The role of voice and word order in incremental sentence processing

Studies on sentence production and comprehension in Tagalog and German

Proefschrift ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen op gezag van de rector magnificus prof. dr. J. H. J. M. Krieken, volgens besluit van het college van decanen in het openbaar te verdedigen op

dinsdag 10 october 2017

om 16.30 uur precies

door

Sebastian Sauppe

geboren op 23 maart 1987

te Borna (Duitsland)

Promotoren
Prof. dr. Stephen C. Levinson
Prof. dr. Robert D. Van Valin, Jr (Heinrich-Heine-Universität Düsseldorf, Duitsland & University at Buffalo, The State University of New York, Verenigde Staten)

Copromotor
Dr. Elisabeth Norcliffe (Max-Planck-Institut für Psycholinguistik, Nijmegen)

Manuscriptcommissie
Prof. dr. Antje S. Meyer
Prof. dr. Nikolaus P. Himmelmann (Universität zu Köln, Duitsland)
Prof. dr. T. Florian Jaeger (University of Rochester, Verenigde Staten)

Cover graphic generated using http://qrohlf.com/trianglify/
Printed and bound by Ipskamp Printing, Enschede

ACKNOWLEDGEMENTS

A great number of people helped me the one way or the other during my time as a PhD student. Even if I don't mention all of them here to keep it short, they can be assured that I am grateful to all of them.

First of all, I would like to thank Stephen C. Levinson for giving me the opportunity to conduct my PhD research at the Language and Cognition department of the Max Planck Institute for Psycholinguistics and for giving me huge academic freedom, allowing me to explore my ideas, yet always pushing me back on track when I wandered to far astray. Without his ideas, insights and support this research would not have been possible.

I also thank Elisabeth Norcliffe for being the most wonderful mentor and supervisor I could have ever hoped for. I thank Robert D. Van Valin, Jr, for his advice, patiently answering the many questions that I had over the years and reading the drafts of all the papers in this thesis in basically no time.

I thank T. Florian Jaeger for the discussions on theories, methods and statistics that we had over the years and that helped to improve the research in this thesis greatly. I thank Xaver Koch for introducing me to pupillometry and for many very helpful discussions. His raised eyebrows led me to overthink many methodological and statistical decisions. I thank Agnieszka Konopka for introducing me to eye tracking, many helpful discussions and answering innumerable questions about sentence production theories and about eye tracking. I thank Anja Latrouite for her insights into Tagalog grammar and for helping me make sense of some of syntactic phenomena.

I am deeply grateful to Philip A. Rentillo, who helped me to conduct the Tagalog production experiments and who has become a dear friend since I first met him in 2012. He patiently answered all my questions about how to say this or that in Tagalog; without his translation help, it would not have been possible to conduct these experiments. I also thank Inger M. Montemor for translating various materials into Tagalog and for answering my questions about the language, Ronald Fischer and Reiner Dirksmeyer for their technical support, and Frédérique Schless and Emiel van den Hoven for help in processing the data I collected.

I thank Edith Sjoerdsma for her help with so many practicalities and administrative things. I thank Frédérique Schless, Julija Baranova and Edith Sjoerdsma for help with the translation of the Dutch summary of the thesis results. My thank goes also to Luise Dorenbusch for invaluable help with the typesetting of this thesis.

I warmly thank Gabriela Garrido Rodriguez and Mathias Barthel for sharing the apartment or the office with me, respectively, during my time in Nijmegen, for their critical questions along the way of carrying out and writing up the research in this thesis, and for being my *paranymphoi*.

Finally, I thank Julia and Friederike for being there and for motivating me to be good at what I do.

CONTENTS

Part I

INTRODUCTORY REMARKS

1

INTRODUCTORY REMARKS

> *Die Frage, was der Satz sei, in welchem Verhältnis er zum Wort, zu den psychischen Vorgängen der Verbindung der Vorstellungen und zu den logischen Urteilsakten stehe, — diese Fragen gehören zu den meistumstrittenen oder gelegentlich wohl auch zu den mit Vorliebe vermiedenen der neueren Sprachwissenschaft.*
>
> Wilhelm Wundt

Language processing is fundamentally incremental. Both during the production and the comprehension of sentences, humans process information in pieces rather than as a whole.

When formulating a sentence, speakers do not wait until they have planned all of the phonological and phonetic detail of all words in their utterance before they start speaking. Rather, they may start to articulate the sentence as soon as enough information has been encoded. Similarly, during sentence comprehension, listeners do not wait until the end of their interlocutor's sentence before assigning any grammatical structures or semantic interpretations to the words that they have heard. Rather, they start to interpret the incoming speech continuously and make predictions about upcoming input as soon as possible.

The papers in this thesis are concerned with three fundamental issues revolving around how the interaction of the grammatical structures of languages and the planning and understanding of event structure and the relations among agents and patients affect incremental sentence production and comprehension.

Specifically, two studies on Tagalog (an Austronesian language of the Philippines) and German are reported that investigate the influence of verb position and voice marking on what kind of information has to be encoded before speaking can begin, i.e., what constitutes "enough information" when the verb is the first, the second or the last word in a sentence (Chapters 2 and 3). Next, the influence of the different voice marking systems of the two languages on cognitive processing load during sentence formulation is investigated by analyzing pupillometric data from these production studies (Chapter 4). Finally, a comprehension experiment studied the influence of verb-initial word order and verbal morphology on listeners' use of information about the verb's meaning and potential arguments to anticipate upcoming linguistic input (Chapter 5).

In their combination, these different studies allow us to look at what roles word order, voice marking and the semantics of events play in sentence production and comprehension, providing insights into whether similar or different mechanisms are at play in production and comprehension because the tasks that listeners and speakers have to perform may be different. Thus, all four studies included in this thesis are concerned with how and when (a) information about actions and events is processed and (b) morphological and word order information is integrated into the representations used in the production and comprehension of sentences.

The remainder of this introduction gives a brief overview of incrementality in sentence processing and the studies reported in Chapters 2–5 are briefly motivated to set the scene for detailed descriptions of the experiments in the individual chapters, including the theoretical backgrounds and hypotheses for each of the studies.

1.1 INCREMENTALITY IN SENTENCE PRODUCTION

Theories of sentence production generally assume that the formulation process consists of several stages. For example, Bock and Levelt's (1994) model assumes that speakers start with the generation of a preverbal message that encodes what concepts they want to convey in their utterance. This is followed by a grammatical encoding stage in which lemmas are selected, syntactic functions are assigned and word order is determined. Finally, the planned material is phonologically encoded and articulated.

Sentence production theories also generally assume that formulation proceeds incrementally (e.g., De Smedt, 1994; F. Ferreira & Swets, 2002; Iwasaki, 2011; Kempen & Hoenkamp, 1987). This means that "that the next processing component in the general flow of information can start working on the still incomplete output of the current processor" (Levelt, 1999, 88), i.e., as soon as a piece of information or an increment is processed at one stage it can be passed on to next stage and it is not necessary that all processing in that stage is finished first. The incremental nature of production may thus help speakers to maintain fluency (e.g., Brown-Schmidt & Konopka, 2015; Griffin, 2001, 2003; Jaeger, Furth, & Hilliard, 2012; Myachykov, Scheepers, Garrod, Thompson, & Fedorova, 2013).

There are different ways of how sentence production may be incremental, i.e., how messages are transformed into ordered strings of words, and thus which encoding operations are carried out first.

Accounts of linearly incremental production assume that the ease of encoding individual parts of a message drives structure building, syntactic function assignment and the linear ordering of words (e.g., V. S. Ferreira & Dell, 2000; Gleitman, January, Nappa, & Trueswell, 2007; Kempen & Hoenkamp, 1987; Levelt, 1989). Concepts that are more accessible conceptually or lexically (e.g., animates as compared to inanimates or entities that were mentioned already earlier in discourse) may be easier to encode conceptually and are thus available for syntactic function assignment earlier (Bock & Warren, 1985; Branigan, Pickering, & Tanaka, 2008; Prat-Sala & Branigan, 2000). The utterance and its grammatical structure is thus built bottom-up by retrieving one word after the other.

By contrast, hierarchically incremental accounts assume that the formulation of sentences begins with the generation of a representation that incorporates information about the described event and about the involved participants, their semantic roles (e.g., agent and patient) and to which syntactic functions they are mapped (Bock, Irwin, Davidson, & Levelt, 2003; Griffin & Bock, 2000; Lee, Brown-Schmidt, & Watson, 2013). This representation then guides lexical retrieval and linguistic encoding (e.g., Norcliffe, Konopka, Brown, & Levinson, 2015; van de Velde, Meyer, & Konopka, 2014). Under this view, incremental production proceeds "word by word" only with respect to linguistic encoding. The building of grammatical structures and the linearization of concepts is carried out beforehand and is assumed to operate independently of lexical processes (e.g., Chang, Dell, & Bock, 2006; Konopka & Meyer, 2014).

