditions by adopting and modifying the paradigm in Koring et al. (2012).

2. Method

Subjects

Forty native speakers of Dutch were recruited from the University of Groningen to participate for course credit or as volunteers. They all reported that they have normal or corrected to normal vision and normal hearing.

Materials

Selection of verbs

All the unaccusative verbs in this study were selected based on the following criteria: (1) they select the auxiliary *be* in perfectives; (2) they are ungrammatical in impersonal passive construction; (3) they are possible to be prenominal past participle; (4) they are impossible to have -er nominalization, and (5) they are possible to add *vanzelf* ‘by itself’. The unergative verbs were selected based on the following criteria: (1) they select the auxiliary *have*; (2) they are grammatical in the impersonal passive construction; (3) their participle forms cannot be used as
prenominal modifiers; (4) they can be nominalized with the morpheme -er; (5) they are impossible work with the vanself (‘by self’), and (6) they can appear in a resultative construction only with a reflexive. Like the original study (Koring, Mak, & Reuland, 2012), all the verbs are non-alternating intransitive verbs. Seven of the 20 unaccusative verbs and 6 of the 20 unergative verbs from our study were taken from Koring et al. (2012). The rest of the verbs were selected based on the criteria listed above.

We controlled for the frequency of the verbs, as Koring et al. (2012) did. The lemma frequencies of unaccusative verbs and unergative verbs were determined by SUBTLEX-NL (Keuleers, Brysbaert, & New, 2010). The log transformed mean frequencies of unaccusative and unergative verbs (7.24 and 6.63 respectively) do not differ significantly (t = 1.02, p = 0.31).

_Sentences and images_

All our test and control sentences have the same structure as those in the English studies:

Mary/John said that + the subject noun + a prepositional phrase modifying the argument + an adverb + main verb (+ a prepositional phrase) + a temporal clause

In each sentence, a modifier of the noun and an adverb appeared between the subject noun and the verb. Together they were 6-10 words (mean=7.8), 10-15 syllables long (mean=12.7). The modifier and the adverb were not semantically related to the subject, the verb, the target image or the distractors. After the verb, there were a temporal clause and a post-verbal prepositional phrase. Together they were 6-12 words (mean=8.4), 11-16 syllables long (mean=13.8). Neither the temporal clause nor the prepositional phrase was semantically associated with the critical argument, the target image or the distractors. We also made sure that the verbs were not semantically related to either the critical argument or the target image. All the
syllable counts are similar to or longer than those of Koring et al. (2012) to make sure that there is enough time for deactivation.

<table>
<thead>
<tr>
<th>Verb</th>
<th>Sentence</th>
<th>Match</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>unaccusative</td>
<td>Marie zei dat de architect met een gele hoed en een blauwe paraplu meteen wegsloop toen de blonde serveerster de hoge rekening bracht.</td>
<td>compass</td>
<td>magnifying glass</td>
</tr>
<tr>
<td>unergative</td>
<td>Marie zei dat de architect met een gele hoed en een blauwe paraplu meteen aarzelde toen de blonde serveerster de hoge rekening bracht.</td>
<td>building plan</td>
<td>magnifying glass</td>
</tr>
<tr>
<td>unaccusative</td>
<td>Marie zei dat de detective met een gele hoed en een blauwe paraplu meteen wegsloop toen de blonde serveerster de hoge rekening bracht.</td>
<td>magnifying glass</td>
<td>building plan</td>
</tr>
<tr>
<td>unergative</td>
<td>Marie zei dat de detective met een gele hoed en een blauwe paraplu meteen aarzelde toen de blonde serveerster de hoge rekening bracht.</td>
<td>magnifying glass</td>
<td>building plan</td>
</tr>
</tbody>
</table>

Figure 8 “Mary said that the architect/detective with a yellow hat and a blue umbrella disappeared/hesitated immediately when the blonde waitress took the expensive bill.”

All subject nouns were terms for occupations like in the English study. Some of the nouns differ between these two studies, because some occupations do not have a translatable equivalent in Dutch. In each trial, there are four images on the screen. The images are all black and white line drawings from Bank of Standardized Stimuli (Brodeur, Guérard, & Bouras, 2014), Bonin, Peereman, Malardier, Méot, and Chalard (2003) and Szekely et al. (2004). In the relatedness judgment test conducted on Amazon Mechanical Turk (AMT), the mean relatedness between the argument and the related image was 4.72 on a scale from 0 to 5 and the relatedness between the argument and the paired non-related image was 0.36 on a scale from 0 to 5.

In the test condition, all the critical arguments have an image that has a schematic relation to it (for example, detective and magnifying glass). For example, when the participants hear the sentence with the detective, they see a magnifying glass as in Figure 9. In the control condition,
there is no image that is related to the critical argument. For example, a building plan substitutes the magnifying glass and the other three images stay unaltered as in Figure 9. The image that is not related to the argument is related to the alternate argument for that pair of verbs (for example, the building plan is related to the architect). Therefore, when the participants hear the sentences with the architect, the magnifying glass in Figure 9 becomes the control condition and the building plan in Figure 9 becomes the test condition.

Because the number of the unaccusative verbs is limited, especially after controlling for the frequency and imageability, we created two different sentence frames and two sets of visual stimuli for each pair of the verbs in order to have the same number of sentences as Koring et al. (2012). If the unaccusative verb in a pair is selected for sentence frame 1, then we used the unergative verb in the same pair for sentence frame 2 in the same list so that the participant does not see the same verb twice. Four lists were created so that each participant only heard the similar sentence and saw the similar visual stimuli once. Our filler items, like those in Koring et al. (2012), had transitive verbs and were paired with displays in which one image depicted the
subject of the sentence. There were 40 fillers which remained the same across all four lists, resulting in a total of eighty sentences in each list (see Appendix 4).

Because the sentence frames and the visual stimuli are the same between the unaccusative condition and the unergative condition and across the test and control conditions, we can ensure that any differences we find between the unaccusative and unergative conditions are due to the differences between these two verb categories, rather than uncontrolled differences between the images, subjects or the rest of the sentences.

We conducted a norming study to control for the relatedness between the critical images and critical arguments. In this judgement test, participants need to rate the relatedness of image and the noun between 1 to 5. According to the results of this test, the mean relatedness between the argument and the related image is 4.5 on a scale of 1 to 5 and the relatedness between the argument and the paired non-related image is 0.4.

We also controlled the plausibility of the verb given the argument as the subject for that verb (4.58 and 4.69 respectively; Mann-Whitney-Wilcoxon Test, p=0.14).

The audio stimuli were recorded at a normal speaking rate by a female native speaker of Dutch, sampled at 44.1 kHz.

Procedure

We based our procedure on Koring et al. (2012). Participants were seated comfortably in front of a monitor. Their eye movements were measured by a Tobii T120. Each session started with a calibration procedure with five fixation points. Participants were told that they would hear some sentences and look at some pictures. During the study, they needed to answer questions. The questions were displayed with two choices. The participants needed to touch the
right answer to proceed with the study. To reduce noise in the data, each trial started with a centrally-located fixation dot. Participants were instructed to look at the dot briefly. There was a 1 second preview of the display before the onset of the spoken sentence. After the end of the spoken sentence, there was another 1 second of silence before the fixation dot appeared. The entire experiment lasted about 20 min. During the study, 16 questions appeared randomly, asking about the sentences or the images just displayed to the participants.

3. Results

Before analyzing the data, we removed the trials where the validity of more than half of the samples is higher than 3 (higher numbers indicate lower validity). To make the results comparable with Koring et al. (2012), we first analyzed our data with the growth curve analysis (Mirman, Dixon, & Magnuson, 2008) they used for their data analysis.

3.1. Growth curve analysis

In this analysis, the change of the proportion of looks to the target was modeled over time. The dependent variable was the differential score (the proportion of looks to the target picture in the test condition minus the proportion of looks to the target picture in the control condition). We initially had only time polynomial factors as predictors and gradually added the condition and interaction on the different terms. This is different from Koring et al. (2012) where in each step they added all the interactions with the higher order time terms at the same time. Orthogonal power polynomials were included as predictors for higher-order time terms. The dependent variable was the difference score, i.e. proportion of looks to the target in the test condition minus proportion of looks to the control condition.
Following Koring et al, we analyzed our data in two frames. The verb frame starts 600 ms before the verb offset and ends 1000 ms after the verb offset. The post-verb frame starts 200 ms after the verb offset and ends 1700 ms after the verb offset.

Verb frame

![Verb Frame](image)

Figure 10

In this window, the different score was modeled by condition, linear term, quadratic term and their interactions with condition at individual participant level (i.e. participants as random effect). Unnergative was taken as the baseline condition in the analysis. The results of model comparison are summarized in Table 12. As can be seen, adding the condition and its interaction with the linear time term improved the model. However, adding the interaction of the quadratic term and condition did not improve the model.
Table 12 model comparison (random slopes in all models includes all time terms in the fixed effect)

<table>
<thead>
<tr>
<th>model</th>
<th>AIC</th>
<th>Chisq</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>-4097.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>linear+quadratic</td>
<td>-4334.00</td>
<td>244.62</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+condition</td>
<td>-4379.50</td>
<td>47.53</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+condition+linear*condition</td>
<td>-4390.10</td>
<td>12.64</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+condition+linear<em>condition+quadratic</em>condition</td>
<td>-4388.90</td>
<td>0.73</td>
<td>.39</td>
</tr>
</tbody>
</table>

We chose to report the results of the full quadratic model, i.e. the same model used in Koring et al. (2012), although this model is not significantly different from the model without the quadratic condition interaction. The results are summarized as Table 13. Our results are different from those in Koring et al. (2012) in many terms. In our study, there was a significant interaction of condition and linear term but no significant interaction of condition and quadratic term. In the original Koring et al. (2012) data, they found a significant negative quadratic term in the unergative condition which means a rise followed by a fall. They interpreted it as an increase of looks to the target picture followed by a decrease. They argued that this showed that the participants looked to the target picture when they first heard the verb and then looked away from it. In our data, we did not find this effect because the quadratic interaction was not significant.
Table 13

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Current Experiment</th>
<th>Koring et al (UE-UA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.23</td>
<td>7.85</td>
</tr>
<tr>
<td>Linear*condition</td>
<td>0.13</td>
<td>3.56</td>
</tr>
<tr>
<td>Quadratic*condition</td>
<td>0.03</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Post-verb frame

Figure 11

In this window, the difference score was modeled by condition, linear term, quadratic term, cubic term, quartic term and their interaction with condition at individual participant level. Unergative was taken as the baseline condition in the analysis. As can be seen in Table 14, gradually adding the condition and its interaction with the higher order time terms improved the
model. Therefore, we chose to report the model with the full interaction terms as use in Koring et al. (2012).

Table 14 model comparison (random slopes in all models includes all time terms in the fixed effect)

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Chisq</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear+quadratic+cubic</td>
<td>-4617.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic</td>
<td>-4732.30</td>
<td>126.80</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition</td>
<td>-4745.00</td>
<td>14.68</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear*condition</td>
<td>-4789.70</td>
<td>46.72</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear<em>condition+quadratic</em>condition</td>
<td>-4793.50</td>
<td>5.82</td>
<td>.05*</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear<em>condition+quadratic</em>condition+cubic*condition</td>
<td>-4807.50</td>
<td>15.97</td>
<td>.001***</td>
</tr>
<tr>
<td>linear+quadratic+cubic+quartic+condition+linear<em>condition+quadratic</em>condition+cubic<em>condition+quartic</em>condition</td>
<td>-4817.4</td>
<td>11.91</td>
<td>.001***</td>
</tr>
</tbody>
</table>

Similar to Koring et al (2012), we found a significant interaction between the linear component and the condition and the interaction between the quartic component and the condition (p’s<0.001). We also found a significant interaction of the quadratic term and condition, the cubic term and condition, as can be seen in Table 15.

After comparing the results with those of Koring et al. (2012) as in Table 13 and Table 15, we can conclude that we did not replicate that study in either time frame. In the verb frame, instead of getting a quadratic trend, we have a significant linear trend. In the post-verb frame, all the time terms interact with condition in our results while Koring et al. (2012) only had a significant linear and quartic interaction. Given the inconsistency in the results between the two
studies and the fact that so many effects were significant, we may question if the effects we found are reliable results or if they are false positives.

Table 15

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Current Experiment</th>
<th>Koring et al (UE-UA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.20</td>
<td>6.42</td>
</tr>
<tr>
<td>Linear*condition</td>
<td>-0.23</td>
<td>-6.87</td>
</tr>
<tr>
<td>Quadratic*condition</td>
<td>-0.08</td>
<td>-2.42</td>
</tr>
<tr>
<td>Cubic*condition</td>
<td>0.14</td>
<td>4.00</td>
</tr>
<tr>
<td>Quartic*condition</td>
<td>0.12</td>
<td>3.45</td>
</tr>
</tbody>
</table>

3.2. Resampling on growth curve analysis

We found different effects in the growth curve analyses compared to those of Koring et al. (2012). This is not surprising given that I have already shown that growth curve models result in a high rate of false positives.

As the specific reason for false positives is yet to be investigated, to verify that there is a problem of false positives in the current data set we conducted a resampling analysis as in Chapter 2, repeated as below. In each resampling procedure, we created a null distribution by randomly assigning the condition labels. Then we perform statistical test on the resampling sample and recorded the result for each term in the model. The analysis followed the following procedure:

1. Loop through all 40 items. For each item \(i\) (a set of sentences like in Figure 8), randomly decide whether to switch the condition label (unaccusative and unergative) or not.
2. Process the data as described in the previous section and use the new condition label as the condition predictor in the growth curve analysis and record the relevant parameters.

3. Repeat 1 and 2 1000 times

To review what each step does: Step 1 created a null distribution as the condition label is randomly assigned. Step 2 calculated the p-value of each parameter from the growth curve models. After Step 3, we created a distribution of p-values. We calculated the percentage of the p-values that are below the threshold 0.05. Because the 1000 samples are null distributions, we would expect the percentage to be around 5%. As can be seen in Table 16, all the parameters from the growth curve analysis are below 0.05 about 30-60% of the time which are significantly different from 5% (Fisher’s exact test all p’s<0.001). This is similar to what we saw in English data. The results show that the growth curve analysis is highly anticonservative for the current data set, i.e., the ratio of false positives of any single parameter is 30%-60% when the threshold is set to 5%. This explains why we did not replicate the results of Koring et al. (2012).

Table 16

<table>
<thead>
<tr>
<th>term</th>
<th>false positives</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>linear*Cond</td>
<td>60%</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>quadratic*Cond</td>
<td>58%</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>Post-verb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>linear*Cond</td>
<td>66%</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>quadratic*Cond</td>
<td>55%</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>cubic*Cond</td>
<td>40%</td>
<td>&lt;0.001***</td>
</tr>
<tr>
<td>quartic*Cond</td>
<td>34%</td>
<td>&lt;0.001***</td>
</tr>
</tbody>
</table>
3.3. Logistic regression models

As we have seen from the resampling analysis, the growth curve analysis results in a high proportion of false positives. Therefore, the significant results we had from the models are likely to be false positives. To test whether the unaccusative and unergative conditions were truly different, we analyzed the data with logistic mixed effect models which has been shown to be more reliable in Chapter 2. The dependent variable is whether the participants look at the target in each time window in each trial. If in that time window, the participant looked at target more than 30% of the time, it is determined that they looked at the target. Otherwise, they are categorized as not looking at the target (1 vs. 0). The independent variables are: verb type (unaccusative vs. unergative coded as 1 and -1), match condition (match vs. mismatch coded as 1 and -1) and their interaction. Participant and item are the random intercepts, and the match and verb type as random slopes (The full random effect model did not converge. Therefore, we chose to report the model with the model best fit, which is one with verb type and match as random slopes, and verb and match as random slopes).

To compare with the results of the growth curve models, we analyzed our data with the same windows. They are: 600 ms before the verb offset to 1000 ms after the verb offset and 200 ms after the verb offset to 1700 ms after the verb offset.

There is a significant match effect in both windows (p’s<0.001), showing that the participants looked to the target picture more in the test condition than the control condition. The verb type and its interaction with match are not significant in any window (all p’s>0.05). This means that the proportions of looks to the target picture do not differ between the two verb conditions in general nor do they differ when the match condition is controlled for. The detailed statistics are shown in Table 17.
As we can see, we did not find a difference between unaccusative and unergative conditions in the logistic mixed effect models. Because the dependent variable is the average of the looks to the target in the window, as the windows are about 1.5 seconds long, one may argue that there may be a verb effect in a shorter time window that is lost with the averaging. One may also argue that the windows chosen are somewhat arbitrary and may have missed the critical time points. To rule out these two possibilities, we conducted a cluster analysis.

### 3.4. Cluster analysis

One advantage of using eye-tracking for the study is that it has a better temporal resolution than cross-modal lexical priming which only has limited probe positions. Koring et al.
(2012) have argued along with Mirman et al. (2008) that we should take the advantage of the better temporal resolution. However, we have shown that the growth curve models they argued for are not appropriate for our dataset. In this section, we took a different approach. Instead of modeling the trajectory of the curves, we tested a series of consecutive time bins. The rationale is that, if there is a series of consecutive time bins that show a significant difference between unaccusative and unergative verbs, and if the cluster is larger than what we would observe in a null distribution, we can claim that unaccusative sentences are different from the unergative sentences in that series of time bins. This method is helpful when we do not have a priori prediction as to where the effect will show up.

The window we chose is from 600 ms prior to the verb offset to 2000 ms after the verb offset. Six hundred milliseconds is the average duration of the verb. We would not expect the effect to show up before the verb because the sentences are the same until that point. We also had no reason to believe that the effect will show up two seconds after the verb. Because the stimuli sentences differ in length, to make sure that we would not get any random difference due to the difference in durations, and we decided to stop at 2000 ms after the verb. This was already beyond the windows used in the previous literature (c.f. Friedmann et al., 2008; Koring et al., 2012). To find the clusters, we first grouped our data into consecutive time bins with the size of 100 ms. Then, we performed the logistic mixed effect model described above on each time bin. There was no cluster at the alpha level of 0.05.

Another advantage of this method is that the significant threshold is empirically defined by a permutation test. Therefore, we can adjust our alpha level without increasing Type I error (false positive rate). This can help us find a shallow but long-lasting effect. When we set the threshold at 0.2, we had a cluster of size 2 at the verb offset. Now, the question is how likely it is
to have cluster of size 2 at the level of 0.2 in a null distribution (where two groups belong to the same underlying population and no difference should be observed). If we are unlikely to have such an effect with a null distribution, we can tell that this is a true effect.

To test how likely it is to find such an effect in a null distribution, we performed a permutation test. In each iteration, we randomly assigned the condition labels to each item, such that pair could either retain the correct label (unergative or unaccusative) or the labels could be switched. As a result, on any resampling, there is a 50% chance that any given verb would be assigned the unaccusative or unergative label. This captures the temporal structure of each trial and the nature of the counterbalancing while removing any signal attributable to the verb class itself.

We ran the permutation 1000 times and created 1000 samples of null distributions. For each sample, we ran the same test on each time bin. Among the 1000 samples, there are 1014 clusters of size 2 or larger where the p value is smaller than 0.2. This means it is very likely to find a cluster of size 2 at 0.2 level in a null sample. Therefore, the cluster we found is highly likely to be due to chance alone and does not reflect any meaningful difference between unaccusative and unergative conditions.

4. Discussion

In this chapter, we have shown that we did not find a difference between unaccusative and unergative sentences in the subject reactivation pattern in Dutch. The discrepancy between our conclusions and the conclusions of Koring et al. (2012) can be explained in three possible ways. First, it is possible that the effect is real but that we failed to detect it due to experimenter errors or chance alone. We find this explanation unlikely. Several of the putative effects in Koring had very low p-values and thus would seem to be easily replicable. Several of our
(unstable) effects were similar in this respect. In addition, in our experiments, we saw large robust effects of semantic relatedness which shifted in a sensible way with the unfolding sentence (e.g., emerging shortly after the subject noun), validating our experimental methods and the integrity of our data pipeline. In addition, our results for the English and Dutch study both failed to find a reliable processing difference between the unaccusative and unergative conditions. These two versions of the studies were conducted with different stimuli and in different experimental environments; therefore, it is unlikely that the failure of replication is due to chance or experimenter errors.

Another possibility is that we failed to find a processing difference because of the experimental materials. The classification of unaccusative and unergative verbs is not uncontroversial. Different diagnostics classify the verbs differently (more on this point in Chapter 4). If Koring et al. (2012) happened to choose more real unaccusative verbs than us, they may be able to find an effect that we cannot find with our stimuli. We find this explanation unlikely as well. We used the same criteria as Koring et al. (2012) in Dutch and Friedmann et al. (2008) in English to choose our verbs. There is no theoretical discrepancy between how we define our unaccusative verbs and how they define theirs. If we believe that the same diagnostic classifies the verbs consistently, we have no reason to believe that the previous studies have more real unaccusative verbs in their stimuli.

Another, and the simplest, explanation is that Koring results are simply a false positive attributable to the instability of growth curve analyses. In our resampling analyses we learned that when we randomly shuffled the condition labels (arbitrarily labeling verbs as unaccusative or unergative) we would get apparently reliable results on each of the relevant parameters roughly half the time. Since the analyses each had two to four parameters capturing differences
in the verbs, this meant that most of the random models produced false positives. In other words, if we assume that Koring's data is similar, then the findings presented in that paper are consistent with what we would expect if there were absolutely no differences between the two verb classes.

Given the English results in Chapter 2, this is not surprising. Although we had reasons to believe that Dutch may have a clearer pattern of unaccusaitivity, it is inconsistent with the original Unaccusative Hypothesis, which assumes that unaccusativity is universal across languages. If we found a difference between Dutch and English, we would have to explain why there is an underlying syntactic difference in one language but not another. Such a difference will be hard to explain given that it cannot be innate (we cannot assume that Dutch- and English-speaking children are born with different grammars) but is implicit (we cannot see the movement or the trace on the surface structure.).

If this third explanation is accurate, then there is currently no evidence from the visual world paradigm for a processing difference between unaccusative and unergative verbs. There are, however, a handful of experiments using other methods that seem to show such differences. As we have mentioned in the previous chapter, there are a few studies with the cross-modal lexical decision priming paradigm (Burkhardt et al., 2003; Friedmann et al. 2008) that found a processing difference between unaccusative and unergative verbs. To recap, both studies found a reactivation of unaccusative verbs about 700 ms after the verb. However, they differed in whether the subject of an unergative verb was reactivated. Burkhardt et al. (2003) found a reactivation of unergative subjects 100 ms after the verb while Friedmann et al. (2008) found no evidence of reactivation for unergative subject right after the verb. One possible explanation is that it takes some time to process the information and reactivate the unergative subject and therefore the probe right after the verb in the Friedmann study cannot catch this effect. These
measures rely on semantic priming mediated by the subject noun. The present study also relies on this mechanism. If these studies provide a stable insight into processing, we should see evidence for this mediated semantic priming 100 ms after the verb offset in unergative verbs and around 700 ms after the verb offset for unaccusative verbs. However, in our two analysis windows which covered these timepoints, we did not find any difference between unaccusatives and unergatives. We can rule out the possibility that effects in the visual world paradigm are simply delayed due to the time required to program an eye movement. A fine-grained analysis over a large time frame (i.e. our cluster analysis) did not show any differences between these two types of verbs between 600 ms before to 2000 ms after the verb offset. Therefore, we think it is unlikely that we missed the critical window of the effect. Another possibility is that cross-modal lexical priming is simply more sensitive than a visual world paradigm. We think this is unlikely; like Friedmann et al. (2008), we found large robust effects of semantic relatedness at the subject, demonstrating that the visual world paradigm is sensitive to priming.

We believe that the differences between these studies are best explained by looking at the potential confounds in the stimuli. The cross-modal priming studies did not equate their unaccusative and unergative stimuli for the full set of potential confounds that we controlled. For example, neither of the two studies reports controlling for the imageability of the verbs, animacy of the subject or holding constant the sentence frames across the two verb classes (see Chapter 2 for more details). If less imageable verbs or less animate subjects lead to slower argument reactivation (see e.g., Paivio, 1991) this could account for the differences observed in these studies.

Besides the lexical priming studies, there are also a few studies with functional magnetic resonance imaging (Shetreet, Friedmann, & Hadar, 2010; Agnew et al., 2014). The results from
these experiments are inconsistent. While Shetreet and colleagues (2010) found that unaccusative verbs have different patterns of activation than unergative and transtive verbs verbs, Agnew and colleagues (2014) found that unaccusatives were more similar to transitives than to unergatives. It is unclear whether either experiment controlled for the imageability of the verb or the animacy of the subject. Thus it is possible that these divergent patterns reflect factors other than verb class. In summary, the psycholinguistic studies do not provide clear support for the claim that unaccusatives are processed differently than unergatives. While future studies, using tighter controls and more robust statistics, may produce such evidence, for now we must remain agnostic.

What does this mean for the unaccusative hypothesis more broadly? The failure to find processing effects in no way excludes the possibility that there is an underlying syntactic difference between unaccusative and unergative verbs. If processing time largely reflects predictability and frequency rather than structural complexity (Levy, 2008), then online measures would be the wrong place to look for answers to syntactic questions. For example, perhaps the syntax of the unaccusative verb is a tad more complex, but this complexity is built into the lexical entry and adds no measurable cost to retrieval. In the absence of any strong processing data, support for the Unaccusative Hypothesis must come from its ability to provide unique predictions about the syntactic distribution of verbs across a range of languages. Linguists have long argued that there is a principled cluster of syntactic phenomena that distinguish unaccusative verbs from unergative verbs, and suggest that unaccusative subjects are close kin to transitive objects (Perlmutter, 1978; Burzio, 1981, 1986; Rosen, 1984; Levin & Rappaport, 1986; a.o.). Critically, the Unaccusative Hypothesis, predicts that these diagnostics should divide verbs
cleanly into the same two categories (across tests and across languages). I will return to these phenomena in Chapter 4.

Beyond unaccusativity, what can we learn about movement? There are two types of movement in the syntactic theory: A-movement in which the constituent moves from an argument position to an argument position, such as the movement in the unaccusative sentences and passive sentences; and A’-movement in which the constituent moves to a non-argument position, such as wh-movement. Both types of movement have been claimed to be supported with experimental evidence. In this chapter, I have shown that the evidence for unaccusativity is not reliable. Another place to look for existing processing evidence of A-movement is in passive sentences. The results from this literature is unclear. There is some work that has found a reactivation of the antecedent in passive sentences (Osterhour & Swinney, 1993). However, other researchers failed to find such a reactivation (Osterhout, unpublished, mentioned in Nicol & Swinney, 1989). Because of the inconsistency of the literature on passive sentences and potential publication bias (that only significant results get published), it is likely that the results of reactivation found in passive sentences is not reliable either.

On the other hand, there is a richer literature on wh-movement (e.g. Love, & Swinney, 1996 for relative clause; Sussman & Sedivy, 2003 for wh-question; Phillips & Wager, 2007 for a review). My personal work on relative clauses with Visual World Paradigm similar to the one presented in this dissertation also reveals a reactivation effect of relative clauses antecedent at the gap site. It suggests that either A-movement and A’-movement are two very different processes with two different types of traces and have different processing signatures, or only A’-movement has a trace position and A-movement does not create a trace position or is not a real movement.
These results are consistent with linguists’ observation that only A’-traces but not A-traces block contraction (Lightfoot, 1976), as shown in (20).

(20)  Who_i do you want t_i to come?

*Who do you wanna come?

John_i is going t_i to come.

John is gonna come.

If A-traces and A’-traces are very different in nature, or if only A’-traces exist, we would expect them to behave differently in language processing, which is consistent with the existing data mentioned above and in this dissertation.
1. Introduction

In Chapters 2 and 3, we have seen that no reliable processing evidence for the Unaccusative Hypothesis was found in English or Dutch. A question one may ask is: are there other types of evidence that can support the Unaccusative Hypothesis? It has been claimed that unaccusative and unergative verbs appear in different linguistic constructions. Researchers have argued that the difference in the distribution signals the difference in the syntactic structure (Perlmutter, 1978; Rosen, 1984; Burzio, 1986; a.o.). Therefore, it serves as an independent reason for the dichotomy of the intransitive verbs apart from the processing evidence. In this chapter, I will take English as an example to evaluate the distributional difference between unaccusative and unergative verbs.

There are several linguistic phenomena that have been claimed to be related to unaccusativity in English. These diagnostics can be classified as deep vs. surface, syntactic vs. semantic, morphologically derived vs. syntactically derived, based on different standards. Levin and Rappaport Hovav (1995) made a distinction between surface unaccusativity and deep unaccusativity. Surface unaccusativity refers to the phenomena where the sole argument of the intransitive verb can appear as a direct object in the surface structure. Deep unaccusativity refers to the phenomenon where the sole argument of the intransitive verb shows some object-like
properties but appears as the subject of the sentence. Orthogonal to the surface-deep unaccusativity distinction, the phenomena linked to unaccusativity could be classified (controversially) based on how they are formed: syntactic transformation or morphological derivation. They could also be classified based on what aspect of difference they express: syntactic or semantic. In the following sections, I will discuss three types of diagnostics with judgment data. These three diagnostics were chosen to represent each aspect of the dimension (deep vs. surface; different formation processes; differences expressed by the diagnostics). They are not the only diagnostics proposed for English. However, they serve as representative examples to show how different types of diagnostics work. For each diagnostic, I first review why the phenomenon is linked to unaccusativity (if available), and spell out the prediction on the distribution of the linguistic data assuming such a link is true. Then, I present data from my online judgment experiments to test if those predictions are carried out. After that, I discuss the potential problems of these diagnostics. The last section discusses the implication of the findings and summarizes the chapter.

2. *There*-insertion construction

   There are only two candidates for diagnostics of surface unaccusativity in English: *there*-insertion construction and locative inversion construction. Locative inversion is the phenomenon where a locative expression appears preverbally and the argument of the verb appears after the verb (21). *There*-insertion is the phenomenon where the subject of the sentence appears in the object position, and the subject position is occupied with an expletive *there*. It has been claimed that only unaccusative and not unergative verbs can appear in such a construction, see (22). In this section, we take *there*-insertion as an example to discuss the theoretical explanation and the data pattern of surface unaccusative diagnostics.
(21) In the distance appeared the towers and spires of a town … [Bromfiled, The Farm, 124]

(from Levin & Rappaport Hovav, 1995).

(22)  
    a. An old man arrived.
    b. There arrived an old man.
    c. An old man laughed.
    d. *There laughed an old man.

In the above example, (22b) is the transformed sentence of (22a), (22d) is the transformed sentence of (22c). The sentence with arrive, an unaccusative verb, is good after the transformation but the sentence with laugh, an unergative verb is bad after the transformation.

This ungrammaticality of (22d) can be explained by the assumption that the subject and the expletive compete for a single specifier position, Spec, TP, in English, a language lacking Transitive Expletive Constructions (TECs) (Alexiadou & Schäfer, 2010). To be more specific there is merged into Spec, TP to check the strong D-feature on T required by the Extended Projection Principle (EPP, i.e., a sentence has to have a subject) (Chomsky, 1995). For an unergative verb, the subject that is base generated in Spec, vP is required to move to Spec, TP (Alexiadou & Anagnostopoulou, 2001). This results in two constituents competing for the same position. For an unaccusative verb, the argument is base generated in the object position. (The position differs depending on specific theories. The critical point is that it is not in the Spec. vP position and is not required to move to Spec, TP.) Therefore, it does not result in the same competition.

As this is a purely syntactic account, we would expect all unaccusative verbs to be able to undergo there-insertion and all unergative verbs to be ungrammatical in this construction.
However, it has been found that there-insertion only works for a subset of unaccusative verbs (Levin, 1993). For example, both arrive and break are unaccusative verbs. However, it is fine to say (23a) but not (23b).

(23)  
   a. There arrived a man in the garden.
   b. *There broke a vase in the garden.

Researchers have explained this by assuming different underlying structures within the unaccusative class. Deal (2009) argues that if an argument is projected in Spec, vP, there cannot be inserted. In her account, the reason is semantic. When the verbalizer requires a thematic argument, there cannot be inserted because it cannot carry any thematic roles. The verbalizer CAUSE for inchoative verbs (change of state verbs) requires a thematic argument, and therefore, change of state verbs such as break do not allow there-insertion. Alexiadou and Schäfer (2010) has a different account. They argue that the theme argument of unaccusative verbs can occupy different syntactic positions within vP. If the argument is at the Spec, vP, it will compete with there, and therefore there-insertion is not allowed (24b). On the other hand, if the argument is at Spec, ResultP, there-insertion is allowed (24a). The theme of change of state verbs occupies Spec vP while the theme of change of location verbs occupies ResultP. Therefore, there-insertion is blocked for change of state verbs. Although the details differ among different researchers, the general idea is similar, i.e. whether a verb allows there-insertion depends on its specific structure within vP.

(24)  
   a. [vp there [resultP theme]]
   b. [vp *there/theme [resultP ]]  

(Alexiadou & Schäfer, 2010)
If the above account is true, we would expect some unaccusative verbs to be bad with *there*-insertion. More importantly, as the account is still a syntactic one, we would expect all unergative verbs to be bad with *there*-insertion. In the following sections, I show an experiment to test this prediction.

In the experiment, we exclude one structure which has been considered as an exception in the literature: outside verbal *there*-sentences such as (25).

(25) Suddenly there flew through the window that shoe on the table.

(Milsark, 1974)

Deal (2009) argues that such sentences are formed by VP fronting. The landing site of the fronted VP does not have an external argument and *there* can be inserted in that position. In such cases, *there* is available independent of the unaccusativity of the verb. Therefore, in the experiment, we exclude cases of outside verbal *there*-sentences.

2.1. Online judgement test

2.1.1. Subjects

Forty Participants were recruited through Amazon Mechanic Turk (AMT). Three were excluded because they did not provide information about their native language. The rest were all self-reported native speakers of English.

2.1.2. Materials

In the study, 80 verbs were selected from the linguistics literature on unaccusativity and verb classifications (Levin, 1993; Levin & Rappaport Hovav, 1995; Sorace, 2000), among which 42 were unaccusative verbs and 38 were unergative verbs.
Two lists were created. Each list consisted of 80 test sentences with both types of verbs. One of the lists contained plain sentences such as (22a) and (22c) and the other contained sentences with the same verbs in there-insertion construction (22b) and (22d). The plain sentences served as a baseline. We wanted to make sure that the there-insertion sentences were not unacceptable due to semantic or pragmatic reasons. Forty fillers were added to each list. All the fillers are ungrammatical sentences with a transitive verb and only one argument like (26). The fillers give the participants a baseline for ungrammatical sentences.

(26) *The electrician fixed when the lightbulb broke.

2.1.3. Procedure

The participants were randomly assigned to one of the two lists. They were asked to provide acceptability judgment on a 1-7 scale for the sentences, where 1 means completely unacceptable and 7 means totally acceptable. After excluding potential non-native speakers, there were 19 participants remaining in the plain list and 18 remaining in the there-insertion list.

2.1.4. Results

2.1.4.1. Overall group difference

If there-insertion tests the syntactic differences between unaccusative and unergative verbs, we would expect that all our unaccusative verbs are compatible with there-insertion and none of our unergative verbs are. First, we test whether there is a group difference between our unaccusative and unergative conditions. Our dependent variable is the mean rating for each verb in each condition across the participants. Because the dependent variables in the two verb conditions do not follow normal distributions (p’s<0.001), we chose to use Wilcoxon rank sum test (the non-parametric test which does not assume a normal distribution) instead of t-test. Two
plain sentences were judged lower than 5 and caused a marginal difference between unergative and unaccusative conditions \((p=0.06)\). We excluded them and their *there*-insertion counterpart in further analysis. After the exclusion, there is no significant difference between unaccusative and unergative sentences in the baseline condition (Wilcoxon rank sum test \(p=0.10\)). Therefore, we made sure that the sentences are all grammatical and semantically felicitous before *there*-insertion transformation.

As can be seen in Figure 13, all unergative *there*-insertion sentences were judged lower than 5 while a few of the unaccusative sentences were judged 5 to 6. The items judged between 4 to 5 is more in the unaccusative condition than the unergative condition. However, there is no significant difference between unaccusative and unergative conditions (Wilcoxon rank sum test \(p=0.1\)).

![Figure 13](image)

2.1.4.2. Cluster analysis

Having (or lacking) a group difference only tells us whether there is a general difference between unaccusative and unergative verbs. What we really want to know is whether *there*-
insertion can classify unaccusative and unergative verbs into two categories. In this section, we test if the there-insertion can truly classify intransitive verbs into unaccusative and unergative verbs. We use an exploratory clustering analysis to entertain this hypothesis. We first used multidimensional scaling (MDS) by treating the ratings as dissimilarity scores and calculating the distance between the verbs based on their ratings on a two-dimensional space. The distance therefore is a measurement of how different two verbs are. Then, we used hierarchical clustering to find two clusters. Hierarchical clustering works in such a way that every object will first form its own cluster and the two most similar (i.e. closest in Figure 15) clusters will join in the next stage (Figure 14). The procedure is performed iteratively until it forms a unified cluster. We asked the model to generate two clusters based on the hierarchical clusters and the result is shown in Figure 14. If there-insertion does distinguish unaccusative and unergative verbs, the two classes formed by the ratings should coincide with unaccusative and unergative verbs respectively.

Figure 15 shows the configuration plot of the cluster analysis where the distance between two verbs reflect their similarity score and the color shows which verb types they are (red is unergative and black is unaccusative). We can see that the unaccusative and unergative verbs overlap in the plot, meaning that they do not coincide with the unaccusative-unergative classification. We may want to ask how good this classification is in general (i.e. how many verbs are classified correctly). Table 18 shows the classification result of this model. Among all the verbs, 18 unaccusatives and 33 unergatives are classified correctly, 24 of unaccusatives and 3 unergatives are misclassified. The overall accuracy (proportion of correctly classified cases) is 0.65. As we can see, most unergatives are correctly classified while most unaccusatives are misclassified. Is it possible that only a subgroup of unaccusative verbs pass there-insertion
diagnostics? If so, this can explain why some unaccusative verbs are unacceptable with *there*-insertion (although it cannot explain why three unergative verbs are acceptable. They have to be treated as exceptions and have additional accounts).

Table 18 confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UA</td>
</tr>
<tr>
<td>Predicted</td>
<td>UA 18</td>
</tr>
<tr>
<td></td>
<td>UE 3</td>
</tr>
</tbody>
</table>

2.1.4.3. Change of state verbs

As noted by several researchers, change of state verbs are not good with *there*-insertion (Deal, 2009; Alexiadou & Schäfer, 2010). Therefore, we excluded them from the analysis and performed the statistical test again. After the exclusion, there is a significant difference between the unaccusative and unergative conditions (p=0.02). This seems to confirm the intuition that linguists have: *there*-insertion is only acceptable with unaccusative verbs (with an exception of change of state verbs) and not unergative verbs. However, as can be seen in Figure 16, even after excluding change of state verbs, there are still some verbs rated very low within unaccusative condition. If the unaccusative and unergative are truly different from each other, what we would expect is not merely a general group difference but that all unaccusative are acceptable (with exceptions of change of state verbs) and all unergative verbs should be unacceptable with *there*-insertion. This pattern is not borne out.
Figure 16 histogram without verbs of change of state

2.1.4.4. Performance in subgroups

Another possibility is that only some subgroups of unaccusative verbs are acceptable with there-insertion. If so, we would expect them to have higher ratings than other groups. The subgroups are drawn from Sorace (2000), where she classified verbs into subgroups according to their aspectual/thematic characteristics to explain the variability in auxiliary selection. In Sorace (2000) and her other work, Sorace has argued that auxiliary selection in Western European languages is sensitive to a hierarchy of aspectual/thematic verb types and the split in unaccusativity is gradient. To be more specific, some verbs consistently select one auxiliary and others have variable behavior (that is, they sometimes select “BE” and sometimes select “HAVE”). She proposed that the verbs on the two ends of the hierarchy in (27) have more consistent behavior while the verbs in the middle are more variable. It is possible that the diagnostics in English show similar sensitivity to the aspectual/thematic characteristics of the
verbs. If so, we would expect that the rating of there-insertion to vary gradiently along the hierarchy. To test this hypothesis, we classify the verb into subcategories based on Sorace (2000) as is shown in (27).