These accounts of incremental production thus differ with respect to how much weight they give to the encoding of information about the relations among event participants and the action carried out on the one hand and the ease of encoding of individual parts of messages on the other hand. Put differently, events are central to hierarchical incrementality, in which an event representation is the basis of the early generation of an utterance plan guiding lexical retrieval (Griffin & Bock, 2000). Conversely, linear incrementality does not assume that speakers encode event information early; planning to describe the action carried out may rather be deferred until later points in time, such as when a verb has to be retrieved (Gleitman et al., 2007; Schriefers, Teruel, & Meinshausen, 1998).

Evidence in favor of both accounts of incremental sentence production has been brought forth and there is little consensus in the literature regarding the size and nature of planning units. Furthermore, several studies suggest that there is some flexibility with respect to the choice between linearly and hierarchically incremental production, which maybe influenced by both language-internal and language-external factors (e.g., F. Ferreira & Swets, 2002; Konopka, 2012; Swets, Jacovina, & Gerrig, 2014).

Another factor that can influence the choice between planning modes is the grammatical structures of the target language. While most sentence production studies so far have concentrated on a small set of closely related European languages (Jaeger & Norcliffe, 2009), there is a growing body of cross-linguistic research that addresses the influence of different grammatical structures on the time course of sentence planning (Norcliffe, Harris, & Jaeger, 2015).

Most notably, Norcliffe, Konopka, et al. (2015) investigated the influence of verb-initial and subject-initial word orders on incremental planning in Tzeltal, a Mayan language spoken in Mexico. They found that speakers of this language engaged in extensive relational encoding at the outset of formulation when a verb was produced first, but concentrated on the encoding of the subject phrase from early on when producing a subject-initial sentence. This suggests that word order may influence the time course of planning and may determine when different kinds of information are encoded with priority (i.e., relations among event participants vs. individual parts of messages).

The studies in Chapters 2 and 3 take these results as a starting point and investigate the influence of verb position on sentence production in Tagalog and German. Specifically, these two studies are designed to test whether speakers of these languages plan sentences in a hierarchically incremental or linearly incremental fashion, depending on verb position. While in Tagalog (an Austronesian language spoken in the Philippines) the verb is always positioned sentence-initially, it may be placed sentence-medially or sentence-finally in German.

In Chapter 2, two eye tracking experiments on Tagalog are reported that exploit the language's verb-initial basic word order to investigate whether sentence formulation in this language is influenced by the need to express the relationship between event participants early in the sentence with a verb as the first word. In some sentence types, Tagalog verbs also exhibit head-marking morphology (Nichols, 1986), i.e., they agree with one of their arguments in semantic role; in other sentence types there is no head-marking morphology and the verb takes an invariant shape. This grammatical property is exploited to investigate how the formulation process is influenced by the need to compute cross-referencing morphological dependencies on the first word of the sentence.

In this chapter, the accounts of linearly and hierarchically incremental sentence production are described in more detail and it is argued that linear incrementality cannot model verb-initial sentence production because its predictions are focused on the early encoding of accessible nominal concepts. This makes linear incrementality potentially suitable for modeling sentence planning in subject-initial but not necessarily in verb-initial languages. Therefore, Chapter 2 focuses on hierarchical incrementality and its predictions for the production of Tagalog sentences, exploring how structural choices and the time course of formulation are influenced by the head-marking morphology that the sentence-initial verb carries.

In Chapter 3, an eye tracking experiment is reported that makes use of variable word order in German. In this language, the lexical verb may either occur immediately after the subject in sentence-medial position or at the end of the sentence. The experiment explored the timing of relational encoding and verb planning in sentences with different word orders in order to test the predictions of linear and hierarchical incrementality. Specifically, it was investigated whether speakers always generated an utterance plan at the outset of sentence formulation and whether this plan involves only conceptual encoding of the depicted action or also the selection of a verb. To capture these two different possible planning scopes for utterance plans, a distinction between weakly and strongly hierarchical incrementality is introduced. While weakly incremental planning only assumes that speakers always engage in relational encoding to identify the action in a depicted event early (Griffin & Bock, 2000; Hwang & Kaiser, 2014; Kuchinsky, Bock, & Irwin, 2011), the strongly hierarchical version assumes that relational encoding additionally includes or is immediately followed by the selection of a verb because verb lemmas may play a crucial role in syntactic function assignment (Bock & Levelt, 1994; F. Ferreira, 2000). To test the predictions of these different approaches, fixations to agent and patient characters before and after speech onset were compared and the influence of accessibility on the choice between active and passive syntax was investigated.

In Chapter 4, a study is reported that investigates the consequences of the differences between the voice systems of Tagalog and German for sentence production and the development of cognitive processing load over the course of formulation. It has been proposed in the literature that Tagalog exhibits a symmetrical voice system (Foley, 2008). This voice system type is only found in Austronesian languages and is characterized by the fact that all voices are equally marked morphologically, independently of whether the agent, the patient, or some other argument carries the highest syntactic function. German's active/passive voice system, on the other hand, is characterized by an asymmetry: the active is the unmarked voice, both morphosyntactically and in terms of usage frequency, in contrast to the passive. The study tests whether these differences in the grammatical architecture of the voice systems have consequences for sentence production by analyzing the time course of pupil size changes, which indicate changes of mental effort. It is assumed that this allows to test whether similar or different cognitive operations are carried out during during the process of formulating and uttering sentences in Tagalog and German.

1.2 INCREMENTALITY IN SENTENCE COMPREHENSION

Following the first part of this thesis which presented three studies on sentence production, the second part is devoted to sentence comprehension and comprises a study on the comprehension of Tagalog sentences.

Sentence comprehension theories also assume that this is an incremental process during which listeners or readers interpret incoming linguistic input immediately. Comprehenders integrate the words that are heard or read as they are encountered into their mental representation of the sentence, incrementally building up a grammatical structure of the utterance (e.g. Altmann & Steedman, 1988; Boland, Tanenhaus, Garnsey, & Carlson, 1995; Knoeferle, Crocker, Scheepers, & Pickering, 2005; Marslen-Wilson, 1975).

For example, in one of the first eye-tracking experiments on incremental sentence comprehension, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) let participants manipulate real-world objects and showed that listeners direct their gaze towards relevant objects and object locations immediately after hearing the corresponding words.

In addition to incremental integration of the input, most theories assume that language comprehension is anticipatory. This means that listeners use the context of an utterance to make predictions about upcoming words before they are actually encountered (Kamide, 2008). Prediction might be one reason why understanding sentences is in general such an effortless and rapid process (Huettig, 2016). Many studies have shown that prediction and anticipation are a central characteristic of comprehension. For example, words that are predictable from context are fixated for less time during reading as compared to less predictable words (Rayner & Well, 1996; Staub, 2015). Kutas and Hillyard (1980) showed that semantically unpredictable words induce an N400 event-related potential in EEG recordings, signaling the violation of listeners' semantic expectations.

Comprehenders may make predictions about the following input on a number of different levels, including, e.g., phonological form (DeLong, Urbach, & Kutas, 2005), semantic plausibility (Kutas & Hillyard, 1980), or connectivity in discourse (Rohde & Horton, 2014).

Additionally, information provided by verbs may be used to anticipate arguments (e.g., Kamide, Altmann, & Haywood, 2003; Knoeferle et al., 2005; Kukona, Fang, Aicher, Chen, & Magnuson, 2012). In a seminal study Altmann and Kamide (1999) showed that listeners use the semantic information provided by verbs in English to anticipate the object referent before it is mentioned. In a visual world experiment, the authors showed pictures such as that of a boy, a cake, a toy car, and several other objects while auditorily presenting sentences like "The boy will eat the...". When hearing the latter kind of sentences, participants looked at the cake in the display immediately after the verb (i.e., before the cake was mentioned) as this was the only edible object in the display. While this convincingly demonstrates the predictive nature of sentence comprehension, experimentation on subject-initial languages like English and German (Kamide, Scheepers, & Altmann, 2003) does not allow us to tease apart what kind of information listeners use to form their predictions. Usually, these experiments investigate anticipation of the sentence-final syntactic object which follows the verb — conflating linear, syntactic and semantic cues, as the anticipation target in these sentences is at the same time the last word, the syntactic object and the patient of the verb.

With experiments on verb-initial languages, however, it is possible to tease these three alternatives apart because linear, syntactic, or semantic cues can be dissociated. Chapter 5 reports a visual world sentence comprehension experiment on Tagalog that exploited the language's erb-initial word order and its voice marking system to investigate whether listeners anticipate upcoming arguments based on their linear order, their syntactic function, or their semantic role.

In brief, the four studies in this thesis make use of different grammatical properties of Tagalog and German in an attempt to explore (i) how verb position and voice marking influence the time course of sentence formulation and which planning strategies speakers employ, (ii) the psychological reality of the distinction between symmetrical and asymmetrical voice systems as manifested in distinct cognitive processing load profiles of producing German and Tagalog sentences, and (iii) the influence of sentence-initial verb position on the anticipation of arguments.