(27) change of location: come, arrive …  
Change of state: die, happen …  
Continuation of condition/state: stay, remain …  
Verbs of existence (existence of state): exist, sit …  
Uncontrolled processes: tremble, sneeze …  
Controlled motional processes: swim, run …  
Controlled non-motional processes: work, play …

A pairwise Wilcoxon Rank Sum test with Bonferroni correction shows that only one group, i.e. verb of existence (e.g. exist), is significantly different from two other groups: non-motional process (e.g. laugh) and uncontrolled process (e.g. cough) (p’s<0.05). All other subgroups are not significantly different from each other. This means that no other subgroups of unaccusative verbs are better than unergative verbs in there-insertion sentences. As we can see in Figure 17, there is a lot of variation within each subcategory of verbs, but all verbs of existence are rated relatively high.

This result needs to be taken with a grain of salt because splitting verbs into subgroups with reduce the items in each subgroup. Therefore, the null result could simply be due to lack of statistical power.
Figure 17 Ratings of subcategories
The result does show that all verbs of existence perform quite well with the there-insertion construction and differs from two other groups even with the limited statistical power. It is possible that there-insertion has a semantic requirement that the verbs need to have an existential meaning. All other verbs, when acceptable with there-insertion are coerced into an existential reading. With the limited data, we cannot say too much about this possibility. Further research is needed to test this hypothesis.

If it were true that there-insertion only works with verbs of a certain meaning, we may have two possibilities: (1) there-insertion only works with unaccusative verbs with a certain meaning (syntax+semantics); or (2) there-insertion only works with verbs with a certain meaning (only semantics). With both alternatives, we would predict that only a subcategory of verbs (e.g. verbs of existence) works with there-insertion. We cannot necessarily establish a link between there-insertion and unaccusativity of the verb. In fact, it may be redundant to assume an additional syntactic difference if semantics can solely explain the data pattern.

3. Prenominal participles

Prenominal participle refers to the phenomenon that the participle form of the verb can be used as a modifier to a noun, as in (28a). It has been argued that prenominal participle is only possible for unaccusative verbs but not unergative verbs (28b).

(28)  a. The fallen leaves

       b. *the slept boy

Properties of APF

a. Affixation of the passive morpheme -ed

b. Change of category: [ + V, - N] → [ + V, + N]

c. Suppression of the external role of the base verb

d. Externalization of an internal role of the base verb

e. Absorption of Case

f. Elimination of the [NP, VP] position

Their explanation is as follows: when the passive morpheme is affixed to an unergative verb, the external theta role associated with the verb is suppressed due to Rule c. If the passive participle is then converted to an adjective, the adjective is unable to assign an internal theta role as required by Rule d. In fact, because of the suppression, the adjective does not have any theta role. The results in the ungrammaticality of (28b). Contrary to unergative verbs, affixation of the passive morpheme to an unaccusative verb does not affect the assignment of the verb's internal theta role, and, after the participle undergoes conversion to an adjective, this theta role can then be assigned external to the AP headed by the adjectival passive participle. Therefore, the construction is grammatical as in (28a). According to this analysis, this diagnostic is formed by a morphological derivation and expresses a syntactic difference between unaccusative and unergative verbs. This predicts that all unaccusative verbs and no unergative verbs should appear as prenominal modifiers. We tested this prediction with the following experiment.
3.1. Online judgment test

3.1.1. Subjects

Forty Participants were recruited through AMT. Two were excluded because they did not provide information about their native language. The rest were all self-report native speakers of English.

3.1.2. Materials

In the study, the same 80 verbs as those in Section 2 were selected. Two lists were created. Each list consisted of 80 test sentences/phrases with both types of verbs. One of the lists contained plain sentences such as (29a) and (29c) and the other contained phrases with the same verbs in prenominal modifier construction (29b) and (29d). The plain sentences served as a baseline. They are used to make sure that the prenominal modifier constructions were not unacceptable due to semantic or pragmatic reasons.

(29)  

a. The leaves fell.

b. The fallen leaves

c. The boy slept.

d. *the slept boy

3.1.3. Procedure

The participants were randomly assigned to one of the lists. They were asked to provide acceptability judgment on a 1-7 scale for the sentences or phrases, where 1 means completely unacceptable and 7 means totally acceptable. After excluding potential non-native speakers, there were 20 participants remaining in the plain list and 18 remaining the prenominal modifier constructions list.
3.1.4. Results

3.1.4.1. Overall group difference

Our dependent variable is the mean rating for each verb in each condition across the participants. Because the dependent variables of the two verb types do not follow normal distributions (p’s<0.001), we chose to use Wilcoxon rank sum test instead of t-test. Four plain sentences were judged lower than 5. We excluded them and their prenominal modifier counterpart in further analysis. This resulted in the same number of verbs in each verb condition (unaccusatives=unergatives=38). After the exclusion, there is no significant difference between unaccusative and unergative sentences in the baseline condition (Wilcoxon rank sum test p=0.63).

![Figure 18](image)

As can be seen in Figure 18, the unergative prenominal adjective constructions were judged between 1 to 6 with most ratings between 1 to 2. The unaccusative prenominal adjective constructions were judged between 1 to 7, with two peaks at 1 to 2 and 5 to 6. There is a significant difference between the unaccusative and unergative conditions (Wilcoxon rank sum test p=0.63).
test \( p < 0.001 \). This means that the judgments for the prenominal adjective constructions differ between unaccusative and unergative verbs.

3.1.4.2. Cluster analysis

As we have discussed above, finding a group difference does not mean that only unaccusative verbs can be used as prenominal modifiers and the unergative verbs cannot. Similar to the previous section, we tested when we classified the verbs into two groups, whether the groups coincide with unaccusative and unergative verbs respectively. To test this, we performed MDS as described in the previous section.

Figure 19 and Figure 20 shows the results of the clustering. The verbs coded in red are unergative verbs and the verbs coded in black are unaccusative verbs. As can be seen in Figure 20, there are some overlaps between unaccusative and unergative verbs. We compared the results from the real labels of the verbs (unaccusative and unergative based on previous literature), and found the following cases of misclassifications (depending on different researchers, the numbers of expectations vary). Based on Sorace (2000), there are four exceptions: roll, rush, rotate and shake. Based on Levin and Rappaport Hovav (1995), there are two exceptions: rush and shake. They argued that roll and rotate can be unaccusative verbs when their subject is not agentive. In our case, the subjects for these two verbs are inanimate and therefore, they could be unaccusative. No matter which cases count as exceptions, we have exceptions based on either proposal. Therefore, they face the same problems. If the prenominal participle is a purely syntactic unaccusative test, no unergative verbs should appear in this class and all unaccusative verbs should be classified together. However, there are some unergative verbs in this class and 16 unaccusative verbs classified as unergatives.
Figure 19
Figure 20
3.1.4.3. Performance in subgroups

Zaenen (1993) has argued that the unaccusative diagnostics are sensitive to semantic factors in Dutch. Cetnarowska (2002) extended a similar claim to English and argued that verbs of existence are stative and atelic, and therefore their participles are not attributive and unacceptable as prenominal modifiers. To test whether semantic subclass information can explain the distribution of the ratings, we classified the verbs into subclasses based on Sorace, (2000) as it is a hierarchy relevant to aspectual information of the verbs.

If the prenominal participle is a semantically constrained syntactic diagnostic, we would expect that only some subset of unaccusative verbs are acceptable with it and no unergative verbs are acceptable. Figure 21 shows the ratings by subclasses of the verbs. As we can see, continuation of state verbs (e.g. develop) are generally unacceptable as prenominal modifiers. This could be explained by the verbs being stative and atelic. However, even some telic verbs, like change of state (e.g. awake) and change of location (e.g. return), can be unacceptable as prenominal modifiers. As Figure 21 shows, there is a large variance within each subcategory of verbs. Although it made the right prediction that atelic verbs are generally bad with this diagnostic (as can be seen with continuation of state verbs but with some exceptions in verb of existence), subcategorization cannot not fully explain the data pattern.
Figure 21
To summarize, in this section, we have tested two different existing proposals: (1) that prenominal participle construction is a purely syntactic diagnostic and should distinguish unaccusative and unergative verbs; and (2) that prenominal participle construction is a syntactic diagnostic that is constrained by the semantic subcategory of the verbs and therefore only some subcategories of unaccusative verbs can pass it. We have seen that neither holds for our data. As with the results in the there-insertion section, we found that being able to appear as a prenominal participle is neither necessary nor sufficient for a verb to be unaccusative.

4. Nominalization

In this section, I discuss a nominalization processes: -er nominalizations. This diagnostic has been argued to associate with the agentivity of the verb (Sproat, 1985). Therefore, I take it as an example of a diagnostic that characterizes the semantic property of the argument (however, note the counterargument for the relevant of semantic factors, Levin & Rappaport, 1988). Fabb (1984) has argued that the relation between -er nominalization and the verb is characterized by the relation between the external argument and the verb. Keyser and Roeper (1984) extended this argument and proposed that -er formation is blocked by a Move α derivation. Therefore, unaccusative verbs should not be able to have -er nominalization because the subject is the moved internal argument. Levin and Rappaport (1988) inherited the argument that all -er nominals refer to the external argument and therefore it should only be compatible with unergative verbs, as the subject of unaccusative verb is the internal argument. However, they made the observation that verbs of change of state and change of location can also have -er nominals as in (30). They argued that that these verbs can be both transitive and intransitive. The nominals are associated with the transitive uses of the related verbs, not the unaccusative ones.
(30) a. Change of state: opener, dryer, freezer, heater, ...

B. Change of location: rollers, (earth)mover, (pancake)turner, ...

In this section, we test whether -er formation only applies to unergative but not unaccusative verbs.

4.1. Online judgement test

4.1.1. Subjects

Twenty Participants were recruited through Amazon Mechanic Turk (AMT). They were all self-report native speakers of English.

4.1.2. Materials

In the study, the same 80 verbs as in the previous sections were selected. One lists was created. It consisted of 80 test nominals with both types of the verbs as (31) shows.

(31) a. *arriver

b. dancer

4.1.3. Procedure

The participants were asked to provide acceptability judgment on a 1-7 scale for the nominal, where 1 means completely unacceptable and 7 means totally acceptable.

4.1.4. Results

4.1.4.1. Overall group difference

Our dependent variable is the mean rating for each nominal in each condition across the participants. Because the dependent variables of the two verb types do not follow normal distributions (p’s<0.001), we chose to use Wilcoxon rank sum test instead of t-test. There is a
significant difference between unaccusative and unergative nominals (Wilcoxon rank sum test p<0.001).

As Figure 22 shows, most nominalizations of unergative verbs are acceptable while there is a wide range of acceptability for unaccusative verbs.

![Figure 22](image)

4.1.4.2. Cluster analysis

The above result only shows a group difference between unaccusative and unergative verbs. It is not clear if all and only unergative verbs can undergo -er nominalization. To examine what clusters are formed based on the rating, we performed the MDS specified Section 2. The results are shown in Figure 23 and Figure 24. In Figure 24, the red coded verbs are unergatives and the black coded ones are unaccusative. According to the clustering classification, there are following exceptions: *lurk, linger, rise* and *sit* as which are unaccusative verbs are classified as unergatives. Among unergative verbs, there are more exceptions, for example, *cough, rehearse, frown, practice, shiver, pant, sneeze and yawn* are classified as unaccusative. This result is not consistent with the prediction that all and only unergative verbs have -er nominalization.
Figure 23
4.1.4.3. Performance in subgroups

In this section, we test whether the results of nominalization can be predicted by the subgroup information of the verbs.

As Figure 25 shows, most unergative verbs (in the subcategory of uncontrolled process, motional process and non-motional process) are acceptable with -er nominalization. There is a wide range of acceptability within each subcategory of unaccusative verbs. In the next step, we took some noise out of the data by excluding the items that could be acceptable/unacceptable for other reasons. The items a listed below in (32).

(32) a. Nominals that have a lexical entry with different a meaning: ascender, liver, hanger, sleeper, lier(liar)

b. nominals that are spelled differently in the lexicon: surviver

c. nominals that have a transitive counterpart: bender, breaker, developer, opener, returner, roller, shaker

After the exclusion, there are 67 verbs left, 32 of which are unaccusatives and 36 are unergatives. Although there is still a significant difference between unaccusative and unergative nominals (Wilcoxon rank sum test p<0.001), the nominalization diagnostic does not differentiate unaccusative and unergative verbs symmetrically. Figure 26 shows the results after the exclusion. There is still a wide range of most unaccusative subcategories (change of location, continuation of state and verb of existence). In addition, the overall acceptability of the uncontrolled motional process dropped after the exclusion. This result is unexpected given the theory of nominalization described above. The theory predicts that all unergative verbs and no unaccusative verbs should be acceptable with -er nominalization.
Figure 25
Figure 26
4.2. Congruity between diagnostics

Levin and Rappaport (1988) claimed that the classification based on -er nominalization should be consistent with the classification based on acceptability of prenominal past participle, i.e. the verbs that have -er nominalizations should not be able to appear as a prenominal past participle, and vice versa. Fifty-nine verbs remain after combining the results of the two diagnostics. We compared the results of these two diagnostics. Only 9 verbs that receive a high rating as a prenominal participle received a low rating for -er nominalization, and 22 verbs which received a low rating as a prenominal participle received a high rating for -er nominalization. Twenty-eight verbs did not follow the above prediction, in that they either received high ratings or low ratings in both diagnostics.

Table 19

<table>
<thead>
<tr>
<th></th>
<th>Nominalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>prenominal</td>
<td>UA</td>
</tr>
<tr>
<td>UA</td>
<td>9</td>
</tr>
<tr>
<td>UE</td>
<td>27</td>
</tr>
</tbody>
</table>

We can already see that the results of the two diagnostics do not seem to be congruent. Could this be due to random noise in the data? To understand how much mistakes are due to subject noise in these two tests, we randomly split the data into two groups by subjects. Each group has half of the subjects. We classified the verbs based on two halves of the subjects and compared the results between these two classifications. The results will give us a baseline of classification inconsistency due only to subject noise given the same diagnostic.

Table 20 shows the results of the prenominal adjective diagnostic. The accuracy is 0.94. Table 21 shows the results of the nominalization diagnostic. The accuracy is 0.81. Both are much higher
than the accuracy in Table 19, which is 0.46. Assuming that 20% of the misclassification could be due to subject noise in the data, the accuracy of the cross-diagnostic classification is still too low. Another possible source of noise is the item effect. Because the -er nominalization and prenominal modification diagnostics are very different linguistic constructions, it might be possible that the acceptability varies across these two studies because the lack of control of the stimuli. However, because of the nature of these two diagnostics, the stimuli cannot be equated, and the noise is hard to quantify. One way to quantify the item effect, assuming that the form of the linguistic constructions per se does not affect the judgment, is to run the same diagnostic with same verbs but different items. We leave that to future research.

Table 20

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>Class 2</td>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 21

<table>
<thead>
<tr>
<th></th>
<th>Class 1</th>
<th>Class 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>Class 2</td>
<td>13</td>
<td>22</td>
</tr>
</tbody>
</table>

5. Discussion

5.1. Theoretical controversy

The diagnostics tested in this chapter are widely cited and discussed in the linguistic and psycholinguistic literature (Fabb, 1984; Levin and Rappaport, 1988; Friedmann, Taranto, Shapiro, and Swinney, 2008; Deal, 2009; Irwin, 2012; a.o.). However, they remain controversial after decades of discussion. In the previous sections, when describing the theoretical explanations for the diagnostics, I discussed the reasons for supporting the constructions as
unaccusative/unergative diagnostics. These explanations and the diagnostics are nonetheless under debate. In this section, I discuss the counterargument for treating the constructions as true unaccusative diagnostics.

5.1.1. Controversy of *there*-insertion as an unaccusative diagnostic

Levin and Rappaport Hovav (1995) have discussed the problems of locative inversion as an unaccusative diagnostic. The same problems exist in the *there*-insertion construction. The problems are exactly as we discussed in the relevant section. First, not all unaccusative verbs can appear in this construction. Second, the most serious problem with *there*-insertion is that some unergative verbs can enter this construction. Levin and Rappaport Hovav (1995) have argued that the presence of unergative verbs in this construction cannot be explained by multiple meanings for the verb involved (so that under some meanings, the verb can actually be unaccusative), because the verbs in this construction do not have meaning shift.

Levin and Rappaport Hovav (1995) provided an alternative explanation on why locative inversion and *there*-insertion have the unaccusative-diagnostic-like property. Their explanation is a discourse function that restricts the verbs that can appear in it. Based on the previous observations (Hartvigson & Jakobsen, 1974), Levin and Rappaport argued that the verbs in these constructions need to be informationally light, meaning that “if the verb contributes information that was not predictable from the context, it would detract from the newness of the information conveyed by the post-verbal NP.” This discourse function restricts the meaning of the verb that can appear in *there*-insertion and prefers verbs of existence and appearance, which happen to be unaccusative verbs. Because of this coincidence, *there*-insertion appears to pick out unaccusative verbs. However, this construction is not restricted to unaccusative verbs. Levin and Rappaport showed that under the appropriate circumstances, a wide variety of unergative verbs can appear
in this construction as well. They illustrated this point by citing a verb inversion examples from Bolinger (1977) as in (33).

(33)   a. Waving gaily was a bright flag.

       b. ?Burning merrily was an enemy flag.

In the above example, waving is a normal motion associated with the flag and therefore does not contribute much new information than stating the existence of the flag. On the other hand, burning is not the characteristic of the existence of the flag and therefore is not informationally light. In the same paper Bolinger (1977) showed that this observation can be extended to locative inversion as in (34). The example therefore shows how the discourse function can extend to unergative verbs. This account is also compatible with the judgment data we saw in the earlier section where the verbs of existence are generally acceptable with there-insertion.

(34)   Out of the precipice behind waved a profusion of feathery rock-lilies … [R. Praed, Outlaw and Lawmaker, 259]

(adopted from Levin and Rappaport, 1995)

The discourse function approach may be hard to evaluate as it is not clear how to quantify and compare information heaviness. My point is not to argue that the discourse approach is a better theory. I simply want to show an example of the controversy and alternatives that have been raised in the literature of there-insertion. Although widely cited and discussed in the literature, using unaccusativity to explain the acceptability of the verbs in there-insertion has by no means been established as the best approach and is by no means uncontroversial.
5.1.2. Controversy of prenominal past participle as an unaccusative diagnostic

Pesetsky (1996) argued that not all unaccusatives form adjectival past participles. A verb can have an adjectival past participle when it has two participants/arguments: the Theme and a controlled A(mbient) Causer. An A-Causer is the external argument and expresses the internal source/cause of the event, e.g. in the example elapsed time, “something intrinsic to the time causes it to elapse.” In addition, he argued that in a language like English, the phenomenon is not fully productive. Cetnarowska (2002) and McIntyre (2013) argued against this claim and contended that this process is actually productive. McIntyre (2013) observed that although the process is productive, it is constrained: the participle requires “salient, relatively stable result states”. This can explain the contrast of grown, “since men but not trees have culturally recognized full-grown states”.