1.3 READING GUIDE

The following chapters have been written in order for them to be stand-alone papers. The author thus asks the reader to forgive slight differences in writing style, terminology, and reporting of statistical results as well as some repetition in various sections of these papers.

Part II

VOICE AND WORD ORDER IN SENTENCE PRODUCTION

2

HEAD-MARKING MORPHOLOGY INFLUENCES THE TIME COURSE OF VERB-INITIAL SENTENCE PRODUCTION: EYE TRACKING EVIDENCE FROM TAGALOG

Sauppe, S., Norcliffe, E., Konopka, A. E., Van Valin, R. D., and Levinson, S. C. (2013). Dependencies first: Eye tracking evidence from sentence production in Tagalog. In Knauff, M., Pauen, M., Sebanz, N., and Wachsmuth, I., editors, *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, pages 1265–1270, Austin, TX. Cognitive Science Society.

Sauppe, S., Norcliffe, E., Konopka, A. E., Van Valin, R. D., and Levinson, S. C. (2017). Head-marking morphology influences the time course of verb-initial sentence production: Eye tracking evidence from Tagalog. Unpublished manuscript.

ABSTRACT

It is generally assumed that speakers plan their utterances incrementally and that the size of planning units is influenced by a variety of factors, including a language's grammatical structure. Two eye-tracked picture description experiments on Tagalog (Austronesian) tested whether the amount of agreement morphology carried by verbs influences the scope of initial planning units in verb-initial sentence production, as predicted by theories of hierarchical incrementality. Native Tagalog speakers described pictures of transitive events with [verb agent patient] or [verb patient agent] sentences. The verb either carried voice marking morphology (signaling whether the agent or the patient was the subject; Experiment 1) or were voiceless and thus exhibited no head-marking morphology (Experiment 2). Analyses compared the formulation of sentences differing in the kind and amount of verbal head-marking morphology as well as speakers' structural choices across events with more and less accessible characters. The results broadly support hierarchically incremental planning of verb-initial sentences, suggesting the early generation of a structural-relational utterance representation: Characters humanness and semantic role interacted in influencing speakers' voice choice and fixations at the outset of formulation differed between sentences with different voice marking and between voice-marked and voiceless sentences. Furthermore, it is discussed why linear incrementality cannot account for planning processes in verb-initial languages.

2.1 INTRODUCTION

During language production speakers transform abstract ideas and thoughts into linearly ordered strings of words that obey the grammar of their language. Models of sentence production generally assume that sentence planning processes proceed in several stages (Bock & Levelt, 1994; Levelt, 1989): the ideas speakers want to convey are encoded during message formulation, grammatical structures are built and lexical items are retrieved; finally, the product of these processes is sent on to phonological and phonetic encoding.

All models acknowledge that these processes need to result in the production of sentences with language-specific grammatical structures. However, it remains an open question how the different operations during linguistic encoding are coordinated in time and thus how the grammatical properties of individual languages may shape sentence planning.

The world's languages differ greatly in their structural properties. One dimension on which languages vary is their basic word order, i.e., the order of subjects, objects and verbs in transitive sentences. Many frequently studied European languages use SVO word order, and only a small minority of the world's languages (approximately

9%) places the verb first (Dryer, 2013). Another dimension along which languages vary is the locus of morphological marking to signal syntactic functions, e.g., the head or the dependent of a phrase or none at all (Nichols, 1986). English, for example, marks the subject function on the head of a sentence (the verb): full noun phrases do not bear case marking but the verb agrees with the subject. German, by comparison, uses both head- and dependent-marking to signal syntactic functions. The verb agrees with the subject (head-marking) and additionally the subject bears nominative case (dependent-marking). In the current paper, we explore how the formulation of simple transitive sentences is influenced by word order and by the locus of morphological marking in the verb-initial language Tagalog, an Austronesian language spoken in the Philippines. Tagalog is a typologically interesting language to study because it exhibits various sentence types that differ with respect to whether and how syntactic functions are marked both on the verbal predicate and on its arguments.

In the following, we first review theories of sentence production and explain how verb-initial languages provide a novel test case for these theories and how head-marking morphology in verb-initial languages may be hypothesized to influence the time-course of sentence formulation. We then report two eye-tracking experiments investigating the influence of grammatical structure on formulation in Tagalog. We compare structural choices and the time-course of planning for several verb-initial sentence types, varying in their overt morphological dependencies and in argument order.

2.1.1 *Theories of sentence production*

Existing theories of sentence production agree that this process proceeds incrementally, i.e., that speakers do not plan an entire sentence before they begin to speak (V. S. Ferreira & Slevc, 2007;

Kempen & Hoenkamp, 1987; Levelt, 1989). Rather, sentences are planned piecemeal, in small increments. As soon as an increment becomes available at one processing level, it is passed on to the next level, allowing speakers to simultaneously begin planning the following increment (Levelt, 1999). There are at least two accounts of how a message can be transformed into a linearized string of words, focusing mainly on differences in how speakers start planning.

2.1.1.1 *Linear incrementality*

The first account of incremental sentence planning states that formulation is linearly incremental. On this view, the scope of planning is narrow as speakers plan sentences "word by word" with little or no advance planning beyond the first content word of the sentence prior to speech onset (Bock & Ferreira, 2014; V. S. Ferreira & Slevc, 2007). Speakers begin building utterances by encoding only the first element of the message conceptually and linguistically; further message elements are then added to the unfolding sentence piecemeal, resulting in a gradual building up of the sentence and its grammatical structure. The planning units or increments may thus be as small as a single word (Brown-Schmidt & Konopka, 2008, 2015; Griffin, 2001).

For example, when a picture such as that in Figure 2.1 is described in English, speakers might begin sentence formulation by conceptually

Figure 2.1: Example stimulus picture

and linguistically encoding only the first word of their utterance (*bird*) before they start to speak, deferring the encoding of the specific action (*eat*) and further event participants (*worm*) until after speech onset. Gleitman et al. (2007) provide evidence from an eye-tracked sentence production task that English speakers do sometimes start speaking after having encoded only the first message element. Gleitman et al. manipulated the visual salience of individual characters in such events via subliminal cueing and showed that speakers were more likely to begin their picture descriptions with cued than with uncued characters, producing more active sentences when the agent was cued and more passives when the patient was cued. Eye movement patterns observed from trial onset onwards showed that participants immediately directed their gaze to the cued referent and maintained fixation on this character until speech onset. This suggests that only the first-mentioned character was encoded during early formulation whereas the encoding of the second character was deferred until later, possibly as late as after speech onset.

The perceptual salience of referents induced by a method like subliminal cueing described above (cf. also Tomlin, 1995) increased their accessibility (the ease of retrieving and encoding a referent). In spontaneous production, the accessibility of concepts and lexical items may be influenced by a variety of linguistic and extra-linguistic factors. Beyond perceptual salience, linear incrementality also emphasizes the influence of conceptual accessibility, i.e., the relative ease of encoding message elements, on word order (Iwasaki, 2011; Kempen & Hoenkamp, 1987; Levelt, 1989). Conceptual accessibility may be influenced by various features, including a referent's imageability (Bock & Warren, 1985), its givenness (Arnold, Losongco, Wasow, & Ginstrom, 2000), as well as its inherent animacy (Branigan et al., 2008; McDonald, Bock, & Kelly, 1993). Thus, if the agent in Figure 2.1 happens to be more conceptually accessible than the patient, speakers may encode this character first, leading them to produce an active sentence ("The bird eats the worm."). If, however, the patient is more accessible, this character is more likely to be encoded and produced first, resulting in the production of a passive sentence ("The worm is eaten by the bird."). In this sense, linear incrementality assumes that syntactic structures are built bottom-up via lexical retrieval, so that the retrieval of individual lexical items constrains structural processing (Bock & Ferreira, 2014).

The accessibility of nominal concepts has been shown to influence the selection of syntactic subjects in subject-initial languages as well as in languages that allow flexible word orders (e.g., Bock & Warren, 1985; Christianson & Ferreira, 2005). Additionally, evidence from languages that allow multiple word orders also shows that more accessible concepts and lexical items tend to be mentioned earlier in a sentence when holding syntactic functions constant (e.g., Branigan & Feliki, 1999; V. S. Ferreira & Yoshita, 2003).

2.1.1.2 *Hierarchical incrementality*

The second account of incrementality in message and sentence formulation is hierarchical incrementality. On this account, speakers start planning sentences by generating a structural-relational representation of the whole message (i.e., a representation of the relations among event participants and their syntactic functions). It is this message-derived representation that then guides linguistic encoding, i.e. the selection of grammatical structures and lexical retrieval. In the current example (Figure 2.1), this implies that speakers first encode that an agent (*bird*) is acting upon a patient (*worm*) in a physical contact event (*eat* or *catch*). This results in encoding of a representation in which the event characters are linked to specific event roles and in which the concepts of *bird* and *worm* are mapped to the syntactic functions subject and object, respectively. Speakers then begin retrieving the names of the two characters in the order required by the selected structure. Thus, unlike linear incrementality where lexical retrieval guides sentence planning, hierarchical incrementality posits that the formulation of sentences proceeds "word by word" only at the level of linguistic encoding, as the retrieval of individual words is guided by a larger message-level representation that has a scope encompassing several words (Bock & Ferreira, 2014; Bock et al., 2003; Konopka & Meyer, 2014; Lee et al., 2013).

Griffin and Bock (2000) were the first to provide evidence of sentence formulation being guided by a relational structure. In an eye-tracked sentence production experiment in which English speakers described line drawings of two-participant events. Griffin and Bock found that speakers distributed their attention roughly evenly among the event characters until 400 ms after picture onset. Speakers then directed their gaze preferentially to the character that was mentioned first, followed by a gaze shift to the second character approximately around speech onset. Griffin and Bock interpret this early gaze pattern (0–400 ms) as evidence of an initial gist apprehension phase during which speakers encode the relationship between event participants before engaging in any linguistic encoding. This information allows them to build a structural-relational representation of the message and the sentence, and then proceed to fill it in with lexical material after 400 ms. Thus, hierarchical incrementality is compatible with sentence production theories that allow structural planning processes to proceed independently of lexical retrieval (e.g., Chang et al., 2006; Dell, 1986; Konopka & Meyer, 2014).

2.1.1.3 *Variability in planning scope and planning in verb-initial languages*

In sum, the accessibility of individual message elements and the early generation of structural-relational representations play different roles in linear and hierarchical accounts of incremental sentence production because they assume different starting points for planning. Although a variety of evidence has been brought forward for both of these accounts, there is little consensus as to whether formulation primarily favors the early encoding of individual lexical items or the encoding of the relations among message elements. Furthermore, the scope of planning may be influenced by intra-linguistic factors (Konopka, 2012; Konopka & Meyer, 2014; van de Velde et al., 2014) or extra-linguistic factors (F. Ferreira & Swets, 2002; Swets et al., 2014), affecting whether planning proceeds with a broader or narrower scope and suggesting that both hierarchically and lexically incremental planning strategies may be used by speakers in different production contexts.

An additional, vastly under-researched question is to what extent formulation processes are influenced by the grammar of the target language. Fortunately, there is now a growing body of cross-linguistically informed studies that make use of the "natural laboratory" (Evans & Levinson, 2009) provided by the diversity of the world's languages (cf. Christianson & Ferreira, 2005; Jaeger & Norcliffe, 2009; Norcliffe, Harris, & Jaeger, 2015), extending the empirical basis of research on sentence production.

Verb-initial languages are particularly interesting among the lesser-studied languages in this regard because they pose a theoretical challenge for existing linearly incremental sentence production accounts. Linear incrementality was developed based on evidence from a very small set of languages with subject-initial basic word order, such as English. In these languages, "word by word" linearly incremental planning means that speakers can start sentence formulation by encoding the most accessible nominal concept in the message (e.g., either the *bird* or the *worm* in Figure 2.1), which then becomes the syntactic subject of the sentence and is produced first, without the explicit need to encode other nominal concepts or any relational information at this point in the formulation process (cf. Gleitman et al., 2007). In verb-initial languages, by contrast, accessible nominal concepts cannot be immediately placed in sentence-initial position because the first word of the sentence is the verb. Further, verbs encode the relationships between referents rather than describing the referents themselves (like nominal concepts). Therefore, the early position of the verb in verb-initial languages implies that speakers must build a representation of the entire event, including encoding of the action and assignment of syntactic functions to event characters, early on in order to be able to select a suitable verb and to plan its morphological marking. Thus, while hierarchical incrementality can be easily applied to verb-initial sentence planning because it generally assumes that speakers engage in early structural-relational planning, linear incrementality runs into difficulties *a priori* by assuming the early encoding of accessible nominal concepts.

However, if we accept that building up the sentence structure linearly "word by word" could also mean that the verb is encoded first, crucially involving the encoding of the relations between event participants, linear incrementality could also be applied to verb-initial languages. The consequence of "blurring the line" between early encoding of nominal message elements and early encoding of relational message elements is that linearly incremental and hierarchically incremental accounts make the same predictions for verb-initial sentence production — namely, that speakers first generate a relational representation of the event which allows them to select a verb to begin their utterance with.

Evidence from online planning of verb-initial sentences is to the best of our knowledge limited to one eye-tracking study carried out on Tzeltal, a Mayan language spoken in Mexico. In Tzeltal, both subject-initial and verb-initial word orders are possible (e.g., *The bird eats the worm* and *Eats the worm the bird*, respectively) and transitive verbs agree with both of their arguments. Comparing speakers' eye movements during the production of both sentences types, Norcliffe, Konopka, et al. (2015) showed that speakers' uptake of visual information is influenced by word order from the outset of planning: formulation of verb-initial sentences (VOS) began with a window during which speakers distributed their attention between the subject and object characters (0–600 ms after stimulus picture onset), suggesting early planning of relational information. When producing subject-initial sentences, on the other hand, Tzeltal speakers showed the same pattern as speakers of English and Dutch producing subject-initial (SVO) sentences: they quickly fixated the subject character after picture onset and maintained fixations on this character until speech onset, at which point they shifted their gaze to the second character. These results show that the time-course of sentence planning may vary with word order even within a language and is thus largely dependent on sentence structure: when the verb is produced first, relational information is encoded early, and when the sentence subject is produced first, the subject character is encoded with priority early in the formulation process. Therefore, Norcliffe, Konopka, et al.'s (2015) results support the view that planning verb-initial sentences requires the early generation of a relational event representation, whereas this may not be the case for subject-initial sentences.

At the same time, this research leaves open a number of questions. First, our investigation of planning in the verb-initial Austronesian language Tagalog will provide the first general replication of the Tzeltal results, thus providing an opportunity to verify whether Norcliffe, Konopka, et al.'s conclusions generalize to other verb-initial languages. Second, and more importantly, if verb-initial sentence production does indeed involve hierarchically incremental planning, it is important to examine whether and how grammatical differences between languages may plausibly contribute to cross-linguistic variations in hierarchical planning. In the same way that there is some variability in the degree to which planning of any given sentence can be linearly and hierarchically incremental in a language like English or Dutch, grammatical differences between individual verb-initial languages may support different degrees of hierarchical sentence planning. If this proposal is correct, sentence planning processes should follow different time courses, depending on the amount of relational planning at the outset of formulation is required by the make-up of a language's grammar. Besides sentence-initial verb placement, the grammatical properties most relevant for assessing such time course variations in planning concern head- and dependent-marking morphology.

Here, we test whether and how the amount of agreement morphology that speakers have to plan might influence the time course of sentence planning. In Tzeltal, verbs are morphologically quite complex because they always agree with both subject and object. As Norcliffe, Konopka, et al. (2015) suggest, this, together with early verb placement, could necessitate a high degree of relational planning at the outset of formulation in Tzeltal. Other verb-initial languages, such as Tagalog, exhibit less complex agreement marking on the verb. Head-marking morphology on the sentence-initial verb in Tagalog is much simpler: either the verb agrees with one of its arguments or it exhibits no morphological dependencies.

In the following, we report two eye-tracked picture description experiments on Tagalog, investigating how the presence or absence of head-marking morphology influences the encoding of events in order to plan sentence-initial verbs and post-verbal arguments. Given Norcliffe, Konopka, et al.'s (2015) results from Tzeltal, our starting point is the assumption that speakers begin with the generation of a structural-relational representation in verb-initial sentence production, and our experiments will help us to assess how the degree of advance planning is influenced by the presence and type of head-marking morphology. While early relational encoding appears to be necessary to enable speakers to produce a verb first, the amount and type of information about the verb's arguments that needs to be encoded might have an effect on how much grammatical information is planned at the outset of formulation. Hence, we investigate the planning of two verb-initial Tagalog sentence types, one in which the verb agrees with one of its arguments and one in which the verb does not exhibit morphological dependencies but takes an invariant shape.

Next, we describe the most relevant structural properties of Tagalog for the current studies and lay out the hypotheses for the two experiments.

2.1.2 *Tagalog*

The grammar of Tagalog is in many respects very different from that of the European languages that are often studied in psycholinguistics. We therefore introduce some terminological conventions that allow us to acknowledge these differences. Firstly, to circumvent the discussion of lexical categories (noun/verb distinction) in Tagalog (Himmelmann, 2008), we use the term predicate to refer to words that carry aspect (as well as voice and mood) marking and the term argument to refer to case-marked phrases. Secondly, instead of using the terms *subject* and *object* to refer to a predicate's arguments, we use the terms *pivot* and *non-pivot*, respectively, because the notions of subject and object cannot be straightforwardly applied to languages that exhibit so-called symmetrical voice systems like Tagalog (Foley, 2008; Riesberg, 2014b; Chapter 4). In English, by comparison, the syntactic subject can be considered the pivot. We understand pivot and non-pivot to be purely syntactic terms and distinguish them from the terms *agent* and *patient*, which allows us to separate semantic roles and syntactic functions.