Orthogonal to the productivity debate, some questions have been raised about characterizing the phenomenon as passive participle as in Levin and Rappaport (1986). Although Levin and Rappaport treated the participle as passive participle, they pointed out that intransitive verbs generally do not have passive forms and therefore the participle may be the perfect participle and not the passive participle. Although this solution solves the problem of assuming intransitives have passive forms, it also loses the generalization that could be captured in this account. The original account treats the participle as passive participle and therefore can assume that adding the -ed affix causes the suppression of the external role of the base verb and absorption of Case, which characterizes the passive formation. However, if we assume that the -ed affix is associated with perfect participle, these two phenomena cannot be explained with the same account. In addition, we are left to explain why this process has the same effect as passive
formation, because we do not have suppression of the external role of the base verb and absorption of Case for free as in passive formation.

5.1.3. Controversy of -er nominalization as an unergative diagnostic

Kaufmann (1995) argued that the restriction on -er nominalization is not unaccusativity but semantics and pragmatics. She argued that in the nominalization, there is a situation variable to be bound. However, in -er nominalization the base verb does not provide any information so that the situation could be “temporally anchored”. This leads to the restriction on -er nominalization: if no information is given to temporally anchor the situation, the situation variable must be bound by contextual knowledge in the following ways: (1) the referent displays the characteristic property in a specific situation that can be identified by the speaker and the audience, e.g. “the interrupter over there”, (2) the referent displays the characteristic property habitually, i.e. is regularly involved in such an event, e.g. “writer”. In the case of unaccusative verbs, which Kaufmann argued to “encode changes of object defining properties” (e.g. color, shape, etc.), -er nominalization is pragmatically inadequate. In the specific-situation, the verb rarely provides a better description than the noun modified by the present participle of the verb. For example, “falling block” instead of “faller” is much more informative of the situation. And therefore, the -er nominalization is pragmatically inadequate. In the habitual situation, the meaning of the verb (i.e. encoding changes of object defining properties) is not specific enough to give a non-redundant description of the individual. Therefore, in both cases, -er nominalization is not chosen to characterize the participant of an unacusative event. With the above characterizations, Kaufmann can reduce the explanation of -er nominalization to semantics and pragmatics with no reference to syntax. She argued that if the phenomenon can be explained with semantics and pragmatics only, assuming a syntactic difference becomes redundant.
5.2. Empirical problems

In addition to the theoretical controversy, the diagnostics also face empirical challenges. I have shown, in the previous sections, that no diagnostics can truly distinguish unaccusative verbs and unergative verbs. In each diagnostic, we have both types of verbs pass and fail the diagnostic. Even a syntactic explanation coupled with semantic subcategorization cannot explain the data pattern because: (1) if the diagnostics are truly syntactic tests, no unergative verbs (or unaccusative verbs depending what the diagnostic is testing) should pass them; (2) in most diagnostics, semantic subcategorization alone cannot explain the distribution of the data. There is a wide range of variation within each semantic subcategory.

In addition, the results from different diagnostics are not consistent with each other. It has been argued that for English, different diagnostics pick out different subclasses (Levin & Rappaport Hovav, 1995). Nevertheless, there are also diagnostics that are argued to give compatible results such as -er nominalization and prenominal past participle (Levin & Rappaport, 1988). However, I have shown that, even for these two tests, the classifications are quite different. Given the variability in the data and the fact that the diagnostics do not give consistent results, there is no strong reason to believe that there are two unified classes of intransitive verbs.

6. Summary

In this chapter, I reviewed three diagnostics that have been proposed as diagnostics for unaccusativity. I have shown with judgment data that, despite wide-spread acknowledgement of these phenomena as unaccusative diagnostics and the frequent adoption of them in the literature, these diagnostics cannot differentiate unaccusative and unergative verbs. In addition, there are theoretical controversies for each diagnostic that are worth further consideration. Most
alternative accounts try to use semantic factors to account for the distribution of the verbs. In fact, all the theoretical explanations have to combine some semantic factors even starting from the initial proposal in Perlmutter (1978). As Kaufmann (1995) argued, it is not clear what syntax adds to these theory. One difficulty for any explanations that involve syntax is that they should predict that all unergative verbs are not permitted in there-insertion or as prenominal modifiers, but this is not the case. One possibility is that the labels for unaccusative and unergative verbs is wrong. The verbs that passed the diagnostics are actually unaccusive verbs. This not entirely impossible. Actually, different researchers may have different classifications and different semantic subcategorizations (e.g. Levin & Rappaport Hovav, 1995; Sorace 2000). If so, we have to rely on the diagnostics to classify the verbs. However, the establishment of these diagnostics is somewhat circular. It is first found that the phenomena can distinguish some unaccusative from unergative verbs. Then, these phenomena are used to classify verbs in turn. One argument to rescue the practice from circularity is that these phenomena give consistent results and therefore can confirm the classification of other phenomena and serve as independent criteria. However, this is not the case in English. As has been shown above, these diagnostics do not pick two uniformed classes. If we do not have independent criteria to evaluate the diagnostics and must identify them by testing some unaccusative and unergative verbs, we are left with no starting place.

I have chosen these three phenomena as representative examples of unaccusative diagnostics. They each characterize different combinations of dimensions; surface (there-insertion) vs. deep unaccusativity (prenominal participle and -er nominalization); syntactic (there-insertion) vs. morphological derivation (prenominal participle and -er nominalization); syntactic (there-insertion and prenominal participle) vs. semantic (-er nominalization) aspects of
the verb. There are some other phenomena that have been argued as diagnostics for unaccusativity, for example, resultative constructions (35) and causative alternation (36). Causative alternation is restricted to a semantic subcategory of verbs: change of state. Therefore, it is unclear whether the phenomenon is due to the syntax or semantics of the verbs. Resultative constructions are highly restricted by context (which adjective can work in the context). It is easy to get floor effects (all the items judged as acceptable due to the inappropriate adjective). How to evaluate this phenomenon needs more consideration.

(35)  
   a. The river froze (*itself) solid
   b. the boy laughed *(himself) hoarse

(36)  
   a. The boy broke the vase
   b. the vase broke

Due to the limitation of resources, I have limited the evaluation of the diagnostics to English. How unaccusative diagnostics work in other languages awaits further investigation.

7. Limitations of the work

As the first study that attempts to study unaccusative diagnostics with relatively large-scale judgment data from naïve speakers, this work has many limitations and space for improvements. Initially, it was not clear which factors we should control and how important they may be before we conducted the study. If they were purely syntactic diagnostics, we would expect a bimodal distribution. The only factor we would care about is that the stimuli obeys other grammatical restrictions and that they are semantically/pragmatically felicitous. The stimuli therefore have not been carefully controlled, as the attempt was trying to make the sentences are natural as possible. This decision leaves room for many variabilities for semantic and pragmatic
factors which could influence the judgment on the individual items and could result in differences between different items. The results simply serve as pilot data to show the existence of the problem (i.e. the diagnostics do not differentiate unaccusative and unergative verbs as the unaccusative hypothesis predicts).

There are some other limitations on the creation of the stimuli because of the limited resources and uncertainty of the prediction of the theory. One limitation is that the verbs are drawn from the previous literature on unaccusativity or verb classifications. This may create a sampling bias. Most of these verbs have been tested in the literature by previous researchers, and therefore, their performance may conform with the theory better. Using the verbs to test the theory that were generated by studying them also creates a circularity in the argument. Therefore, they may not be representative of all intransitive verbs. A more random set of verbs may show different behaviors. The point of this work is to show that even these verbs which should be more compliant with the theory do not behave as the theory predicts. It should give us to a reason question what we may find with a more random sample of verbs.

Another limitation of the stimuli is that there are some factors that are ignored in this work. For example, it has been observed that the judgment of prenominal modifiers may be improved if there is an adverb before the participle as in (37).

(37)  ?the arrived book

the recently arrived book

However, there is no clear theoretical account about why it is so. In addition, the adverb is context dependent. We cannot insert a single adverb for all the stimuli. The specific adverb may influence the judgment of the item and this creates more variance in the stimuli. Therefore, we chose not to add adverbs in our stimuli. Whether this is the best approach can be debated.
There are also many other design considerations that should be discussed. For example, we separated the two lists in each diagnostic and did not counterbalance the stimuli. This is because the transformed sentences are usually rated lower than the plain sentences. Putting them in the same study may cause the participants to limit the rating of the transformed sentences to a smaller proportion of the scale and this could diminish some potential differences. This decision made the ratings from the plain sentences and the transformed sentences not comparable because the participants may use the scales differently. Therefore, in our analysis, we never try to compare them directly. We simply used the plain sentences as a reason to exclude sentences in the transformed condition. Whether it is better to analyze the data differently so as to use difference scores between these two conditions or to rescale the data, or even to put the ratings in these two conditions into the regression model, are all issues to be discussed in the future work.
Chapter 5

CONCLUSION

1. On the syntactic account for unaccusativity

In the previous chapters, the discussion has been focused on the original Unaccusative Hypothesis, which assumes a syntactic difference between unaccusative and unergative verbs. This hypothesis assumes that the sole argument of unaccusative verbs originates from the object (internal argument) position and is realized as the subject (external argument) position through movement. This has been a popular view (and probably the dominant view within the generative literature), and most psycholinguistic studies have been designed based on the assumption that there is a movement in unaccusative sentences. There are two major types of evidence that support the Unaccusative Hypothesis: (1) the linguistic behaviors of unaccusative and unergative verbs, i.e. that these two types of verbs appear in different linguistic constructions, which has been argued to reflect their structural difference; (2) the processing difference found in the psycholinguistic studies.

In the previous chapters, I have evaluated these two types of supporting evidence for the Unaccusative Hypothesis. In Chapters 2 and 3, I have shown that no processing difference can be found with eye-tracking studies in either English or Dutch. With a careful review of the existing processing evidence from other paradigms, I have shown that the results are not consistent or reliable among these studies. This means that we do not have valid evidence for a processing
difference between unaccusative and unergative verbs. In Chapter 4, I investigated three unaccusative diagnostics in English and showed that these diagnostics, contrary to what has been claimed in the literature, cannot distinguish unaccusative and unergative verbs. They can, at best, pick out some subset of unaccusative verbs (together with a few unergatives) but the verbs picked out from the diagnostics are not consistent. These results, however, cannot be used to support an overall syntactic distinction between unaccusative and unergative verbs. Taken together, no empirical evidence has been found in my investigation to support the Unaccusative Hypothesis. Not finding evidence for the hypothesis does not mean that the hypothesis is wrong. After all, the unaccusative diagnostics do pick out some specific verbs and reveal differences among different verbs. The question is what do they actually do, and whether there are better explanations than the Unaccusative Hypothesis. In addition, we would want to consider what reasons there are to abandon the Unaccusative Hypothesis.

The Unaccusative Hypothesis was proposed in the context of Relational Grammar (Perlmutter, 1978) and has been adopted by the early Government and Binding theory (Burzio 1981, 1986). As has been reviewed in Chapter 1, at the early age of Government and Binding Theory, this hypothesis was consistent with the assumption that the identical thematic relationships between items are represented by identical structural relationships between those items at the level of deep structure (Uniformity of Theta-Assignment Hypothesis, Baker, 1988), which provides a straightforward link between the positions of the arguments and the thematic roles they take. However, with the recent development of the theory, the distinction between the internal and external argument becomes less clear as they are both based generated in the VP domain as the specifiers. The distinction becomes even more unclear in the intransitive cases because there is only one argument and we do not even face the problem of argument ordering,
which could potentially distinguish the arguments based on their orders. It is unclear how, in the recent theory of syntax, the movement account of unaccusativity can account for the distinction between unaccusative and unergative verbs, as they both need to move for feature checking.

The only way to explain the thematic difference between unaccusative and unergative arguments may be that unaccusative and unergative verbs have different functional heads, as assumed in the Minimalist Program. The distinction of these two types of verbs is made by the position where they are based generated (inherited from the early GB). Like in (38) v is the head that introduces the agent role and V is the head that introduce the theme role. An unaccusative does not have the vP projection and therefore, it only takes a theme argument. An unergative does not fill in the specifier of VP, so the only argument goes to the specifier of vP and therefore, the argument is an agent. In this account, movement or the distinction between internal and external argument are not important for unaccusativity.

(38)

As has been mentioned in Chapter 1, we do not necessarily expect a processing difference with such an account. In this case, we would not assume that the theme NP will move to the spec, vP otherwise it will be assigned an agent role which causes a role conflict (the same argument being assigned two thematic roles). Therefore, in this account, both unaccusative and unergative verbs should only have one movement (movement to Spec, TP). As Chapter 1 shows,
there are two slightly different structures posited for unaccusative verbs. In one of them, the argument originates in the complement position whereas in the other, the argument originates in the Spec, VP position. In the latter case, we do not predict a difference in the processing pattern.

In the former case, the situation is more complex. If the unergative argument originates in the specifier position of vP and the unaccusative argument originates from the complement of VP, we seem to predict that the unaccusative argument should be reactivated postverbally and the unergative argument should be reactivated preverbally. However, the participants do not know which types of verbs they encounter before they identify the verb. Therefore, the preverbal position cannot be a place to distinguish the unaccusative condition from the unergative condition. The participants either need to assume that all the verbs in the study are unergatives and reactivate every subject (in which case both unaccusative and unergative subjects are reactivated preverbally), or they need to wait until they hear the verb to reactivate the argument (in which case both unaccusative and unergative subjects are reactivated postverbally – this result is the same as assuming all verbs are unaccusative). The only scenario in which these two conditions could differ is if the verbs are identified much earlier than the end of the verbs. In this case, the unergative subjects could be reactivated earlier than the unaccusative verbs because the unaccusative subjects are reactivated after the verb. Overall, we do not necessarily predict a difference between these two conditions. This is consistent with the data we saw in Chapter 2 and 3 where no processing difference was found between these two types of verbs. Does this mean that this is the correct account for unaccusativity?

This account inherits the Unaccusative Hypothesis in that it preserves a one-to-one mapping between structural position and thematic roles. vP is linked to the agent role and VP is linked to the theme role. One benefit of this account is that it gives us a straightforward account
for causatives. For example, the verb *break* can be both an intransitive and transitive verb. When it is intransitive, it takes a theme argument. Using the above analysis of thematic roles, we can postulate the structures in (39) (details of derivation omitted). When *break* is intransitive, the argument is assigned the theme role by the V and when *break* is transitive, there is a light-verb *cause* that introduces an agent argument.

(39)

a.  

b.  

However, this also has several problems. For example, the unergative structure is the same as a transitive in (38). It has two argument positions to fill. Not filling one of them will cause a violation of theta criterion which requires a one-to-one correspondence between the arguments and thematic roles. One way to solve this problem is to assume that the object argument of the unergative verb is incorporated into the verb (Hale and Keyser, 1993) and therefore this position is not available to be filled. Although this explanation is possible, this stipulative incorporation makes the account less parsimonious which is against the intention to propose the one-to-one mapping rule.

More than the loss of parsimony, this account faces some empirical problems. The above account assumes a structural difference between unaccusative and unergative verbs which predicts that these two types of verbs should perform differently in the syntactic tests. This is not
the pattern we found with the three unaccusative diagnostics we tested. In addition to unaccusative mismatches where different diagnostics pick out different sets of verbs within and across languages (van Voorst, 1985; Levin, 1986; Dowty 1991; a.o.), we found that the diagnostics sometimes pick unergative verbs. In addition, sometimes the same verb can be both unaccusative and unergative. For example, Glushan (2009) argued that the unaccusativity of a verb can change based on the animacy of its argument in Russian. When the argument is animate, the verb is an unergative and when the argument is inanimate, the verb can be unaccusative. As shown in (40), the verb *fall* can only pass the unaccusative po-test in Russian when the subject is inanimate. Such variable behaviors have been noted in other languages such as Italian as in (41).

(40) a. Po jabloku upalo s každogo dereva

Po apple\textit{dat} fell from each tree

‘an apple fell from each of the trees’

B.*po sportsmenu upalo s každogo trenažera

Po sportman\textit{dat} fell from each machine

‘A sportman fell from each of the machines’

(examples form Glushan 2009)

(41) a. Mario ha continuato

Maria has continued

‘Mario continued.’

b. Il dibattito e continuato.

The debate is continued.

‘The debate continued.’

(examples form Rosen 1984)
If we assume a syntactic difference between unaccusative and unergative verbs, we have to assume that each of these verbs has two syntactic structures (or lexical entries). However, it is rather stipulative because the meanings of the verbs are very similar in their unaccusative and unergative usages. Alternatively, we could assume that these two diagnostics actually reveal some semantic differences between those sentences, such as whether the subject is more agentive or theme-like.

To summarize, the new syntactic account for unaccusativity has an advantage, which is that it preserves a one-to-one correspondence between structural position and thematic roles as in the original Unaccusative Hypothesis. Nevertheless, it faces several problems. On the one hand, it needs to make several stipulations which are hard to justify with external evidence (such as an assumed lexical difference) and adds complexity to the theory. On the other hand, and more problematic, the data pattern from the unaccusative diagnostics are not consistent with the prediction of the theory. One option is that we try to explain the data pattern by adding more assumptions and semantic factors (as in Chapter 4). Another option is to consider alternative explanations.

2. On semantic accounts for unaccusativity

An alternative account that has been proposed in the literature is the semantic account. Under this account, no syntactic distinction is necessary to explain the difference between the unaccusative and unergative verbs. The difference should be explained by semantic factors. Even the researchers that support the syntactic account have agreed that unaccusative and unergative verbs differ in their meanings since the onset of the Unaccusative Hypothesis (Perlmutter, 1978). The question is, can we get rid of the syntactic part and use semantics as the sole factor for unaccusativity?
Van Valin (1990) has claimed that all unaccusative diagnostics can be accounted for by semantic factors with no reference to a syntactic distinction between unaccusative and unergative verbs. In his account, the different behaviors of unaccusative and unergative verbs can be explained by semantic factors such as aspectual properties (Aktionsart). For example, he explained the auxiliary selection in Italian by the rule “select essere if the Logical Structure of the verb contains a state predicate”. A state verb is of the form predicate’ (x) or (x, y), for example, stare 'stay' be-at' (x, y). To some degree, the Logical Structure looks like structure of light-verbs. Therefore, this account can be similar to the decomposition account in Minimalist Program. However, in the semantic account, the predicate structure does not need to be realized in the syntactic structure. It is at the semantic/conceptual level. It is only a matter of terminology to call a certain representation syntactic or semantic because the light verb bears meanings. What matters is whether putting it at a different level of representation gives us different predictions. If we assume that the semantic factor has binary values, the prediction is similar to a syntactic distinction, i.e. the verb needs to be either A or B. If we allow the values to be continuous, we would expect more gradient behaviors of the verbs. Sorace (2000) has found that the auxiliary selection is gradient rather than binary, which could be a possible result of a meaning continuum.

So far, we have been working under the assumption that there are two types of intransitive verbs. However, we see that a strict dichotomy is not reflected in the data from the processing studies or from the judgment of diagnostics. An advantage of the semantic account is that it can create rich representations and therefore, the distinction does not have to be binary. There can be multiple subgroups of verbs. This is more consistent with phenomena observed as unaccusative mismatches where different diagnostics pick different subclasses of verbs. It is possible that each semantic factor picks out a unique class of verbs that have a certain meaning.
3. Linking meaning and structure

Both the syntactic and the semantic accounts assume that there are two classes of intransitive verbs: one with agent arguments and the other with theme arguments. In the last section, I introduced another possibility: that is there are multiple subgroups of verbs. Although multiple groups have been proposed (from Permutter 1978 to Levin & Rappaport 1995 to now) in the literature, they are proposed as subgroups under either unaccusative and unergative verbs. There are still two large groups differentiated by the unaccusativity. This is because there are two types of semantic roles (agent and theme) associated with the intransitive verbs.