Basic declarative sentences in Tagalog are predicate-initial. In sentences in perfective and imperfective aspect, one argument of the predicate is marked by *ang*, which will be referred to as the pivot argument. Voice affixes that signal this argument's semantic role are attached to the predicate.[1] Thus, there is a morphosyntactic dependency between the predicate and the pivot argument (Himmelmann, 2002; Nichols, 1986) that is double-marked: the relationship is dependent-marked by *ang* on the pivot argument and head-marked by the voice affix on the predicate. Other arguments that do not have their semantic role indicated at the predicate will be referred to as non-pivot arguments when they are marked by *ng*.

The examples in (2.1)[2] show how a transitive, two-participant event in Figure 2.1 can be described in Tagalog, illustrating the grammatical properties that are relevant for the experiments reported here. Pivot markers and the corresponding voice affixes are printed in boldface.

(2.1)	a.	k<**um**>akain	ng=uod	**ang**=ibon
		<AV>eat	NPVT=worm	PVT=bird
	b.	k<**um**>akain	**ang**=ibon	ng=uod
		<AV>eat	PVT=bird	NPVT=worm
	c.	k<**in**>akain	ng=ibon	**ang**=uod
		<PV>eat	NPVT=bird	PVT=worm
	d.	k<**in**>akain	**ang**=uod	ng=ibon
		<PV>eat	NPVT=worm	PVT=bird
		"The bird eats a/the worm."[3]		

These examples show the double-marking nature of the dependency between the sentence-initial predicate and the pivot argument in perfective and imperfective sentences: While the pivot status of

[1]Pivot arguments are also privileged in a number of syntactic constructions (Kroeger, 1993b; Schachter, 1995a).

[2]The following abbreviations are used here: A = agent, AV = agent voice, IPFV = imperfective aspect, NPVT = non-pivot argument, OBL = oblique argument, P = patient, Pred = predicate, PV = patient voice, PVT = pivot argument, RDP = reduplication, RP = recent perfective aspect. The first line of a glossed example shows the sentence in Tagalog with the relevant morphemes separated, the second line provides a morpheme-by-morpheme translation, and the last line gives a translation of the whole sentence.

[3]The patient is translated as definite or indefinite, depending on the voice marking which puts constraints on interpreting the specificity of *ang*-marked arguments (Adams & Manaster-Ramer, 1988; Latrouite, 2015).

an argument is indicated by ang-marking, the predicate signals that argument's semantic role. In (2.1a) the agent of the event (the bird) functions as the pivot and is marked by *ang*, so the predicate thus takes the agent voice affix <*um*> that signals the semantic role of the pivot is agent. In (2.1b), on the other hand, the patient (the worm) is the pivot argument, which is marked by *ang*; the predicate thus takes the patient voice affix <*in*> that signals that the semantic role of the pivot is patient.[4] When the patient is human, it must be selected as pivot (Latrouite, 2011). The semantic roles of non-pivot arguments (marked by *ng*) are not signaled by the predicate. We focus on the implications that this agreement system has for online sentence planning in Experiment 1.

There are no syntactic constraints on the order of arguments following the predicate (although canonically the pivot argument is positioned sentence-finally; Schachter & Otanes, 1972). In (2.1a,c) the pivot is sentence-final, whereas in (2.1b,d) it is in sentence-medial position. Thus [Pred_{AV} P_{NPVT} A_{PVT}] word order is the canonical order for agent voice sentences and [Pred_{PV} A_{NPVT} P_{PVT}] word order is canonical for patient voice sentences. Also, all sentence types in (2.1) are equally transitive, all of them have one pivot and one non-pivot argument (Riesberg, 2014b). In English, by contrast, actives and passives differ in transitivity with passives being syntactically intransitive. This property of Tagalog will allow us to compare sentences with the same transitivity where only the semantic role of the pivot argument differs.

In addition, Tagalog also exhibits a voiceless construction in which none of the arguments functions as pivot, i.e., neither argument is marked by *ang* nor does the predicate signal the semantic role of any

[4] Tagalog also exhibits a variety of other voices in which the location of an event, the instrument or the beneficient function as pivot arguments. Together with patient voice, these other voices are often subsumed under the label "undergoer voice" because they share a number of semantic and formal characteristics (Himmelmann, 2005b). In the current paper, however, we will only cover patient voice and agent voice, the two most frequent voices.

argument. When the predicate carries recent perfective aspect marking to describe events that just finished (ex:RP-ETTagalog), both agent and patient arguments receive default case marking: agents and indefinite patients are marked as non-pivots by *ng* and definite patients are marked as obliques by *sa* (Kroeger, 1993b; Schachter & Otanes, 1972). As in perfective and imperfective sentences with voice marking (2.1), the predicate is the first word in the sentence (always followed by the adverb *(pa)lang* 'just') and the order of its arguments is free. The canonical order, however, is [Pred_{RP} A P]. We investigate the time-course of formulation in voiceless recent perfective sentences in Experiment 2.

(2.2)	a.	ka-ka∼kain	lang	ng=ibon	ng/sa=uod
		RP-RDP∼eat	just	NPVT=bird	NPVT/OBL=worm
	b.	ka-ka∼kain	lang	ng/sa=uod	ng=ibon
		RP-RDP∼eat	just	NPVT/OBL=worm	NPVT=bird

"A/the bird just ate a/the worm."

The difference between voice-marked perfective and imperfective sentences on the one hand and recent perfective sentences on the other hand is thus whether there is an overtly marked dependency between the sentence-initial predicate and one of its arguments. While the predicate in voice-marked sentences exhibits head-marking morphology indexing the pivot's semantic role, predicates in recent perfective sentences take an invariant shape that does not depend on any of its arguments.

2.1.3 *Current experiments*

In two experiments, Tagalog speakers performed an eye-tracked picture description task, similar to that of previous studies (Griffin & Bock, 2000; Konopka & Meyer, 2014; Norcliffe, Konopka, et al., 2015). Target stimuli were colored line drawings of transitive events (cf. Figure 2.1). In the first experiment, speakers were instructed to describe pictures using voice-marked sentences with overt morphological dependencies between the predicate and the pivot argument. In the second experiment, speakers were instructed to produce recent perfective sentences that did not exhibit head-marking of the predicate.

These two experiments exploit the language-internal contrasts described above, which allow us to investigate three issues about the production of verb-initial sentences. First, the different orders of agent and patient after the predicate that are possible in both voice-marked (2.1) and recent perfective sentences (2.2) allow us to investigate the influence of character-specific variables (like accessibility) on argument order. Second, the different morphological markings for agent voice and patient voice in imperfective and perfective sentences make it possible to test predictions from hierarchical incrementality about the scope of initial planning. On this view, speakers should encode the event relations as well as all information that is necessary for the production of voice morphology at the outset of formulation. In other words, early eye movements across sentences should differ across sentence types whenever head-marking differs, reflecting the specific grammatical requirements of each sentence type. Finally, the voiceless predicates in recent perfective sentences allow us to investigate how the presence or absence of head-marking morphology influences speakers' early structural-relational planning and the degree of advance planning.

In the following, we outline the predictions made by hierarchical incrementality for the production of predicate-initial Tagalog sentences and report the results of two experiments testing these hypotheses.

2.1.3.1 *Predictions for the influence of accessibility on structural choices*

In subject-initial languages, accessible characters are preferentially realized as syntactic subjects (Bock & Warren, 1985; Branigan et al., 2008). However, accessibility effects that have been observed in subject-initial languages often necessarily conflate the influence of accessibility on word order and its influences on syntactic function assignment. The variability in the ordering of agent and patient after the predicate that is possible in Tagalog in voice-marked imperfective and perfective sentences (2.1) and also in voiceless recent perfective sentences (2.2) allows us to investigate whether accessibility influences pivot selection, word order, or both (cf. Tanaka, Branigan, McLean, & Pickering, 2011). Notably, however, positioning an accessible character "early" in Tagalog means that it is mentioned sentence-medially (after the predicate, preceding the second character).

The accessibility of a character can be influenced by perceptual and conceptual factors (among others). Accessibility effects on pivot selection will be tested for in Experiment 1 where Tagalog speakers produced voice-marked sentences; accessibility effects on the order of agent and patient after the predicate will be tested in both Experiments 1 and 2 (voiceless recent perfective sentences).

PIVOT SELECTION In a picture description task, some depicted characters may be perceptually more salient than others, leading speakers to fixate on them first (these first fixations to characters are expected within the first 200 ms after stimulus onset; Allopenna, Magnu, & Tanenhaus, 1998; Duchowski, 2007). In Gleitman et al.'s (2007) experiments, first-fixated characters were more likely to be selected as subjects; Myachykov, Thompson, Garrod, and Scheepers (2012) report similar results. In subject-initial languages, these findings support accounts of linearly incremental planning, where perceptually accessible characters are more likely to be lexically encoded first and thereby placed in the sentence-initial subject position. In Tagalog, first fixations could influence pivot selection so that speakers would be more likely to produce agent voice sentences when they fixate the agent first or patient voice when they fixate the patient first because perceptually accessible characters should be assigned to the highest syntactic function. However, hierarchical incrementality does not predict such an effect because the assignment of syntactic functions is assumed to be based on a relational event representation and is thus not driven by the early availability of individual nominal concepts (Bock et al., 2003).

A second dimension along which characters can differ is conceptual accessibility. In subject-initial languages, animates and humans are more likely to be selected as subjects (Branigan et al., 2008). In Tagalog, an influence of conceptual accessibility on voice choice would indicate a high degree of structural advance planning. Given that the pivot does not come first in the sentence, speakers would have to determine the characters' animacy and use this information to assign syntactic functions well in advance of mentioning them in either sentence-medial or sentence-final position. Notably, in Tagalog, there is in fact a grammatical constraint that can enforce such a process: if there is a human patient in the event, it must be selected as the pivot (i.e., the event must be described in patient voice; Latrouite, 2011). This suggests *a priori* that speakers of Tagalog might engage in broad structural-relational planning at the outset of formulation in order to satisfy this grammatical requirement. We can, however, also ask to what extent conceptual accessibility may influence structural choices in Tagalog beyond this grammatical constraint. In cases where the patient is not human, is there any independent effect of conceptual accessibility on voice choice?

WORD ORDER We also investigate how perceptual and conceptual accessibility may affect the order in which characters are mentioned in sentences with and without head-marking morphology (Experiments 1 and 2, respectively). It is worthwhile to test for accessibility effects on word order in addition to voice choice because in Tagalog word order is in principle independent from voice marking, and because both [agent patient] and [patient agent] orders are possible with agent voice, patient voice and recent perfective sentences (although these sentence types have canonical word orders). It is thus possible to test whether accessibility affects word order independently of pivot selection in voice-marked sentences.

An effect of perceptual accessibility (e.g., first-fixated characters being mentioned sentence-medially rather than sentence-finally) would suggest that speakers' structural choices can be driven by the early visual salience of agent or patient characters. We hypothesize that the early placement of first-fixated characters in a sentence would mean that linguistic encoding of these characters starts early in the formulation process, potentially even before structural-relational encoding is completed. Likewise, conceptual accessibility may influence word order via a similar mechanism. There could be an influence of character humanness on post-predicate word order, such that human characters are more likely to be mentioned sentence-medially than sentence-finally because they are easier to encode linguistically and therefore might be processed earlier.

Importantly, the voice of a Tagalog sentence determines its canonical word order: pivot arguments are canonically and most frequently positioned sentence-finally. Therefore, any influences of the different kinds of accessibility on word order must persist even when the canonical ordering of arguments is controlled for.

TIME COURSE OF FORMULATION We analyze eye movements during the production of Tagalog sentences to examine what speakers attend to and what they encode with priority at different points in time when preparing their utterances. We first focus on eye movements during early time windows (0–400 ms, 0–600 ms) because they can provide critical insight into the way speakers begin formulation (Griffin & Bock, 2000).

In the Experiment 1, speakers described pictures with voice-marked sentences (2.1). We compare the distribution of fixations to agent characters in sentence types with different voice marking and different orders of agents and patients. We hypothesize that, in line with hierarchical incrementality, Tagalog speakers first generate a structural-relational representation of the event and the utterance (roughly between picture onset and 400–600 ms), which subsequently guides lexical retrieval (after 400–600 ms). To do this, speakers need to encode relational information very early in order to select a predicate (Norcliffe, Konopka, et al., 2015); Dobel, Glanemann, Kreysa, Zwitserlood, and Eisenbeiß (2011) show that this kind of information can be extracted within 100 ms of stimulus picture onset (cf. also Hafri, Papafragou, & Trueswell, 2013; Potter, 1976). By hypothesis, because the predicate carries a voice affix that signals the pivot argument's semantic role, speakers will also need to select one event participant to be the pivot at this point in time.

Thus, if Tagalog speakers first encode relational and structural information, they should preferentially fixate the character that will become the pivot argument shortly after picture onset. This would indicate the generation of a representation of the event and the utterance in which one of the characters is quickly assigned to the pivot function. If lexical retrieval is then guided by this structural-relational representation (Kuchinsky et al., 2011), speakers should fixate the characters in their order of mention after 400–600 ms, i.e., first the agent and then the patient when the word order after the predicate is [A P], as in (2.1b,c), and vice versa when it is [P A], as in (2.1a,d). The transition from initially fixating the pivot character to fixating the first-mentioned character should take place as soon as structural planning is completed (Griffin & Bock, 2000). Fixations to the pivot character in this early time window are not expected to differ between sentences with different word orders because speakers need to plan the predicate's voice marking irrespectively of whether the agent or the patient will be mentioned first.

In the Experiment 2, speakers described sentences in recent perfective aspect, where the predicate lacked head-marking morphology (2.2). Here, we also predict that speakers first generate a structural-relational representation at the outset of formulation (between stimulus picture onset and approximately 400–600 ms). Formulation of recent perfective sentences with [Pred A P] order and with [Pred P A] order should result in similar eye movement patterns after picture onset (reflecting relational encoding) as the predicate takes the same form in both of these sentence types. We again predict that the structural-relational representation guides lexical retrieval after 400–600 ms, leading speakers to fixate characters in the order of mention.

Finally, to explore how the time course of sentence planning is influenced by the head-marking morphology on voice-marked predicates in Tagalog, we run a joint analysis on sentences from Experiments 1 and 2. We compare speakers' fixations during the production of agent voice sentences from (Experiment 1) and recent perfective voice sentences from (Experiment 2) with the same [A P] order after the predicate. These sentence types only differ in whether or not the predicate exhibits a morphological dependency with the agent argument. We predict that the need to assign the pivot function to the agent leads to more fixations to that character early in the formulation of voice-marked sentences (2.1) as compared to the formulation of recent perfective sentences (2.2) in which the predicate's morphological form does not depend on either argument. In other words, speakers' visual uptake of information should be influenced in the earliest phase of planning (within 400–600 ms) by the requirement to select a pivot and plan a voice affix in imperfective and perfective sentences.

2.2 EXPERIMENT I

In the first experiment, Tagalog speakers described transitive events using predicate-initial and voice-marked sentences in imperfective or perfective aspect. We examine the influence of different kinds of accessibility on pivot selection and on post-predicate word order, and how differences in voice marking influence the time course of sentence formulation.

2.2.1 *Method*

2.2.1.1 *Participants*

Fifty-three native speakers of Tagalog (13 males, mean age = 17.5 years, age range = 15–28 years) participated for payment. All of them reported speaking Tagalog for at least five hours per day and to at least one of their parents. All participants had normal or corrected-to-normal vision.

2.2.1.2 *Materials and design*

Stimuli were colored line drawings of transitive and intransitive events, including 44 pictures of transitive events that were used as targets (cf. Figure 2.1 for an example and Appendix 2.A for a full list of depicted events). Based on a norming study conducted with 20 Tagalog speakers (who did not participate in the main experiments reported here), we selected 44 target pictures that elicited descriptions with agent voice, patient voice, and pictures that showed a similar bias for agent and patient voice (20 pictures with an agent voice preference, 19 pictures with a patient voice preference, and 5 pictures with no preference). The target pictures were interspersed among 76 filler pictures of intransitive events and at least one filler intervened between any two targets. Two versions of each target picture were created by mirror-reversing the pictures. All the stimulus pictures were arranged in four lists by creating two pseudo-randomized orders of targets and fillers that were crossed with the two mirror-reversed versions of each target to avoid effects due to the order of stimuli and the position of characters in the pictures. Participants' eye movements were recorded with a Tobii T120 eye tracker (120 Hz sampling rate). The picture descriptions were audio-recorded together with the eye movement data by the Tobii Studio software.

2.2.1.3 *Procedure*

Participants first completed a questionnaire about their linguistic background and then read the instructions in Tagalog. The experimenter (a native speaker of Tagalog) also repeated the instructions verbally. The participants' task was to describe the events shown in the pictures in one predicate-initial sentence, naming all event participants, as accurately and quickly as possible. Stimuli were presented in two blocks lasting 10–15 minutes each. Eye-tracker calibration was performed before each block.

The experiment began with a practice phase to familiarize participants with the task. Participants saw eleven example pictures accompanied by pre-recorded descriptions presented one at a time on the screen. Next, participants saw the same pictures again one at a time and were asked to describe them themselves. The experimenter provided feedback if the participant used non-predicate-initial structures (e.g., existential constructions) or started speaking very late after picture onset. On each trial, the stimulus was preceded by a fixation dot displayed randomly in one of five positions at the top of the screen.