However, the semantic roles are defined by the meaning and have a certain arbitrariness. It is not clear whether they are the most structure-relevant definitions. For example, the definition below from Haegeman (1994) is one possible definition of agent and theme. As can be seen, theme is the umbrella term for a state or anything affected by an action. These are two different relationships to the predicate. It is possible that these two types of themes differ linguistically.

AGENT/ACTOR: the one who intentionally initiates the action expressed by the predicate.

THEME: the entity affected by the action or state expressed by the predicate.

Instead of treating agent and theme as two distinct thematic roles, Dowty (1991) treated them as prototypes. He proposed several properties for a prototypical agent and prototypical theme as shown in (42). Under this proposal, the arguments which have an agent role can satisfy different agent properties.
(42) contributing properties for the agent proto-role:

a. volitional involvement in the event or state

b. sentence (and/or perception)

c. causing an event or change of state in another participant

d. movement (relative to the position of another participant)

(e. exists independently of the event named by the verb)

Contributing properties for the Patient Proto-Role:

a. undergoes change of state

b. incremental theme

c. causally affected by another participant

d. stationary relative to movement of another participant

(e. does not exist independently of the event, or not at all)

Therefore, instead of assuming that different unaccusative diagnostics pick out different subclasses of unaccusative verbs, we could instead assume that each diagnostic picks out a class with a distinct semantic property. No larger classes are needed for this analysis. It is possible that this semantic factor coincides with a contributing property of a proto-role. However, it does not mean that the larger categories defined by the two proto-roles are relevant to the behavior of the verb in the diagnostic.

Which aspects of meaning are relevant to the structure is still an open question. However, from our investigation, we can learn that a syntactic or semantic dichotomy cannot explain the
behaviors of the intransitive verbs in various linguistic constructions. There is no simple mapping between the semantic role and the structure of the verbs.

We can have two possible mapping rules. One is that there is a one-to-one correspondence between the arguments introduced by the light-verbs and the thematic roles they take. This has been discussed in the previous chapters and sections. Another is that there is no one-to-one correspondence between structural position and thematic roles. The argument can take different roles in the same position depending on the meaning of the verb. That is, laugh and fall do not need to have different structures. Their arguments simply have different thematic roles. These two possibilities cannot be differentiated by the data currently available.

4. Future directions

The data we have do not show a clear dichotomy of intransitive verbs. It may well be that there are multiple classes instead of two. It is an empirical question whether postulating multiple classes explains the data better than two classes. It is also an empirical question which factors are the most relevant to the syntactic behaviors of the verbs (such as what construction they can appear in). The data collected from judgments of the diagnostics can help answer some of these questions. We can try to separate the verbs into multiple classes and see if the grouping could reduce the variance within each group and if the groupings are more meaningful. As the verbs used in this dissertation were selected from the literature of the unaccusative diagnostics, they may not give the most reliable results because there is some sampling bias (the verbs may support the classification in the literature). It will be helpful to apply unaccusative diagnostics to a larger sample of verbs that are selected randomly and test what kind of classes they form.

To test whether intransitive verbs are classified into two large classes and what properties they have, we could decompose the meaning of the verbs by asking participants to judge whether
a verb has a certain property (e.g. defined in Dowty, 1991) (see Kako, 2006 for the application to the transitive verbs). The correlation of the results from this study and the data from the diagnostics could tell us which aspects of meaning are relevant to their behavior in the diagnostics. It is possible that a more fine-grained theory of meaning-form mapping could account for the behavior of the verbs in the unaccusative diagnostics that a two-way distinction cannot explain. Whether this approach has better explanatory power awaits further investigation.
Appendix 1: Stimuli for the English Visual World Paradigm Study

1. The clown from an eastern province in Canada secretly arrived when the committee announced the winner of the annual contest.

2. The cook from an eastern province in Canada secretly arrived when the committee announced the winner of the annual contest.

3. The clown from an eastern province in Canada secretly sighed when the committee announced the winner of the annual contest.

4. The cook from an eastern province in Canada secretly sighed when the committee announced the winner of the annual contest.

5. The baker from a small town in North Carolina supposedly lurked around the bank after the power suddenly went out.

6. The sailor from a small town in North Carolina supposedly lurked around the bank after the power suddenly went out.

7. The baker from a small town in North Carolina supposedly snuck around the bank after the power suddenly went out.

8. The sailor from a small town in North Carolina supposedly snuck around the bank after the power suddenly went out.

9. The librarian from a working-class suburb of Los Angeles abruptly awoke in the dark when the thunderstorm suddenly struck.

10. The astronomer from a working-class suburb of Los Angeles abruptly awoke in the dark when the thunderstorm suddenly struck.

11. The librarian from a working-class suburb of Los Angeles abruptly shouted in the dark when the thunderstorm suddenly struck.

12. The astronaut from a working-class suburb of Los Angeles abruptly shouted in the dark when the thunderstorm suddenly struck.

13. The athlete from a big family in Florida coincidentally resided in London when the 2012 Olympics took place.

14. The singer from a big family in Florida coincidentally resided in London when the 2012 Olympics took place.

15. The athlete from a big family in Florida coincidentally worked in London when the 2012 Olympics took place.
16. The singer from a big family in Florida coincidentally worked in London when the 2012 Olympics took place.
17. The poet from the most prestigious area of Paris quickly emerged from the building as the parade passed by.
18. The tourist from the most prestigious area of Paris quickly emerged from the building as the parade passed by.
19. The poet from the most prestigious area of Paris quickly rushed from the building as the parade passed by.
20. The tourist from the most prestigious area of Paris quickly rushed from the building as the parade passed by.
21. The architect with a fashionable bag and an umbrella immediately disappeared when the waitress brought the expensive bill.
22. The nurse with a fashionable bag and an umbrella immediately disappeared when the waitress brought the expensive bill.
23. The architect with a fashionable bag and an umbrella immediately hesitated when the waitress brought the expensive bill.
24. The nurse with a fashionable bag and an umbrella immediately hesitated when the waitress brought the expensive bill.
25. The captain in a grey coat and dark blue tie abruptly rose when the senator entered the lecture hall.
26. The judge in a grey coat and dark blue tie abruptly rose when the senator entered the lecture hall.
27. The captain in a grey coat and dark blue tie abruptly whispered when the senator entered the lecture hall.
28. The judge in a grey coat and dark blue tie abruptly whispered when the senator entered the lecture hall.
29. The farmer in the center of the Crimean Peninsula surprisingly remained when the troops evacuated the residents in the area.
30. The fisherman in the center of the Crimean Peninsula surprisingly remained when the troops evacuated the residents in the area.
31. The farmer in the center of the Crimean Peninsula surprisingly cried when the troops evacuated the residents in the area.
32. The fisherman in the center of the Crimean Peninsula surprisingly cried when the troops evacuated the residents in the area.

33. The banker from Detroit, the city of decay, unexpectedly stayed when most families moved to the Northeast.

34. The lumberjack from Detroit, the city of decay, unexpectedly stayed when most families moved to the Northeast.

35. The banker from Detroit, the city of decay, unexpectedly complained when most families moved to the Northeast.

36. The lumberjack from Detroit, the city of decay, unexpectedly complained when most families moved to the Northeast.

37. The astronaut from the mountainous area of Colorado unobtrusively appeared when the long-winded chairman was giving his annual speech.

38. The driver from the mountainous area of Colorado unobtrusively appeared when the long-winded chairman was giving his annual speech.

39. The astronaut from the mountainous area of Colorado unobtrusively slept when the long-winded chairman was giving his annual speech.

40. The driver from the mountainous area of Colorado unobtrusively slept when the long-winded chairman was giving his annual speech.

41. The musician with no interest in the theory of justice conspicuously vanished when the philosopher started a lengthy criticism of John Rawls' theory.

42. The magician with no interest in the theory of justice conspicuously vanished when the philosopher started a lengthy criticism of John Rawls' theory.

43. The musician with no interest in the theory of justice conspicuously whined when the philosopher started a lengthy criticism of John Rawls' theory.

44. The magician with no interest in the theory of justice conspicuously whined when the philosopher started a lengthy criticism of John Rawls' theory.

45. The biologist with the big black sport utility vehicle instantly departed when the Yellowstone National Park closed for the day.

46. The photographer with the big black sport utility vehicle instantly departed when the Yellowstone National Park closed for the day.

47. The biologist with the big black sport utility vehicle instantly grumbled when the Yellowstone National Park closed for the day.
48. The photographer with the big black sport utility vehicle instantly grumbled when the Yellowstone National Park closed for the day.

49. The butcher with a great passion for the Super Bowl suddenly collapsed when the Denver Broncos scored in the last minute of the game.

50. The hairdresser with a great passion for the Super Bowl suddenly collapsed when the Denver Broncos scored in the last minute of the game.

51. The butcher with a great passion for the Super Bowl suddenly swore when the Denver Broncos scored in the last minute of the game.

52. The hairdresser with a great passion for the Super Bowl suddenly swore when the Denver Broncos scored in the last minute of the game.

53. The fireman from a small village in Akron, Pennsylvania dramatically died when the tornado destroyed the entire neighborhood.

54. The lifeguard from a small village in Akron, Pennsylvania dramatically died when the tornado destroyed the entire neighborhood.

55. The fireman from a small village in Akron, Pennsylvania dramatically yelled when the tornado destroyed the entire neighborhood.

56. The lifeguard from a small village in Akron, Pennsylvania dramatically yelled when the tornado destroyed the entire neighborhood.

57. The geographer with a loud voice and quick temper suddenly fell when the boat lurched violently because of the storm.

58. The conductor with a loud voice and quick temper suddenly fell when the boat lurched violently because of the storm.

59. The geographer with a loud voice and quick temper suddenly screamed when the boat lurched violently because of the storm.

60. The conductor with a loud voice and quick temper suddenly screamed when the boat lurched violently because of the storm.

61. The dentist from the poorest neighborhood in Brooklyn, New York dramatically tumbled down the stairs when the party was about to end.

62. The gardener from the poorest neighborhood in Brooklyn, New York dramatically tumbled down the stairs when the party was about to end.

63. The dentist from the poorest neighborhood in Brooklyn, New York dramatically jumped down the stairs when the party was about to end.
64. The gardener from the poorest neighborhood in Brooklyn, New York dramatically jumped down the stairs when the party was about to end.

65. The businessman from a small town in Massachusetts unexpectedly returned before the financial crisis was resolved.

66. The electrician from a small town in Massachusetts unexpectedly returned before the financial crisis was resolved.

67. The businessman from a small town in Massachusetts unexpectedly resigned before the financial crisis was resolved.

68. The electrician from a small town in Massachusetts unexpectedly resigned before the financial crisis was resolved.

69. The mailman from a rural town in Montana surprisingly left before the winner of the prize was announced.

70. The janitor from a rural town in Montana surprisingly left before the winner of the prize was announced.

71. The mailman from a rural town in Montana surprisingly bragged before the winner of the prize was announced.

72. The janitor from a rural town in Montana surprisingly bragged before the winner of the prize was announced.

73. The painter from the most conservative part of Maine continuously stood while the city councilor convened the meeting.

74. The plumber from the most conservative part of Maine continuously stood while the city councilor convened the meeting.

75. The painter from the most conservative part of Maine continuously spoke while the city councilor convened the meeting.

76. The plumber from the most conservative part of Maine continuously spoke while the city councilor convened the meeting.

77. The carpenter from the Northern part of Hamburg, Germany, suddenly arose when the choir began to sing the memorial hymn.

78. The hunter from the Northern part of Hamburg, Germany, suddenly arose when the choir began to sing the memorial hymn.

79. The carpenter from the Northern part of Hamburg, Germany, suddenly paused when the choir began to sing the memorial hymn.
80. The hunter from the Northern part of Hamburg, Germany, suddenly paused when the choir began to sing the memorial hymn.

81. The clown with curly red hair and freckles suddenly arrived when the famous actress started the performance by greeting the audience.

82. The cook with curly red hair and freckles suddenly arrived when the famous actress started the performance by greeting the audience.

83. The clown with curly red hair and freckles suddenly sighed when the famous actress started the performance by greeting the audience.

84. The cook with curly red hair and freckles suddenly sighed when the famous actress started the performance by greeting the audience.

85. The baker with the conspicuous comb-over and wrinkled face noticeably lurked around the house before the burglar alarm suddenly went off.

86. The sailor with the conspicuous comb-over and wrinkled face noticeably lurked around the house before the burglar alarm suddenly went off.

87. The baker with the conspicuous comb-over and wrinkled face noticeably snuck around the house before the burglar alarm suddenly went off.

88. The sailor with the conspicuous comb-over and wrinkled face noticeably snuck around the house before the burglar alarm suddenly went off.

89. The librarian on a voluntary six-month hiatus from work dramatically awoke after the storm caused a tree to smash through the window.

90. The astronomer on a voluntary six-month hiatus from work dramatically awoke after the storm caused a tree to smash through the window.

91. The librarian on a voluntary six-month hiatus from work dramatically shouted after the storm caused a tree to smash through the window.

92. The astronomer on a voluntary six-month hiatus from work dramatically shouted after the storm caused a tree to smash through the window.

93. The athlete from a war-torn region of the middle east temporarily resided in the U.S. while the economy slowly recovered.

94. The singer from a war-torn region of the middle east temporarily resided in the U.S. while the economy slowly recovered.

95. The athlete from a war-torn region of the middle east temporarily worked in the U.S. while the economy slowly recovered.
96. The singer from a war-torn region of the middle east temporarily worked in the U.S. while the economy slowly recovered.

97. The poet with a strong interest in Brazilian pop music suddenly emerged from the cathedral after the loud cannon went off in the yard.

98. The tourist with a strong interest in Brazilian pop music suddenly emerged from the cathedral after the loud cannon went off in the yard.

99. The poet with a strong interest in Brazilian pop music suddenly rushed from the cathedral after the loud cannon went off in the yard.

100. The tourist with a strong interest in Brazilian pop music suddenly rushed from the cathedral after the loud cannon went off in the yard.

101. The architect from a middle-class neighborhood in Los Angeles surprisingly disappeared when the keynote speech was supposed to start.

102. The nurse from a middle-class neighborhood in Los Angeles surprisingly disappeared when the keynote speech was supposed to start.

103. The architect from a middle-class neighborhood in Los Angeles surprisingly hesitated when the keynote speech was supposed to start.

104. The nurse from a middle-class neighborhood in Los Angeles surprisingly hesitated when the keynote speech was supposed to start.

105. The captain with a mosquito bite on his left forearm unobtrusively rose before the singer began a rendition of the national anthem.

106. The judge with a mosquito bite on his left forearm unobtrusively rose before the singer began a rendition of the national anthem.

107. The captain with a mosquito bite on his left forearm unobtrusively whispered before the singer began a rendition of the national anthem.

108. The judge with a mosquito bite on his left forearm unobtrusively whispered before the singer began a rendition of the national anthem.

109. The farmer from a rural village in Central America unnecessarily remained after the party was canceled due to bad weather.

110. The fisherman from a rural village in Central America unnecessarily remained after the party was canceled due to bad weather.

111. The farmer from a rural village in Central America unnecessarily cried after the party was canceled due to bad weather.
112. The fisherman from a rural village in Central America unnecessarily cried after the party was canceled due to bad weather.

113. The banker with a brown coat and brown shoes apparently stayed after the police officers broke up the protest.

114. The lumberjack with a brown coat and brown shoes apparently stayed after the police officers broke up the protest.

115. The banker with a brown coat and brown shoes apparently complained after the police officers broke up the protest.

116. The lumberjack with a brown coat and brown shoes apparently complained after the police officers broke up the protest.

117. The astronaut with big ears and an asymmetrical haircut conspicuously appeared as the professor began to discuss French literature.

118. The driver with big ears and an asymmetrical haircut conspicuously appeared as the professor began to discuss French literature.

119. The astronaut with big ears and an asymmetrical haircut conspicuously slept while the professor began to discuss French literature.

120. The driver with big ears and an asymmetrical haircut conspicuously slept while the professor began to discuss French literature.

121. The magician in a thick winter jacket with a fur-lined hood dramatically vanished when the politicians announced their working plan.

122. The musician in a thick winter jacket with a fur-lined hood dramatically vanished when the politicians announced their working plan.

123. The magician in a thick winter jacket with a fur-lined hood dramatically whined when the politicians announced their working plan.

124. The musician in a thick winter jacket with a fur-lined hood dramatically whined when the politicians announced their working plan.

125. The biologist from a prestigious family in Eastern Europe immediately departed after the opposing team scored the winning goal.

126. The photographer from a prestigious family in Eastern Europe immediately departed after the opposing team scored the winning goal.

127. The biologist from a prestigious family in Eastern Europe immediately grumbled after the opposing team scored the winning goal.
128. The photographer from a prestigious family in Eastern Europe immediately grumbled after the opposing team scored the winning goal.

129. The butcher with the shiny watch and designer shoes immediately collapsed when the fastidious superintendent left the room.

130. The hairdresser with the shiny watch and designer shoes immediately collapsed when the fastidious superintendent left the room.

131. The butcher with the shiny watch and designer shoes immediately swore when the fastidious superintendent left the room.

132. The hairdresser with the shiny watch and designer shoes immediately swore when the fastidious superintendent left the room.

133. The fireman from a very remote part of Appalachia dramatically died after the furious rhinos trampled the caravan.

134. The lifeguard from a very remote part of Appalachia dramatically died after the furious rhinos trampled the caravan.

135. The fireman from a very remote part of Appalachia dramatically yelled after the furious rhinos trampled the caravan.

136. The lifeguard from a very remote part of Appalachia dramatically yelled after the furious rhinos trampled the caravan.

137. The geographer with a surplus of paid vacation days immediately fell when the train to Andermatt began to shake severely.

138. The conductor with a surplus of paid vacation days immediately fell when the train to Andermatt began to shake severely.

139. The geographer with a surplus of paid vacation days immediately screamed when the train to Andermatt began to shake severely.

140. The conductor with a surplus of paid vacation days immediately screamed when the train to Andermatt began to shake severely.

141. The dentist with the expensive silver convertible unexpectedly tumbled when the ferry hit a rough patch of water and severely lurched.

142. The gardener with the expensive silver convertible unexpectedly tumbled when the ferry hit a rough patch of water and severely lurched.

143. The dentist with the expensive silver convertible unexpectedly jumped when the ferry hit a rough patch of water and severely lurched.
144. The gardener with the expensive silver convertible unexpectedly jumped when the ferry hit a rough patch of water and severely lurched.

145. The businessman in a dark turtleneck and worn-out khaki pants unexpectedly returned after the political climate finally improved.

146. The electrician in a dark turtleneck and worn-out khaki pants unexpectedly returned after the political climate finally improved.

147. The businessman in a dark turtleneck and worn-out khaki pants unexpectedly resigned after the political climate finally improved.

148. The electrician in a dark turtleneck and worn-out khaki pants unexpectedly resigned after the political climate finally improved.

149. The mailman from a small liberal arts college in New England unexpectedly left when the conversation turned to famous alumni.

150. The janitor from a small liberal arts college in New England unexpectedly left when the conversation turned to famous alumni.

151. The mailman from a small liberal arts college in New England unexpectedly bragged when the conversation turned to famous alumni.

152. The janitor from a small liberal arts college in New England unexpectedly bragged when the conversation turned to famous alumni.

153. The painter with a potbelly and a thick Southern accent distractingly stood as the lights in the conference room gradually dimmed.

154. The plumber with a potbelly and a thick Southern accent distractingly stood as the lights in the conference room gradually dimmed.

155. The painter with a potbelly and a thick Southern accent distractingly spoke as the lights in the conference room gradually dimmed.

156. The plumber with a potbelly and a thick Southern accent distractingly spoke as the lights in the conference room gradually dimmed.

157. The carpenter in the rented tuxedo with a tear in the sleeve unexpectedly arose after the plan was announced to the public.

158. The hunter in the rented tuxedo with a tear in the sleeve unexpectedly arose after the plan was announced to the public.

159. The carpenter in the rented tuxedo with a tear in the sleeve unexpectedly paused after the plan was announced to the public.
160. The hunter in the rented tuxedo with a tear in the sleeve unexpectedly paused after the plan was announced to the public.

**Fillers**

161. The pirate with a large beard and an eyepatch suddenly found the treasure when the corsair entered the dark narrow cave.

162. The woman with a high-pitched voice unsurprisingly annoyed the guests after the party went on for an hour.

163. The soldier in a washed-out military uniform happily hugged the signaler after the victory was announced.

164. The baby with a chubby face and adorable dimples almost touched the fireplace when the nanny was cleaning the kitchen.

165. The queen with a reputation of being uncooperative surprisingly signed the document after the cabinet gave up their hope.

166. The boy with a Boston Red Sox hat heroically saved the old lady before the building was burnt down.

167. The bride with a pleasant smile warmly welcomed all the guests before the wedding ceremony started.

168. The aristocrat in a light pink Victorian lace dress reluctantly greeted the knight after the duke introduced him as a hero.

169. The king with an authoritative look gladly praised the general after the allies defeated their ruthless enemy.

170. The band with tremendous teamwork problems finally released a new album after the last well-received one came out ten years ago.

171. The waiter with great interpersonal skills successfully handled the angry customer after the chef made a silly mistake.

172. The cowboy with a leather hat and brown boots proudly amazed the visitors after showing all the horseback riding skills.

173. The policeman with many legendary stories immediately identified the robber after the bank staff described his characteristics.

174. The priest with a calming voice immediately comforted the local people after the hurricane destroyed one third of the county.
175. The dancer with graceful manners gently dismissed the reporters when the award ceremony resumed after the break.

176. The Santa Claus with a giant bag on the back dotingly patted the kid after passing him the Christmas gift.

177. The motorcyclist with a shiny helmet and trendy sunglasses successfully repaired the engine before the national competition took place.

178. The wrestler with a tough personality unfortunately lost the game after all the cheers and encouragement from the audience.

179. The witch with an oak magic wand readily helped the residents after the drought caused the famine.

180. The bird with a delicate yellow mouth and beautiful feathers caught the worm when the first beam of morning light penetrated the forest.

181. The lion with golden thick hair authoritatively expelled the invader after the lionesses sensed the danger and began to roar.

182. The squirrel with a round body and a fluffy tail diligently buried the chestnuts before the long and fierce winter came.

183. The wolf with gorgeous white fur courageously protected the puppies when the hyenas attacked the pack.

184. The monkey with a funny looking face ferociously bit the tourists when the baby began to cry.

185. The parrot with no experience of mimicking quickly aroused the customer's interest when he entered the shop.

186. The beaver with tiny fingers and a big nose rapidly built the dam before the predators entered the area.

187. The tiger with beautiful stripes on the face persistently chased the rabbit before it eventually managed to run away.

188. The shark with frightening sharp teeth quickly devoured the huge cod when it carelessly swam by.

189. The peacock with a dazzling green train proudly defeated the competitors when the female was choosing her mate.

190. The camel with impressively huge humps quickly drank a little water when the travelers entered the oasis after a long journey.
191. The elephant with an abnormally long nose successfully reached the banana while the other animals' attempts failed.

192. The hedgehog with cute little ears eagerly collected the berries when the grassland was covered with them.

193. The goose with brown feathers and big wings hurriedly fed the gosling as it was crying out in hunger.

194. The mouse with little pink ears and small eyes bravely challenged the cat when it pounced on the cheese.

195. The moose with the prettiest antlers fortuitously discovered a new meadow when the old inhabitant ran out of food.

196. The panda with a big face and small paws clumsily threw the ball when the trainer gave the signal.

197. The penguin with a bright orange ring around the neck unexpectedly ignored the fish after the zoo-keeper refilled the food tray.

198. The deer with many little white dots on the back licked the wound as the hunter was making his way with a gun.

199. The turkey with a red neck and bald head threatened the fox when it tried to steal the baby.

200. The spider with long legs and a narrow body immediately killed the fly when it was captured by the net.
## Appendix 2: English Eye-Tracking Study Verb Information

<table>
<thead>
<tr>
<th>Verb</th>
<th>Log frequency</th>
<th>Imageability</th>
<th>Verb</th>
<th>Log frequency</th>
<th>Imageability</th>
</tr>
</thead>
<tbody>
<tr>
<td>to brag</td>
<td>1.63</td>
<td>3.16</td>
<td>to appear</td>
<td>5.48</td>
<td>3.55</td>
</tr>
<tr>
<td>to complain</td>
<td>3.82</td>
<td>3.71</td>
<td>to lurk</td>
<td>1.85</td>
<td>3.65</td>
</tr>
<tr>
<td>to cry</td>
<td>4.26</td>
<td>5.40</td>
<td>to arise</td>
<td>3.46</td>
<td>3.35</td>
</tr>
<tr>
<td>to grumble</td>
<td>1.58</td>
<td>3.23</td>
<td>to arrive</td>
<td>4.74</td>
<td>3.87</td>
</tr>
<tr>
<td>to hesitate</td>
<td>2.88</td>
<td>3.29</td>
<td>to awake</td>
<td>1.86</td>
<td>4.55</td>
</tr>
<tr>
<td>to jump</td>
<td>4.23</td>
<td>5.61</td>
<td>to collapse</td>
<td>3.10</td>
<td>4.30</td>
</tr>
<tr>
<td>to pause</td>
<td>3.42</td>
<td>3.23</td>
<td>to depart</td>
<td>2.49</td>
<td>3.48</td>
</tr>
<tr>
<td>to resign</td>
<td>3.05</td>
<td>2.84</td>
<td>to die</td>
<td>5.48</td>
<td>4.77</td>
</tr>
<tr>
<td>to rush</td>
<td>3.81</td>
<td>3.90</td>
<td>to disappear</td>
<td>4.03</td>
<td>3.39</td>
</tr>
<tr>
<td>to scream</td>
<td>3.73</td>
<td>4.74</td>
<td>to emerge</td>
<td>4.26</td>
<td>3.58</td>
</tr>
<tr>
<td>to shout</td>
<td>3.72</td>
<td>4.81</td>
<td>to fall</td>
<td>5.45</td>
<td>5.26</td>
</tr>
<tr>
<td>to sleep</td>
<td>4.43</td>
<td>5.39</td>
<td>to leave</td>
<td>6.36</td>
<td>3.94</td>
</tr>
<tr>
<td>to sigh</td>
<td>3.19</td>
<td>4.00</td>
<td>to remain</td>
<td>5.39</td>
<td>2.52</td>
</tr>
<tr>
<td>to speak</td>
<td>5.65</td>
<td>4.48</td>
<td>to reside</td>
<td>2.41</td>
<td>2.94</td>
</tr>
<tr>
<td>to swear</td>
<td>3.11</td>
<td>3.35</td>
<td>to return</td>
<td>5.29</td>
<td>3.10</td>
</tr>
<tr>
<td>to sneak</td>
<td>2.62</td>
<td>3.71</td>
<td>to rise</td>
<td>4.87</td>
<td>4.00</td>
</tr>
<tr>
<td>to whine</td>
<td>1.70</td>
<td>3.87</td>
<td>to stand</td>
<td>5.82</td>
<td>5.35</td>
</tr>
<tr>
<td>to whisper</td>
<td>3.61</td>
<td>4.35</td>
<td>to stay</td>
<td>5.46</td>
<td>3.26</td>
</tr>
<tr>
<td>to work</td>
<td>6.65</td>
<td>4.81</td>
<td>to tumble</td>
<td>2.34</td>
<td>4.39</td>
</tr>
<tr>
<td>to yell</td>
<td>3.45</td>
<td>4.81</td>
<td>to vanish</td>
<td>2.85</td>
<td>3.47</td>
</tr>
</tbody>
</table>
## Appendix 3: Imageability of the Friedman et al. Verbs

<table>
<thead>
<tr>
<th>Verb</th>
<th>Imageability</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>to appear</td>
<td>4.07</td>
<td>UA</td>
</tr>
<tr>
<td>to arise</td>
<td>3.93</td>
<td>UA</td>
</tr>
<tr>
<td>to arrive</td>
<td>4.11</td>
<td>UA</td>
</tr>
<tr>
<td>to bark</td>
<td>5.86</td>
<td>UE</td>
</tr>
<tr>
<td>to bloom</td>
<td>4.93</td>
<td>UA</td>
</tr>
<tr>
<td>to crawl</td>
<td>5.66</td>
<td>UE</td>
</tr>
<tr>
<td>to cry</td>
<td>6.07</td>
<td>UE</td>
</tr>
<tr>
<td>to depart</td>
<td>4.24</td>
<td>UA</td>
</tr>
<tr>
<td>to disappear</td>
<td>4.16</td>
<td>UA</td>
</tr>
<tr>
<td>to emerge</td>
<td>4.07</td>
<td>UA</td>
</tr>
<tr>
<td>to escape</td>
<td>4.00</td>
<td>UE</td>
</tr>
<tr>
<td>to exist</td>
<td>3.07</td>
<td>UA</td>
</tr>
<tr>
<td>to fall</td>
<td>5.72</td>
<td>UA</td>
</tr>
<tr>
<td>to flower</td>
<td>5.28</td>
<td>UA</td>
</tr>
<tr>
<td>to hesitate</td>
<td>3.38</td>
<td>UE</td>
</tr>
<tr>
<td>to jump</td>
<td>5.86</td>
<td>UE</td>
</tr>
<tr>
<td>to laugh</td>
<td>6.16</td>
<td>UE</td>
</tr>
<tr>
<td>to occur</td>
<td>2.28</td>
<td>UA</td>
</tr>
<tr>
<td>to remain</td>
<td>3.21</td>
<td>UA</td>
</tr>
<tr>
<td>to rise</td>
<td>5.03</td>
<td>UA</td>
</tr>
<tr>
<td>to scream</td>
<td>5.90</td>
<td>UE</td>
</tr>
<tr>
<td>to shout</td>
<td>5.62</td>
<td>UE</td>
</tr>
<tr>
<td>to sing</td>
<td>5.86</td>
<td>UE</td>
</tr>
<tr>
<td>to sleep</td>
<td>6.24</td>
<td>UE</td>
</tr>
<tr>
<td>to smile</td>
<td>6.43</td>
<td>UE</td>
</tr>
<tr>
<td>to stood</td>
<td>3.21</td>
<td>UA</td>
</tr>
<tr>
<td>to tremble</td>
<td>5.28</td>
<td>UE</td>
</tr>
<tr>
<td>to vanish</td>
<td>3.78</td>
<td>UA</td>
</tr>
<tr>
<td>to wave</td>
<td>6.07</td>
<td>UE</td>
</tr>
<tr>
<td>to wink</td>
<td>5.90</td>
<td>UE</td>
</tr>
</tbody>
</table>
Appendix 4: Stimuli for the Dutch Visual World Paradigm Study

Dutch sentence

1. Marie zei dat de architect met een gele hoed en een blauwe paraplu meteen wegsloop toen de blonde serveerster de hoge rekening bracht.

2. Marie zei dat de detective met een gele hoed en een blauwe paraplu meteen wegsloop toen de blonde serveerster de hoge rekening bracht.

3. Marie zei dat de architect uit een kleine gemeente in Limburg opeens wegsloop toen de burgemeester van Maastricht met zijn speech begon.

4. Marie zei dat de detective uit een kleine gemeente in Limburg opeens wegsloop toen de burgemeester van Maastricht met zijn speech begon.

5. Marie zei dat de architect met een gele hoed en een blauwe paraplu meteen aarzelde toen de blonde serveerster de hoge rekening bracht.

6. Marie zei dat de detective met een gele hoed en een blauwe paraplu meteen aarzelde toen de blonde serveerster de hoge rekening bracht.

7. Marie zei dat de architect uit een kleine gemeente in Limburg opeens aarzelde toen de burgemeester van Maastricht met zijn speech begon.

8. Marie zei dat de detective uit een kleine gemeente in Limburg opeens aarzelde toen de burgemeester van Maastricht met zijn speech begon.

9. Jan zei dat de brandweerman uit Duitsland in het midden van de nacht aankwam nadat het hele dorp bijeen was gekomen voor het feest.

10. Jan zei dat de badmeester uit Duitsland in het midden van de nacht aankwam nadat het hele dorp bijeen was gekomen voor het feest.

11. Jan zei dat de brandweerman met een opvallend kapsel en een rode bril opeens aankwam toen het alarm plotseling erg luid afging.

12. Jan zei dat de badmeester met een opvallend kapsel en een rode bril opeens aankwam toen het alarm plotseling erg luid afging.

13. Jan zei dat de brandweerman uit Duitsland in het midden van de nacht mompelde nadat het hele dorp bijeen was gekomen voor het feest.

14. Jan zei dat de badmeester uit Duitsland in het midden van de nacht mompelde nadat het hele dorp bijeen was gekomen voor het feest.
15. Jan zei dat de brandweerman met een opvallend kapsel en een rode bril opeens mompelde toen het alarm plotseling erg luid afging.


17. Marie zei dat de clown uit een klein afgelegen bergachtig gebied 's middags verscheen toen de koningin haar speech tijdens het diner gaf.

18. Marie zei dat de kok uit een klein afgelegen bergachtig gebied 's middags verscheen toen de koningin haar speech tijdens het diner gaf.

19. Marie zei dat de clown met grote oren en een asymmetrisch kapsel 's morgens verscheen toen de professor een lezing over bijen gaf.

20. Marie zei dat de kok met grote oren en een asymmetrisch kapsel 's morgens verscheen toen de professor een lezing over bijen gaf.

21. Marie zei dat de clown uit een klein afgelegen bergachtig gebied 's middags sliep toen de koningin haar speech tijdens het diner gaf.

22. Marie zei dat de kok uit een klein afgelegen bergachtig gebied 's middags sliep toen de koningin haar speech tijdens het diner gaf.

23. Marie zei dat de clown met grote oren en een asymmetrisch kapsel 's morgens sliep toen de professor een lezing over bijen gaf.

24. Marie zei dat de kok met grote oren en een asymmetrisch kapsel 's morgens sliep toen de professor een lezing over bijen gaf.

25. Jan zei dat de goochelaar met interesse in wiskundige theorieën meteen verdween toen de filosoof kritiek op de systeemtheorie uitte.

26. Jan zei dat de dirigent met interesse in wiskundige theorieën meteen verdween toen de filosoof kritiek op de systeemtheorie uitte.

27. Jan zei dat de goochelaar met een zwarte jas met een rode kraag meteen verdween nadat de Tweede Kamer had gedebateerd.

28. Jan zei dat de dirigent met een zwarte jas met een rode kraag meteen verdween nadat de Tweede Kamer had gedebateerd.

29. Jan zei dat de goochelaar met interesse in wiskundige theorieën meteen zeurde toen de filosoof kritiek op de systeemtheorie uitte.

30. Jan zei dat de dirigent met interesse in wiskundige theorieën meteen zeurde toen de filosoof kritiek op de systeemtheorie uitte.
31. Jan zei dat de goochelaar met een zwarte jas met een rode kraag meteen zeurde nadat de Tweede Kamer had gedebateerd.

32. Jan zei dat de dirigent met een zwarte jas met een rode kraag meteen zeurde nadat de Tweede Kamer had gedebateerd.

33. Jan zei dat de piloot uit een oosterlijke provincie in Frankrijk luidruchtig arriveerde toen de film over de historie van geld begon.

34. Jan zei dat de muzikant uit een oosterlijke provincie in Frankrijk luidruchtig arriveerde toen de film over de historie van geld begon.

35. Jan zei dat de piloot met een rode jas en een witte broek onverwachts arriveerde toen de baas plotseling het kantoor verliet.

36. Jan zei dat de muzikant met een rode jas en een witte broek onverwachts arriveerde toen de baas plotseling het kantoor verliet.

37. Jan zei dat de piloot uit een oosterlijke provincie in Frankrijk luidruchtig zuchtte toen de film over de historie van geld begon.

38. Jan zei dat de muzikant uit een oosterlijke provincie in Frankrijk luidruchtig zuchtte toen de film over de historie van geld begon.

39. Jan zei dat de piloot met een rode jas en een witte broek onverwachts zuchtte toen de baas plotseling het kantoor verliet.

40. Jan zei dat de muzikant met een rode jas en een witte broek onverwachts zuchtte toen de baas plotseling het kantoor verliet.

41. Marie zei dat de zakenman met een zwarte trui en een versleten broek onverwachts terugkwam nadat de politieke onrust enorm verbeterd was.

42. Marie zei dat de electricien met een zwarte trui en een versleten broek onverwachts terugkwam nadat de politieke onrust enorm verbeterd was.

43. Marie zei dat de zakenman uit een klein dorp in Zuid-Holland meteen terugkwam toen de vrouw begon te huilen tijdens de voorstelling.

44. Marie zei dat de electricien uit een klein dorp in Zuid-Holland meteen terugkwam toen de vrouw begon te huilen tijdens de voorstelling.

45. Marie zei dat de zakenman met een zwarte trui en een versleten broek onverwachts klaagde nadat de politieke onrust enorm verbeterd was.

46. Marie zei dat de electricien met een zwarte trui en een versleten broek onverwachts klaagde nadat de politieke onrust enorm verbeterd was.
47. Marie zei dat de zakenman uit een klein dorp in Zuid-Holland meteen klaagde toen de vrouw begon te huilen tijdens de voorstelling.

48. Marie zei dat de electricien uit een klein dorp in Zuid-Holland meteen klaagde toen de vrouw begon te huilen tijdens de voorstelling.

49. Jan zei dat de dichter met een interesse in Britse popmuziek plotseling opdook toen er een kanonkogel werd afgeschoten op het veld.

50. Jan zei dat de toerist met een interesse in Britse popmuziek plotseling opdook toen er een kanonkogel werd afgeschoten op het veld.

51. Jan zei dat de dichter uit het noorden van de Franse hoofdstad opeens opdook toen de fanfare in de kleine hoofdstraat voorbij kwam.

52. Jan zei dat de toerist uit het noorden van de Franse hoofdstad opeens opdook toen de fanfare in de kleine hoofdstraat voorbij kwam.

53. Jan zei dat de dichter met een interesse in Britse popmuziek plotseling zwaide toen er een kanonkogel werd afgeschoten op het veld.

54. Jan zei dat de toerist met een interesse in Britse popmuziek plotseling zwaide toen er een kanonkogel werd afgeschoten op het veld.

55. Jan zei dat de dichter uit het noorden van de Franse hoofdstad opeens zwaide toen de fanfare in de kleine hoofdstraat voorbij kwam.

56. Jan zei dat de toerist uit het noorden van de Franse hoofdstad opeens zwaide toen de fanfare in de kleine hoofdstraat voorbij kwam.

57. Marie zei dat de boer met een geruit overhemd en een pet gisteren verdwaalde toen de voorzitter van de raad het buurthuis opende.

58. Marie zei dat de dokter met een geruit overhemd en een pet gisteren verdwaalde toen de voorzitter van de raad het buurthuis opende.

59. Marie zei dat de boer uit een groot gezin met tien kinderen 's nachts verdwaalde toen de klok van de oude kerk het middernachtsuur sloeg.

60. Marie zei dat de dokter uit een groot gezin met tien kinderen 's nachts verdwaalde toen de klok van de oude kerk het middernachtsuur sloeg.

61. Marie zei dat de boer met een geruit overhemd en een pet gisteren gluurde toen de voorzitter van de raad het buurthuis opende.

62. Marie zei dat de dokter met een geruit overhemd en een pet gisteren gluurde toen de voorzitter van de raad het buurthuis opende.
63. Marie zei dat de boer uit een groot gezin met tien kinderen 's nachts gluurde toen de klok van de oude kerk het middernachtsuur sloeg.

64. Marie zei dat de dokter uit een groot gezin met tien kinderen 's nachts gluurde toen de klok van de oude kerk het middernachtsuur sloeg.