Participants were debriefed after the experiment. The experimental session lasted approximately 40 minutes.

2.2.1.4 *Sentence scoring*

Utterances produced on target trials were transcribed and scored as agent voice or patient voice sentences or as other constructions (e.g., existentials and coordinated sentences). Only agent voice or patient voice sentences with both arguments realized overtly were included in the analyses. Trials containing pauses or other disfluencies were included because these are normal in spontaneous speech production; however, trials in which participants corrected themselves or restarted their sentences were excluded. Trials were also excluded when the first fixation in that trial fell on the agent or patient character instead of the fixation dot at the top of the screen, if the first fixation to the agent or patient occurred later than 800 ms after picture onset, and in cases of track loss (trials where two consecutive fixations were longer than 600 ms apart). Additionally, trials were excluded if the speech onset was longer than 3 standard deviations away from a participant's grand mean. The final data set consisted of 1266 sentences.

2.2.1.5 *Analyses*

Analyses were carried out in R using (generalized) linear mixed effects regression (Bates, Mächler, Bolker, & Walker, 2015; Pinheiro & Bates, 2000; R Core Team, 2015). Categorical predictors were deviation-coded (-0.5 and 0.5); Time as continuous predictors was mean-centered. Statistical significance was assessed by Wald Z tests (Agresti, 2007; Jaeger, 2008). Parsimonious random effects structures were determined following the procedure proposed by (Bates, Kliegl, Vasishth, & Baayen, 2015). More information about the individual models is provided below. Graphs were produced using the ggplot2 and plotrix packages (Lemon, 2006; Wickham, 2009).

2.2.2 *Results*

2.2.2.1 *Distribution of responses*

Participants produced 402 sentences with agent voice marking and [Pred_{AV} P_{NPVT} A_{PVT}] word order (2.1a), 63 sentences with agent voice marking and [Pred_{AV} A_{PVT} P_{NPVT}] word order (2.1b), 787 sentences with patient voice marking and [Pred_{PV} A_{NPVT} P_{PVT}] word order (2.1c), and 14 sentences with patient voice marking and [Pred_{PV} P_{PVT} A_{NPVT}] word order (2.1d). Analyses of the time course of fixations and eye-voice spans were limited to the first three sentence types (2.1a–c). In 58.6% of the trials, speakers first fixated the agent character.

2.2.2.2 *Pivot selection*

The first analysis compared the effects of perceptual and conceptual accessibility on pivot selection (which determines whether the predicate carries agent voice or patient voice marking) in a binomial mixed effects regression. To assess the influence of perceptual accessibility, the identity of the first fixated character (agent or patient) was included as a predictor. Conceptual accessibility was operationalized as the humanness of agent characters. Patient voice was produced for all pictures with a human patient, following the grammatical constraint that demands human patients to be selected as pivots (Latrouite, 2011). This suggests that speakers always verified whether the patient was human or not before selecting the pivot. Consequently, the regression model was restricted to target pictures with non-human patients (Table 2.1, 35 pictures).

Numerically, speakers produced more patient voice sentences when both agent and patient were non-human compared to when the agent was human and the patient was non-human (Figure 2.2); however, this effect was not statistically significant (Table 2.1). First

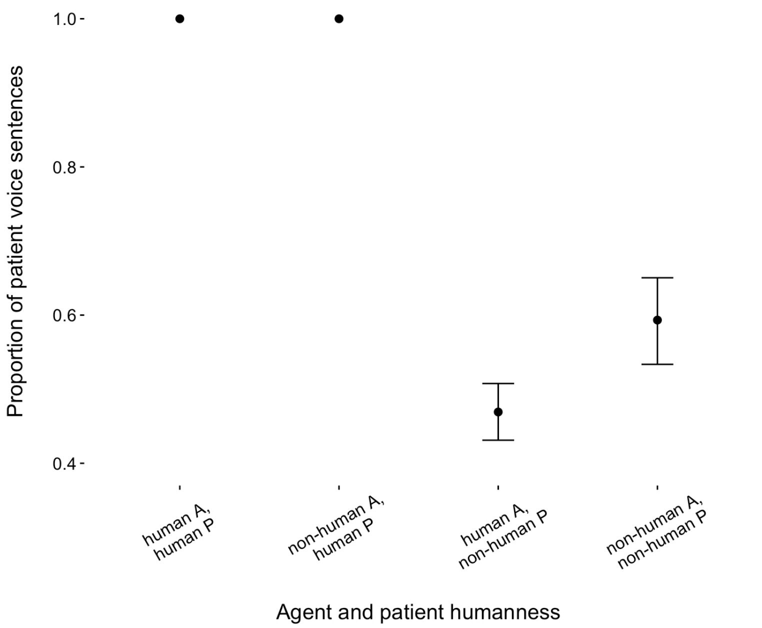

Figure 2.2: Proportions of patient voice sentences produced in Experiment 1 as a function of agent (A) and patient (P) humanness.

Table 2.1: Binomial mixed effects regression results for the effects of perceptual and conceptual accessibility on voice and pivot selection (only stimulus pictures with non-human patients).

	$\hat{\beta}$	$\|Z\|$	95% Wald CI
Intercept	-0.88	0.81	[-2.86, 1.10]
First-fixated Character (= patient)	0.47	1.05	[-1.23, 4.43]
Agent Humanness (= non-human)	1.85	1.17	[-0.42, 1.28]

Note: response = patient pivot / patient voice fixation location did also not influence speakers' voice choice, suggesting that the order in which agents and patients were fixated (reflecting the perceptual properties of the characters) did not influence planning directly.

2.2.2.3 *Word order*

The second analysis tested whether perceptual and conceptual accessibility influenced the order of agent and patient after the predicate above and beyond the effect of voice marking (Table 2.2). The predicate's voice marking was included as a control predictor because voice marking determined the canonical, pivot-final word order. For agent voice sentences, the canonical order was [Pred$_{AV}$ P A$_{PVT}$], whereas for patient voice sentences it was [Pred$_{PV}$ A P$_{PVT}$]. From grammatical descriptions of Tagalog (e.g., Himmelmann, 2005b) it was expected that speakers would produce sentences with the canonical word order more often than sentences with the non-canonical order.

Voice was the only significant predictor of word order (main effect of Voice), reflecting the fact that speakers indeed produced mainly agent-final sentences when the predicate carried agent voice marking and patient-final sentences when it carried patient voice marking (Figure 2.3).

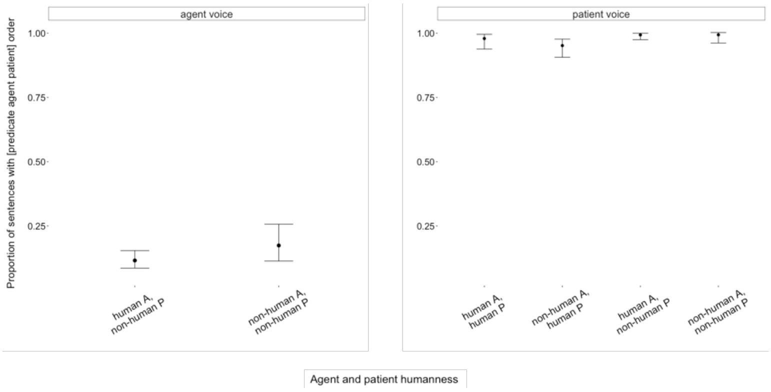

Figure 2.3: Proportions of [Pred A P] word order produced in Experiment I as a function of voice marking and agent (A) and patient (P) humanness.

Table 2.2: Binomial mixed effects regression results for the effects of voice marking and perceptual and conceptual accessibility on the order of agent and patient after imperfective and perfective predicates.

	$\hat{\beta}$	$\|Z\|$	95% Wald CI	
Intercept	-3.09	5.83	[-4.13, -2.05]	***
Voice (= agent voice)	8.98	7.29	[6.57, 11.39]	***
First-fixated Character (= patient)	0.36	1.12	[-0.27, 0.99]	
Agent Humanness (= non-human)	-0.37	0.64	[-1.51, 0.77]	
Patient Humanness (= non-human)	-2.36	1.46	[-5.51, 0.80]	
Agent Humanness × Patient Humanness	-1.13	0.56	[-5.51, 0.80]	

Note: response = [Pred P A] word order; *** $p < 0.001$

2.2.2.4 *Time course of sentence formulation*

The likelihood of fixations to agent characters over patient characters was compared between voice-marked sentence types using binomial linear mixed effects regression (Donnelly & Verkuilen, 2017). In order to reduce the temporal autocorrelation between eye tracking samples (cf. Barr, 2008), fixations were aggregated into 100 ms time bins for each trial and the number of samples with fixations to the agent and to the patient character in the respective previous time bin were also included as a nuisance variables. We selected three time windows for comparison: 100–600 ms, 600–1700 ms, and 1700–2700 ms after picture onset, respectively. The first time window, 100–600 ms, is based on Griffin and Bock's (2000) estimate of event apprehension of 0–400 ms. In their study, this estimate was based on the observation that speakers began fixating characters in the order of mention after 400 ms, which was interpreted as the onset of linguistic encoding. We follow Norcliffe, Konopka, et al. (2015) in extending this window to 600 ms, as visual inspection of the fixation patterns in our dataset suggests that speakers largely began fixating the characters in the order of mention after 600 ms.[5] We did not include fixations during the first 100 ms because it is unlikely that participants moved their eyes in response to stimulus presentation so quickly (Duchowski, 2007).