65. Jan zei dat de jager uit een klein dorp in het noorden van Friesland opeens vertrok voordat de winnaar van de wedstrijd bekend werd gemaakt.

66. Jan zei dat de timmerman uit een klein dorp in het noorden van Friesland opeens vertrok voordat de winnaar van de wedstrijd bekend werd gemaakt.

67. Jan zei dat de jager met een grote passie voor zwart-wit foto's plotseling vertrok toen de jury de foto's van de wedstrijd moest beoordelen.

68. Jan zei dat de timmerman met een grote passie voor zwart-wit foto's plotseling vertrok toen de jury de foto's van de wedstrijd moest beoordelen.

69. Jan zei dat de jager uit een klein dorp in het noorden van Friesland opeens opschepte voordat de winnaar van de wedstrijd bekend werd gemaakt.

70. Jan zei dat de timmerman uit een klein dorp in het noorden van Friesland opeens opschepte voordat de winnaar van de wedstrijd bekend werd gemaakt.

71. Jan zei dat de jager met een grote passie voor zwart-wit foto's plotseling opschepte toen de jury de foto's van de wedstrijd moest beoordelen.

72. Jan zei dat de timmerman met een grote passie voor zwart-wit foto's plotseling opschepte toen de jury de foto's van de wedstrijd moest beoordelen.

73. Jan zei dat de bakker uit een zeer kleine stad in Zwitserland dramatisch stierf toen de tornado de volledige woonwijk verwoestte.

74. Jan zei dat de zeeman uit een zeer kleine stad in Zwitserland dramatisch stierf toen de tornado de volledige woonwijk verwoestte.

75. Jan zei dat de bakker met een opvallende baard en krullen onverwachts stierf nadat de tijger de caravan in het park vernielde.

76. Jan zei dat de zeeman met een opvallende baard en krullen onverwachts stierf nadat de tijger de caravan in het park vernielde.

77. Jan zei dat de bakker uit een zeer kleine stad in Zwitserland dramatisch gilde toen de tornado de volledige woonwijk verwoestte.

78. Jan zei dat de zeeman uit een zeer kleine stad in Zwitserland dramatisch gilde toen de tornado de volledige woonwijk verwoestte.
79. Jan zei dat de bakker met een opvallende baard en krullen onverwachts gilde nadat de tijger de caravan in het park vernielde.

80. Jan zei dat de zeeman met een opvallende baard en krullen onverwachts gilde nadat de tijger de caravan in het park vernielde.

81. Marie zei dat de kapitein met een lange jas en een blauwe sjaal plotseling opstond toen de nieuwe minister van Financiën de zaal in liep.

82. Marie zei dat de rechter met een lange jas en een blauwe sjaal plotseling opstond toen de nieuwe minister van Financiën de zaal in liep.

83. Marie zei dat de kapitein met een dikke muggenbult op zijn arm opeens opstond toen de zanger het Nederlandse volkslied begon te zingen.

84. Marie zei dat de rechter met een dikke muggenbult op zijn arm opeens opstond toen de zanger het Nederlandse volkslied begon te zingen.

85. Marie zei dat de kapitein met een lange jas en een blauwe sjaal plotseling fluisterde toen de nieuwe minister van Financiën de zaal in liep.

86. Marie zei dat de rechter met een lange jas plotseling fluisterde toen de nieuwe minister van Financiën de zaal in liep.

87. Marie zei dat de kapitein met een dikke muggenbult op zijn arm opeens fluisterde toen de zanger het Nederlandse volkslied begon te zingen.

88. Marie zei dat de rechter met een dikke muggenbult op zijn arm opeens fluisterde toen de zanger het Nederlandse volkslied begon te zingen.

89. Jan zei dat de slager met een grote passie voor voetbal onverwachts schrok toen de tegenpartij in de laatste minuut scoorde.

90. Jan zei dat de kapper met een grote passie voor voetbal onverwachts schrok toen de tegenpartij in de laatste minuut scoorde.

91. Jan zei dat de slager met een horloge en dure schoenen onmiddelijk schrok toen de baas met een boze blik de winkel betrad.

92. Jan zei dat de kapper met een horloge en dure schoenen onmiddelijk schrok toen de baas met een boze blik de winkel betrad.

93. Jan zei dat de slager met een grote passie voor voetbal onverwachts vloekte toen de tegenpartij in de laatste minuut scoorde.

94. Jan zei dat de kapper met een grote passie voor voetbal onverwachts vloekte toen de tegenpartij in de laatste minuut scoorde.
105. Jan zei dat de bioloog met een dure zwarte sportauto meteen wegging toen het natuurpark bekend maakte eerder te gaan sluiten.

106. Jan zei dat de apotheker met een dure zwarte sportauto meteen wegging toen het natuurpark bekend maakte eerder te gaan sluiten.

107. Jan zei dat de bioloog uit een rijke Europese familie meteen wegging nadat de tegenpartij het winnende doelpunt scoorde.

108. Jan zei dat de apotheker uit een rijke Europese familie meteen wegging nadat de tegenpartij het winnende doelpunt scoorde.

109. Jan zei dat de bioloog met een dure zwarte sportauto meteen mopperde toen het natuurpark bekend maakte eerder te gaan sluiten.

110. Jan zei dat de apotheker met een dure zwarte sportauto meteen mopperde toen het natuurpark bekend maakte eerder te gaan sluiten.
111. Jan zei dat de bioloog uit een rijke Europese familie meteen mopperde nadat de tegenpartij het winnende doelpunt scoorde.
112. Jan zei dat de apotheker uit een rijke Europese familie meteen mopperde nadat de tegenpartij het winnende doelpunt scoorde.
113. Marie zei dat de atleet met een luid stem en een strenge blik meteen viel toen het treinstel heen en weer schommelde vanwege de storm.
114. Marie zei dat de zanger met een luid stem en een strenge blik meteen viel toen het treinstel heen en weer schommelde vanwege de storm.
115. Marie zei dat de atleet met een passie voor postmoderne kunst opeens viel toen de directeur de kamer luidruchtig binnen kwam.
116. Marie zei dat de zanger met een passie voor postmoderne kunst opeens viel toen de directeur de kamer luidruchtig binnen kwam.
117. Marie zei dat de atleet met een luid stem en een strenge blik meteen schreeuwde toen het treinstel heen en weer schommelde vanwege de storm.
118. Marie zei dat de zanger met een luid stem en een strenge blik meteen schreeuwde toen het treinstel heen en weer schommelde vanwege de storm.
119. Marie zei dat de atleet met een passie voor postmoderne kunst opeens schreeuwde toen de directeur de kamer luidruchtig binnen kwam.
120. Marie zei dat de zanger met een passie voor postmoderne kunst opeens schreeuwde toen de directeur de kamer luidruchtig binnen kwam.
121. Jan zei dat de bibliothecaris uit een arme buurt in Amsterdam 's avonds struikelde toen het feestje bijna was afgelopen.
122. Jan zei dat de politieman uit een arme buurt in Amsterdam 's avonds struikelde toen het feestje bijna was afgelopen.
123. Jan zei dat de bibliothecaris met een duur zilveren horloge 's middags struikelde toen de boot hevig schommelde door het noodweer op zee.
124. Jan zei dat de politieman met een duur zilveren horloge 's middags struikelde toen de boot hevig schommelde door het noodweer op zee.
125. Jan zei dat de bibliothecaris uit een arme buurt in Amsterdam 's avonds ronddoolde toen het feestje bijna was afgelopen.
126. Jan zei dat de politieman uit een arme buurt in Amsterdam 's avonds ronddoolde toen het feestje bijna was afgelopen.
127. Jan zei dat de bibliothecaris met een duur zilveren horloge ’s middags ronddoolde toen de boot hevig schommelde door het noodweer op zee.

128. Jan zei dat de politieman met een duur zilveren horloge ’s middags ronddoolde toen de boot hevig schommelde door het noodweer op zee.

129. Marie zei dat de duiker uit een grote conservatieve gemeente meteen vluchtte toen bekend werd dat de hoorzitting nog lang zou duren.

130. Marie zei dat de tuinman uit een grote conservatieve gemeente meteen vluchtte toen bekend werd dat de hoorzitting nog lang zou duren.

131. Marie zei dat de duiker met een dikke buik en Amsterdams accent opeens vluchtte toen de lichten in de vergaderkamer werden gedimd.

132. Marie zei dat de tuinman met een dikke buik en Amsterdams accent opeens vluchtte toen de lichten in de vergaderkamer werden gedimd.

133. Marie zei dat de duiker uit een grote conservatieve gemeente meteen gaapte toen bekend werd dat de hoorzitting nog lang zou duren.

134. Marie zei dat de tuinman uit een grote conservatieve gemeente meteen gaapte toen bekend werd dat de hoorzitting nog lang zou duren.

135. Marie zei dat de duiker met een dikke buik en Amsterdams accent opeens gaapte toen de lichten in de vergaderkamer werden gedimd.

136. Marie zei dat de tuinman met een dikke buik en Amsterdams accent opeens gaapte toen de lichten in de vergaderkamer werden gedimd.

137. Jan zei dat de chauffeur uit een kleine provincie in Duitsland hard vooruit ging tijdens het schaatsen op het grote Friese meer.

138. Jan zei dat de astronaut uit een kleine provincie in Duitsland hard vooruit ging tijdens het schaatsen op het grote Friese meer.

139. Jan zei dat de chauffeur met een grijze baard en een kaal hoofd hard vooruit ging toen de kermisattractie begon te bewegen.

140. Jan zei dat de astronaut met een grijze baard en een kaal hoofd hard vooruit ging toen de kermisattractie begon te bewegen.

141. Jan zei dat de chauffeur uit een kleine provincie in Duitsland hard nadacht tijdens het schaatsen op het grote Friese meer.

142. Jan zei dat de astronaut uit een kleine provincie in Duitsland hard nadacht tijdens het schaatsen op het grote Friese meer.
143. Jan zei dat de chauffeur met een grijze baard en een kaal hoofd hard nadacht toen de kermisattractie begon te bewegen.

144. Jan zei dat de astronaut met een grijze baard en een kaal hoofd hard nadacht toen de kermisattractie begon te bewegen.

145. Marie zei dat de tandarts uit een heel erg klein dorp in Friesland opeens wegliep toen de postbode heel erg vriendelijk lachte.

146. Marie zei dat de verpleegster uit een heel erg klein dorp in Friesland opeens wegliep toen de postbode heel erg vriendelijk lachte.

147. Marie zei dat de tandarts met een glimmende rode auto plotseling wegliep op het moment dat iemand de tuin betrad.

148. Marie zei dat de verpleegster met een glimmende rode auto plotseling wegliep op het moment dat iemand de tuin betrad.

149. Marie zei dat de tandarts uit een heel erg klein dorp in Friesland opeens zong toen de postbode heel erg vriendelijk lachte.

150. Marie zei dat de verpleegster uit een heel erg klein dorp in Friesland opeens zong toen de postbode heel erg vriendelijk lachte.

151. Marie zei dat de tandarts met een glimmende rode auto plotseling zong op het moment dat iemand de tuin betrad.

152. Marie zei dat de verpleegster met een glimmende rode auto plotseling zong op het moment dat iemand de tuin betrad.

153. Jan zei dat de houthakker met een voorliefde voor de Spaanse kust onverwachts wegliep toen de gastheer de gasten welkom heette.

154. Jan zei dat de bankier met een voorliefde voor de Spaanse kust onverwachts wegliep toen de gastheer de gasten welkom heette.

155. Jan zei dat de houthakker uit een kleine provincie in Frankrijk plotseling wegglipte op het moment dat het toneelstuk had moeten beginnen.

156. Jan zei dat de bankier uit een kleine provincie in Frankrijk plotseling wegglipte op het moment dat het toneelstuk had moeten beginnen.

157. Jan zei dat de houthakker met een voorliefde voor de Spaanse kust onverwachts praatte toen de gastheer de gasten welkom heette.

158. Jan zei dat de bankier met een voorliefde voor de Spaanse kust onverwachts praatte toen de gastheer de gasten welkom heette.
159. Jan zei dat de houthakker uit een kleine provincie in Frankrijk plotseling praatte op het moment dat het toneelstuk had moeten beginnen.

160. Jan zei dat de bankier uit een kleine provincie in Frankrijk plotseling praatte op het moment dat het toneelstuk had moeten beginnen.

**Fillers**

161. Marie zei dat het groene en bruine onkruid in de tuin onverwachts lachtte toen het meisje de straat oprende met een fles water in haar hand.

162. Jan zei dat de roos naast de voordeur van het vakantiehuis eindelijk begon te zingen nadat de eigenaren wanhopig waren geworden.

163. Marie zei dat de taart van oma geheimzinning sprak toen hij in de keuken op de keukentafel stond voordat de gasten binnen kwamen.

164. Jan zei dat de brandweerman met een passie voor televisie kijken verwelkte nadat Siamese kittens waren gered uit het brandende huis.

165. Marie zei dat de dichter uit een kleine Italiaanse provincie onverwachts plette nadat de gasten uit het hotel waren verdwenen.

166. Jan zei dat de timmerman met een grote zwart-wit fotocollectie gebeurde in de hoek van het huis voor zeventien uur gedurende het onweer.

167. Marie zei dat de visser uit het zuiden van het land toenam toen de helicopter boven hem vloog op het strand vorige week dinsdag.

168. Jan zei dat de advocaat die van snoep houdt en de neiging heeft om aan te komen wilde toen een padvinder hem koekjes probeerde te verkopen.

169. Marie zei dat de vaas op de schoorsteenmantel in de gang fronste toen de operazanger op de radio te horen was.

170. Jan zei dat het bijennest in de boom van de buren nieste toen het plotseling begon te regenen vorige week dinsdag.

171. Marie zei dat de dakdekker met een strooien hoed, drie hamers en een beitel gehate na de beste gereedschapwinkel in de stad werd gesloten.

172. Jan zei dat de psycholoog met een diploma van een universiteit in Minnesota verteld wanneer de assistent draaide de airco op volle toeren aanzetten.

173. Marie zei dat de chirurg uit Oklahoma doelbewust de schuld bij de assistent goot een volle fles chloroform op de linoleum vloer.
174. Jan zei dat de schrijver met een grote groene Jeep schold na de drie weerbarstige peuters maakte een puinhoop in de woonkamer.

175. Marie zei dat de conciërge met een oranje zijden broek uit Rotterdam geprezen als een mooi ogende vrouw met een gebloemde jurk kwam de kamer binnen.

176. Jan zei dat de ballerina met een blauwe panty zei in de voorkant van het toejuichen publiek na de show eindigde.

177. Marie zei dat de luitenant met een goed gestreken overhemd en glanzend loafers smaakte toen de serveerster de dure wijn bracht.

178. Jan zei dat de kleermaker uit het oosten van Amsterdam genegeerd wanneer het tijd om de smokings en jurken passen voor de deelnemers aan de bruiloft was.

179. Marie zei dat de agent van het koor van de kerk genoten in de elfde gezinshereniging in het herenhuis.

180. Jan zei dat de verhuurder van het gebouw op de Koningsstraat pakte toen de nieuw ondertekende huurcontract bijna vloog weg in het licht 's avonds briesje.
Appendix 5: Verb Information for the Dutch Eye-Tracking Study

<table>
<thead>
<tr>
<th>Dutch verb</th>
<th>Translation</th>
<th>English</th>
<th>Koring</th>
<th>Imageability</th>
<th>Log Frequency</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>aankomen</td>
<td>arrive</td>
<td>0</td>
<td>0</td>
<td>2.73</td>
<td>8.17</td>
<td>UA</td>
</tr>
<tr>
<td>aarzelen</td>
<td>hesitate</td>
<td>1</td>
<td>0</td>
<td>2.40</td>
<td>5.95</td>
<td>UE</td>
</tr>
<tr>
<td>arriveren</td>
<td>arrive</td>
<td>1</td>
<td>1</td>
<td>3.00</td>
<td>6.77</td>
<td>UA</td>
</tr>
<tr>
<td>fluisteren</td>
<td>whisper</td>
<td>1</td>
<td>0</td>
<td>3.67</td>
<td>6.32</td>
<td>UE</td>
</tr>
<tr>
<td>gapen</td>
<td>yawn</td>
<td>0</td>
<td>1</td>
<td>3.93</td>
<td>4.60</td>
<td>UE</td>
</tr>
<tr>
<td>gillen</td>
<td>yell</td>
<td>1</td>
<td>0</td>
<td>3.87</td>
<td>6.72</td>
<td>UE</td>
</tr>
<tr>
<td>gluren</td>
<td>lurk</td>
<td>1</td>
<td>0</td>
<td>3.27</td>
<td>5.57</td>
<td>UE</td>
</tr>
<tr>
<td>klagen</td>
<td>complain</td>
<td>1</td>
<td>0</td>
<td>2.67</td>
<td>7.50</td>
<td>UE</td>
</tr>
<tr>
<td>mompelen</td>
<td>mumble</td>
<td>0</td>
<td>0</td>
<td>3.80</td>
<td>4.68</td>
<td>UE</td>
</tr>
<tr>
<td>mopperen</td>
<td>grumble</td>
<td>1</td>
<td>1</td>
<td>3.40</td>
<td>4.20</td>
<td>UE</td>
</tr>
<tr>
<td>nadenken</td>
<td>thinking</td>
<td>0</td>
<td>0</td>
<td>2.13</td>
<td>8.37</td>
<td>UE</td>
</tr>
<tr>
<td>ontwaken</td>
<td>awake</td>
<td>1</td>
<td>1</td>
<td>3.67</td>
<td>6.11</td>
<td>UA</td>
</tr>
<tr>
<td>opduiken</td>
<td>emerge</td>
<td>1</td>
<td>0</td>
<td>3.07</td>
<td>6.23</td>
<td>UA</td>
</tr>
<tr>
<td>opscheppen</td>
<td>brag</td>
<td>1</td>
<td>0</td>
<td>3.27</td>
<td>5.34</td>
<td>UE</td>
</tr>
<tr>
<td>opstaan</td>
<td>rise</td>
<td>1</td>
<td>0</td>
<td>3.40</td>
<td>7.74</td>
<td>UA</td>
</tr>
<tr>
<td>praten</td>
<td>talk</td>
<td>1</td>
<td></td>
<td>3.87</td>
<td>10.63</td>
<td>UE</td>
</tr>
<tr>
<td>roepen</td>
<td>shout</td>
<td>1</td>
<td>0</td>
<td>4.00</td>
<td>8.70</td>
<td>UE</td>
</tr>
<tr>
<td>ronddolen</td>
<td>wander</td>
<td>0</td>
<td>0</td>
<td>2.80</td>
<td>3.81</td>
<td>UE</td>
</tr>
<tr>
<td>schreeuwen</td>
<td>scream</td>
<td>1</td>
<td>1</td>
<td>3.87</td>
<td>7.88</td>
<td>UE</td>
</tr>
<tr>
<td>schrikken</td>
<td>be shocked</td>
<td>0</td>
<td>1</td>
<td>3.40</td>
<td>7.80</td>
<td>UA</td>
</tr>
<tr>
<td>slapen</td>
<td>sleep</td>
<td>1</td>
<td>0</td>
<td>3.67</td>
<td>9.74</td>
<td>UE</td>
</tr>
<tr>
<td>sterven</td>
<td>die</td>
<td>1</td>
<td>0</td>
<td>3.27</td>
<td>9.79</td>
<td>UA</td>
</tr>
<tr>
<td>struikelen</td>
<td>stumble</td>
<td>0</td>
<td></td>
<td>3.93</td>
<td>5.84</td>
<td>UA</td>
</tr>
<tr>
<td>terugkomen</td>
<td>return</td>
<td>1</td>
<td>0</td>
<td>3.00</td>
<td>8.47</td>
<td>UA</td>
</tr>
<tr>
<td>vallen</td>
<td>fall</td>
<td>1</td>
<td>1</td>
<td>3.80</td>
<td>9.94</td>
<td>UA</td>
</tr>
<tr>
<td>verdwalen</td>
<td>get lost</td>
<td>0</td>
<td>1</td>
<td>3.13</td>
<td>7.22</td>
<td>UA</td>
</tr>
<tr>
<td>verdwijnen</td>
<td>vanish</td>
<td>1</td>
<td>1</td>
<td>2.87</td>
<td>8.83</td>
<td>UA</td>
</tr>
<tr>
<td>Dutch Verb</td>
<td>English Equivalent</td>
<td>Total Occurrences</td>
<td>First Appearance</td>
<td>Last Appearance</td>
<td>Location</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>verschijnen</td>
<td>appear</td>
<td>1</td>
<td>1</td>
<td>2.80</td>
<td>7.30</td>
<td>UA</td>
</tr>
<tr>
<td>vertrekken</td>
<td>leave</td>
<td>1</td>
<td>0</td>
<td>3.27</td>
<td>9.19</td>
<td>UA</td>
</tr>
<tr>
<td>vloeken</td>
<td>swear</td>
<td>1</td>
<td>1</td>
<td>3.60</td>
<td>6.03</td>
<td>UE</td>
</tr>
<tr>
<td>vluchten</td>
<td>flee</td>
<td>0</td>
<td>0</td>
<td>2.93</td>
<td>7.58</td>
<td>UA</td>
</tr>
<tr>
<td>vooruitgaan</td>
<td>set forward</td>
<td>0</td>
<td>0</td>
<td>2.80</td>
<td>4.38</td>
<td>UA</td>
</tr>
<tr>
<td>weggaan</td>
<td>depart</td>
<td>1</td>
<td>0</td>
<td>3.67</td>
<td>8.70</td>
<td>UA</td>
</tr>
<tr>
<td>wegglippen</td>
<td>slip away</td>
<td>0</td>
<td>0</td>
<td>3.40</td>
<td>3.89</td>
<td>UA</td>
</tr>
<tr>
<td>wegsluipen</td>
<td>disappear</td>
<td>1</td>
<td>0</td>
<td>3.47</td>
<td>3.40</td>
<td>UA</td>
</tr>
<tr>
<td>zeuren</td>
<td>whine</td>
<td>1</td>
<td>1</td>
<td>3.60</td>
<td>7.02</td>
<td>UE</td>
</tr>
<tr>
<td>zuchten</td>
<td>sigh</td>
<td>1</td>
<td>0</td>
<td>4.13</td>
<td>4.20</td>
<td>UE</td>
</tr>
<tr>
<td>zwaaien</td>
<td>waving</td>
<td>0</td>
<td>0</td>
<td>4.00</td>
<td>6.75</td>
<td>UE</td>
</tr>
<tr>
<td>weglopen</td>
<td>to walk away</td>
<td>0</td>
<td>0</td>
<td>3.13</td>
<td>7.43</td>
<td>UA</td>
</tr>
<tr>
<td>zingen</td>
<td>to sing</td>
<td>0</td>
<td>0</td>
<td>4.27</td>
<td>8.62</td>
<td>UE</td>
</tr>
</tbody>
</table>
Appendix 6: Verb Imageability for the Koring et al. Study

<table>
<thead>
<tr>
<th>Verb</th>
<th>English translation</th>
<th>Type</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>verschijnen</td>
<td>(appear)</td>
<td>UA</td>
<td>4.33</td>
</tr>
<tr>
<td>ontstaan</td>
<td>(arise)</td>
<td>UA</td>
<td>2.33</td>
</tr>
<tr>
<td>arriveren</td>
<td>(arrive)</td>
<td>UA</td>
<td>3.89</td>
</tr>
<tr>
<td>ontwaken</td>
<td>(awake)</td>
<td>UA</td>
<td>5.11</td>
</tr>
<tr>
<td>schrikken</td>
<td>(be shocked)</td>
<td>UA</td>
<td>5.67</td>
</tr>
<tr>
<td>mislukken</td>
<td>(be unsuccessful)</td>
<td>UA</td>
<td>3.07</td>
</tr>
<tr>
<td>juichen</td>
<td>(cheer)</td>
<td>UE</td>
<td>6.26</td>
</tr>
<tr>
<td>instorten</td>
<td>(collapse)</td>
<td>UA</td>
<td>5.22</td>
</tr>
<tr>
<td>barsten</td>
<td>(crack)</td>
<td>UA</td>
<td>5.41</td>
</tr>
<tr>
<td>huilen</td>
<td>(cry)</td>
<td>UE</td>
<td>6.33</td>
</tr>
<tr>
<td>fietsen</td>
<td>(cycle)</td>
<td>UE</td>
<td>6.74</td>
</tr>
<tr>
<td>dansen</td>
<td>(dance)</td>
<td>UE</td>
<td>6.48</td>
</tr>
<tr>
<td>uitsterven</td>
<td>(die out)</td>
<td>UA</td>
<td>3.00</td>
</tr>
<tr>
<td>verdwijnen</td>
<td>(disappear)</td>
<td>UA</td>
<td>4.07</td>
</tr>
<tr>
<td>omvallen</td>
<td>(fall over)</td>
<td>UA</td>
<td>5.41</td>
</tr>
<tr>
<td>vallen</td>
<td>(fall)</td>
<td>UA</td>
<td>5.70</td>
</tr>
<tr>
<td>verdwalen</td>
<td>(get lost)</td>
<td>UA</td>
<td>4.11</td>
</tr>
<tr>
<td>groeien</td>
<td>(grow)</td>
<td>UA</td>
<td>4.81</td>
</tr>
<tr>
<td>grommen</td>
<td>(growl)</td>
<td>UE</td>
<td>5.26</td>
</tr>
<tr>
<td>mopperen</td>
<td>(grumble)</td>
<td>UE</td>
<td>4.67</td>
</tr>
<tr>
<td>toeteren</td>
<td>(hoot)</td>
<td>UE</td>
<td>6.00</td>
</tr>
<tr>
<td>huppelen</td>
<td>(hop)</td>
<td>UE</td>
<td>6.04</td>
</tr>
<tr>
<td>toenemen</td>
<td>(increase)</td>
<td>UA</td>
<td>3.48</td>
</tr>
<tr>
<td>springen</td>
<td>(jump)</td>
<td>UE</td>
<td>6.07</td>
</tr>
<tr>
<td>lachen</td>
<td>(laugh)</td>
<td>UE</td>
<td>6.59</td>
</tr>
<tr>
<td>zeuren</td>
<td>(nag)</td>
<td>UE</td>
<td>4.30</td>
</tr>
<tr>
<td>hijgen</td>
<td>(pant)</td>
<td>UE</td>
<td>5.85</td>
</tr>
</tbody>
</table>
Appendix 6: Verb Imageability for the Koring et al. Study (continued)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>stijgen</td>
<td>(rise)</td>
<td>UA</td>
<td>5.07</td>
</tr>
<tr>
<td>rennen</td>
<td>(run)</td>
<td>UE</td>
<td>6.48</td>
</tr>
<tr>
<td>schreeuwen</td>
<td>(scream)</td>
<td>UE</td>
<td>6.19</td>
</tr>
<tr>
<td>krijsen</td>
<td>(shriek)</td>
<td>UE</td>
<td>5.58</td>
</tr>
<tr>
<td>zinken</td>
<td>(sink)</td>
<td>UA</td>
<td>5.56</td>
</tr>
<tr>
<td>glimlachen</td>
<td>(smile)</td>
<td>UE</td>
<td>6.44</td>
</tr>
<tr>
<td>wandelen</td>
<td>(stroll)</td>
<td>UE</td>
<td>6.41</td>
</tr>
<tr>
<td>vloeken</td>
<td>(swear)</td>
<td>UE</td>
<td>4.89</td>
</tr>
<tr>
<td>opzwellen</td>
<td>(swell up)</td>
<td>UA</td>
<td>5.04</td>
</tr>
<tr>
<td>praten</td>
<td>(talk)</td>
<td>UE</td>
<td>6.04</td>
</tr>
<tr>
<td>verwelken</td>
<td>(wilt)</td>
<td>UA</td>
<td>4.85</td>
</tr>
<tr>
<td>lukken</td>
<td>(work out well)</td>
<td>UA</td>
<td>2.74</td>
</tr>
<tr>
<td>gapen</td>
<td>(yawn)</td>
<td>UE</td>
<td>6.56</td>
</tr>
</tbody>
</table>
Appendix 7: Stimuli for the unaccusative diagnostics

There-insertion

Plain sentences

1. The Saint Patrick's Parade advanced on the Fifth Avenue.
2. Many elders ambled along the riverbank.
3. A blemish appeared on the surface of the vase.
4. A troop of three thousand soldiers approached.
5. A problem arose in the research plan.
6. A man wearing a red hat arrived at the ceremony.
7. The chairman's voice ascended.
8. The lion with a large territory awoke.
9. The lamp pole with vintage patterns bent.
10. The naughty boy blinked at his friend.
11. A vase with delicate decoration broke.
13. The roof collapsed on top of the kid.
14. A bus full of tourists came into the station.
15. Many customers complained about the quality of the product.
16. The patient with a terrible cold coughed.
17. The seven-month-old baby cried in the bedroom.
18. The pop singer danced on the stage.
19. The luxury cruise ship departed for Germany.
20. The aircraft descended into the desert.
21. A strange closeness developed between the two competitors.
22. A beautiful blue bird died in the garden.
23. A student disappeared in the forest.
24. A hyena emerged from the jungle.
25. Five prisoners escaped from the jail.
27. A heart-shaped leaf fell off the tree.
28. The armed bank robbers fled from the police.
29. A dove flew in the sky.
30. The picky teacher frowned as he graded the homework.
31. The CEO of the company went to the conference room.
32. A haze of smoke hung below the ceiling.
33. A miracle happened at Christmas.
34. An old couple jogged in the park.
35. A fuzzy furry cat jumped from the tree.
36. The headmaster's speech lasted for an hour.
37. An awkward man laughed during the lecture.
38. The goat leaped across the creek.
39. The officer with a black tie left.
40. A group of deer lay near the river.
41. A visitor from Spain lingered around the museum.
42. The queen lived in the castle.
43. A suspicious looking man lurked around the bank.
44. A riot occurred on the street.
45. A window of the old house opened.
46. The ambitious athlete panted after the intensive training.
47. The amusing talk show played on the TV.
48. The famous basketball player practiced before the game.
49. An atmosphere of crisis prevailed.
50. The dancers rehearsed before the show.
51. An issue remained in the proposal.
52. A famous actor resided in the area.
53. The president returned to the White House.
54. The cracking bell rang with a loud noise.
55. The sun rose over the mountain.
56. The ball rolled down the hill.
57. The giant steel wheels rotated.
58. A cute dog ran across the road.
59. The interviewee rushed into the conference room.
60. A timid girl screamed in the dark room.
61. The two-hundred-year-old building shook in the earthquake.
62. The tiny white rabbit shivered in the shrub.
63. The bus driver shouted to the passengers.
64. A well-known director sat in the hall.
65. A pack of wolves slept under the sun.
66. A pretty little girl smiled to the strangers.
67. The presenter in a black button-down shirt sneezed during his talk.
68. The mayor of the city spoke in front of the City Hall.
69. The car speeded up as it went around the corner.
70. A funny statue stood in the garden.
71. Much of the audience stayed at the stadium.
72. A tourist strolled around the city.
73. Twenty-one passengers survived.
74. A graceful swan swam in the lake.
75. The little kid with a red apple in his hand trembled.
76. The magician vanished from the stage.
77. A stranger walked across the lawn.
78. A staff of sixty worked in the bank.
79. The manager with a high level of anxiety worried about the budget.
80. Many guests at the cocktail party yawned due to boredom.

*There-insertion sentences*

1. There advanced the Saint Patrick's Parade on the Fifth Avenue.
2. There ambled many elders along the riverbank.
3. There appeared a blemish on the surface of the vase.
4. There approached a troop of three thousand soldiers.
5. There arose a problem in the research plan.
6. There arrived a man wearing a red hat at the ceremony.
7. There ascended the chairman's voice.
8. There awoke the lion with a large territory.
There bent the lamp pole with vintage patterns.
There blinked the naughty boy at his friend.
There broke a vase with delicate decoration.
There climbed an adventurous mountaineer.
There collapsed the roof on top of the kid.
There came a bus full of tourists into the station.
There complained many customers about the quality of the product.
There coughed the patient with a terrible cold.
There cried the seven-month-old baby in the bedroom.
There danced the pop singer on the stage.
There departed the luxury cruise ship for Germany.
There descended the aircraft into the desert.
There developed a strange closeness between the two competitors.
There died a beautiful blue bird in the garden.
There disappeared a student in the forest.
There emerged a hyena from the jungle.
There escaped five prisoners from the jail.
There existed an optimal solution.
There fell a heart-shaped leaf off the tree.
There fled the armed bank robbers from the police.
There flew a dove in the sky.
There frowned the picky teacher as he graded the homework.
There went the CEO of the company to the conference room.
32. There hung a haze of smoke below the ceiling.
33. There happened a miracle at Christmas.
34. There jogged an old couple in the park.
35. There jumped a fuzzy furry cat from the tree.
36. There lasted the headmaster's speech for an hour.
37. There laughed an awkward man during the lecture.
38. There leaped a goat across the creek.
39. There left the officer with a black tie.
40. There lay a group of deer near the river.
41. There lingered a visitor from Spain around the museum.
42. There lived the queen in the castle.
43. There lurked a suspicious looking man around the bank.
44. There occurred a riot on the street.
45. There opened a window of the old house.
46. There panted the ambitious athlete after the intensive training.
47. There played the amusing talk show on the TV.
48. There practiced the famous basketball player before the game.
49. There prevailed an atmosphere of crisis.
50. There rehearsed the dancers before the show.
51. There remained an issue in the proposal.
52. There resided a famous actor in the area.
53. There returned the president to the White House.
54. There rang the cracking bell with a loud noise.
There rose the sun over the mountain.

There rolled the ball down the hill.

There rotated the giant steel wheels.

There ran a cute dog across the road.

There rushed the interviewee into the conference room.

There screamed a timid girl in the dark room.

There shook the two-hundred-year-old building in the earthquake.

There shivered a tiny white rabbit in the shrub.

There shouted the bus driver to the passengers.

There sat a well-known director in the hall.

There slept a pack a wolf under the sun.

There smiled a pretty little girl to the strangers.

There sneezed the presenter in a black button-down shirt during his talk.

There spoke the mayor of the city in front of the City Hall.

There speeded up the car as it went around the corner.

There stood a funny statue in the garden.

There stayed much of the audience at the stadium.

There strolled a tourist around the city.

There survived twenty-one passengers.

There swam a graceful swan in the lake.

There trembled the little kid with a red apple in his hand.

There vanished the magician from the stage.

There walked a stranger across the lawn.
78. There worked a staff of sixty in the bank.
79. There worried the manager with a high level of anxiety about the budget.
80. There yawned many guests at the cocktail party due to boredom.

**Prenominal modifiers**

Plain sentences

1. The book arrived.
2. The man came.
3. The mailman approached
4. The girl went.
5. The woman left.
6. The flag ascended.
7. The team departed.
8. The professor returned.
9. The sun rose.
10. The leaf fell.
11. The thief escaped.
12. The prisoner fled.
13. The aircraft descended.
14. The troops advanced.
15. The old man died.
16. The musician appeared.
17. The package disappeared.
18. The magician vanished.
19. A problem arose.
20. The baby awoke.
21. The country developed.
22. A raven emerged.
23. The disturbance happened.
24. The accident occurred.
25. The roof collapsed.
26. The vase broke.
27. The tree bent.
28. The window opened.
29. The business man stayed.
30. The boy remained.
31. The condition lasted.
32. The passenger survived.
33. The girl lingered.
34. The king lived.
35. The birds resided.
36. The fear lurked.
37. The problem existed.
38. The statue stood.
39. The lady sat.
40. The wool hung well.
41. The child lay.
42. An atmosphere of crisis prevailed.
43. The dog shivered.
44. The ground shook.
45. The teacher trembled.
46. The wheel rotated.
47. The student coughed.
48. The cat sneezed.
49. The officer yawned.
50. The athlete panted.
51. The singer blinked.
52. The bell rang.
53. The bunny ran.
54. The duck swam.
55. The monkey climbed.
56. The wolf jumped.
57. The bird flew.
58. The lawyer jogged.
59. The ball rolled.
60. The tourist rushed.
61. The toddler walked.
62. The poet strolled.
63. The performer danced.
64. The car sped.
65. The goat leaped.
66. The artist ambled.
67. The employee worked.
68. The children played.
69. The chairman talked.
70. The dancer rehearsed.
71. The team practiced.
72. The baker complained.
73. The kid slept.
74. The captain shouted.
75. The boss screamed.
76. The infant cried.
77. The landlord laughed.
78. The hairdresser frowned.
79. The biologist smiled.
80. The senator thought.

Prenominal modifiers

1. the arrived book
2. the come man
3. the approached mailman
4. the gone girl
5. the left woman
6. the ascended flag
7. the departed team
8. the returned professor
9. the risen sun
10. the fallen leaf
11. the escaped thief
12. the fled prisoner
13. The descended aircraft
14. the advanced troop
15. the died old man
16. the appeared musician
17. the disappeared package
18. the vanished magician
19. an arisen problem
20. the awaken baby
21. the developed country
22. an emerged raven
23. the happened disturbance
24. the occurred accident
25. the collapsed roof
26. the broken vase
27. the bent tree
28. the opened window
29. the stayed business
the remained boy
the lasted condition
the survived passenger
the lingered girl
the lived king
the resided bird
the lurked fear
the existed problem
the stood statue
the sat lady
the well-hung wool
the lain child
a prevailed atmosphere of crisis
the shivered dog
the shaken ground
the trembled teacher
the rotated wheel
the coughed student
the sneezed cat
the yawned officer
the panted athlete
the blinked singer
the rung bell
the run bunny
the swum duck
the climbed monkey
the jumped wolf
the flown bird
the jogged lawyer
the rolled ball
the rushed tourist
the walked toddler
the strolled poet
the danced performer
the sped car
the leaped goat
the ambled artist
the worked employee
the played children
the talked chairman
the rehearsed dancer
the practiced team
the complained baker
the slept kid
the shouted captain
the screamed boss
76. the cried infant
77. the laughed landlord
78. the frowned hairdresser
79. the smiled biologist
80. the thought senator

**Nominalization**

1. advander
2. ambler
3. appearer
4. approacher
5. ariser
6. arriver
7. ascender
8. awaker
9. bender
10. blinker
11. breaker
12. climber
13. collapser
14. comer
15. complainer
16. cougher
17. crier
18. dancer
19. departer
20. descender
21. developer
22. dier
23. disappearer
24. emerger
25. escaper
26. exister
27. faller
28. fleer
29. flyer
30. frowner
31. goer
32. hanger
33. happener
34. jogger
35. jumper
36. ladder
37. laugher
38. leaper
39. leaver
40. lier
41. lingerer
42. liver
43. lurker
44. occurer
45. opener
46. panter
47. player
48. practicer
49. prevailer
50. rehearser
51. remainer
52. resider
53. returner
54. ringer
55. riser
56. roller
57. rotater
58. runner
59. rusher
60. screamer
61. shaker
62. shiverer
63. shouter
64. sitter
65. sleeper
66. smiler
67. sneezer
68. speeder
69. stander
70. stayer
71. stroller
72. surviver
73. swimmer
74. talker
75. thinker
76. trembler
77. vanisher
78. walker
79. worker
80. yawner
References:


Arregi, K. (2004). The have/be alternation in Basque. Ms. University of Illinois at Urbana Champaign. Chicago


Burzio, L. (1981). *Intransitive verbs and Italian auxiliaries*. Massachusetts Institute of Technology,


Fillmore, C. J. (1968). The grammar of 'hitting' and 'breaking'.


Sproat, R. W. (1985). *On deriving the lexicon*. Massachusetts Institute of Technology,