We chose a second analysis time window of 600–1700 ms, ranging from the end of the first time window until approximately speech onset (grand mean = 1694 ms). Finally, the last analysis time window ranged from 1700–2700 ms, i.e., one second after speech onset. We hypothesize that speakers primarily engage in linguistic encoding of the first character during the 600–1700 ms time window (following event apprehension) and the second character during the 1700–2700 ms time window. In fact, during the former time window

[5] Similar results for all comparisons are found with a 100–400 ms time window.

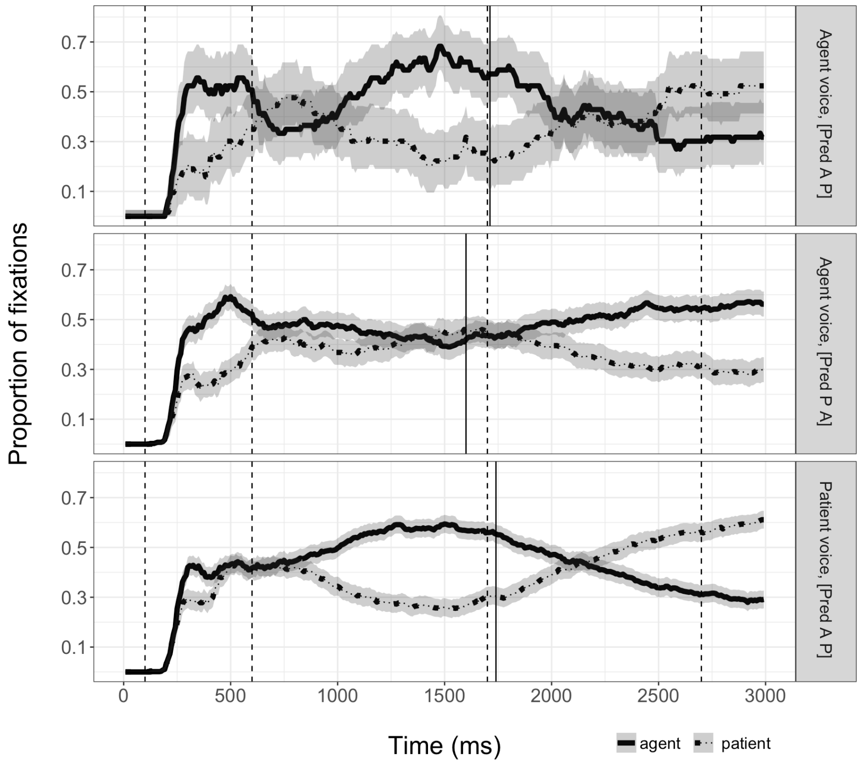

Figure 2.4: Mean proportions of fixations to agents and patients in imperfective and perfective sentence types with different voice marking and word orders; ribbons indicate 95% confidence intervals (Villacorta, 2012); solid vertical lines indicate mean speech onsets; dashed vertical lines mark analysis time windows.

speakers preferentially fixated the character that was mentioned immediately after the predicate; this was independent of its syntactic function (pivot or non-pivot) or its semantic role (agent or patient). After speech onset, during the 1700–2700 ms time window, speakers began to shift their attention to the sentence-final character.

The binomial logistic regression models for each analysis time tested whether and how voice marking and word order predict fixations to agent or patient characters, allowing us to explore what the source of potential differences in speakers' eye movements during the production of voice-marked Tagalog sentences was. Specifically, we tested whether these differences were driven by differences in voice marking (as predicted by hierarchical incrementality) or by the order of agent and patient arguments after the predicate. The results are shown in Table 2.3.

During the first time window (100–600 ms), agent characters were fixated more than patient characters overall (Intercept). In addition, speakers were more likely to fixate the agent when it was the pivot argument, i.e., when the predicate was produced with morphological agent voice marking, and were more likely to fixate the patient in patient voice sentences, i.e., when the patient was the pivot argument (main effect of Voice). The order of agent and patient after the predicate, by contrast, did not influence fixation likelihoods in this early time window.

In the second time window (600–1700 ms), speakers' fixations were contingent on word order; they fixated the character that was to be mentioned immediately after the predicate. In sentences with [Pred A P] word order, speakers turned their visual attention towards agents and in sentences with [Pred P A] word order, they gazed at the patient (interaction of Time and Word Order). Finally, in the third analysis time window (1700–2700 ms), the order in which agent and patient characters were mentioned also determined speakers' fixation behavior. After having fixated preferentially on the sentence-medial

Table 2.3: Binomial mixed effects regression results modelling the influence of voice marking and word order on the time course of fixations towards agent characters vs. patient characters during the formulation of agent voice and patient voice sentences.

	$\hat{\beta}$	$\|Z\|$	95% Wald CI	
100–600 ms				
Intercept	1.99	3.12	[0.74, 3.23]	*
Time	-0.12	0.19	[-1.39, 1.14]	
Voice (= patient voice)	-1.38	2.50	[-2.46, -0.30]	*
Time × Voice	-0.02	0.03	[-1.26, 1.22]	
Word Order (= Pred P A)	-0.81	1.39	[-1.95, 0.33]	
Time × Word Order	-0.21	0.38	[-1.26, 0.85]	
A Fixations in Previous Time Bin	0.25	40.68	[0.24, 0.26]	***
P Fixations in Previous Time Bin	-0.22	32.40	[-0.23, -0.20]	***
600–1700 ms				
Intercept	0.58	1.85	[-0.03, 1.19]	
Time	0.21	2.30	[0.03, 0.39]	*
Voice (= patient voice)	-0.41	1.17	[-1.09, 0.28]	
Time × Voice	-0.09	0.88	[-0.29, 0.11]	
Word Order (= Pred P A)	-0.97	3.08	[-1.59, -0.35]	**
Time × Word Order	-0.24	2.48	[-0.44, -0.05]	*
A Fixations in Previous Time Bin	0.27	80.79	[0.26, 0.28]	***
P Fixations in Previous Time Bin	-0.21	63.86	[-0.22, -0.20]	***
1700–2700 ms				
Intercept	-1.35	4.14	[-2.00, -0.71]	***
Time	-0.34	3.49	[-0.54, -0.15]	***
Voice (= patient voice)	0.48	1.27	[-0.26, 1.23]	
Time × Voice	0.23	1.96	[>-0.01, 0.45]	
Word Order (= Pred P A)	1.00	2.70	[0.27, 1.72]	**
Time × Word Order	0.34	2.82	[0.10, 0.57]	**
A Fixations in Previous Time Bin	0.35	85.38	[0.34, 0.36]	***
P Fixations in Previous Time Bin	-0.23	55.33	[-0.24, -0.22]	***

Note: response = agent fixation; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

character in the previous time window, they turned their gaze towards the last-mentioned character after speech onset (interaction of Time and Word Order).

2.2.3 *Discussion*

Taken together, the results of this experiment provide the first fine-grained insight into the selection of grammatical structures and the time course of sentence formulation in Tagalog.

2.2.3.1 *Pivot selection and word order*

Conceptual accessibility had a strong influence on pivot selection in the sentences produced in Experiment 1. Patient voice was always produced when the patient in the event was human, obeying the grammatical constraint that human patients must be pivots (Latrouite, 2011). This means that speakers always assessed whether the patient was human or non-human early in the formulation process because this determined whether they would have to prepare patient voice marking for the sentence-initial predicate. In the events where it was possible to choose pivots (i.e., for pictures with non-human patients), speakers numerically tended to select human agents to be pivots (but this effect was not statistically significant). Thus, conceptual accessibility — in interaction with characters' semantic roles — affected the assignment of syntactic functions, consistent with previous findings in subject-initial languages (Branigan et al., 2008) and in the verb-initial language Tzeltal (Norcliffe, Konopka, et al., 2015). Conceptually accessible characters were more likely to become pivots although they were only mentioned sentence-finally in the majority of cases (approx. 94% of responses were pivot-final), consistent with accounts that assume that syntactic functions are ordered in a hierarchy and that the most accessible message elements have access to the highest syntactic function first (Bock, 1986; Bock & Warren, 1985).

The finding that pivot selection (and therefore which voice affix was planned for the sentence-initial predicate) was influenced by conceptual accessibility supports hierarchical incrementality. It suggests that Tagalog speakers encoded the event as a whole early on (i.e., by extracting its gist very quickly after picture onset; cf. Dobel, Gumnior, Bölte, & Zwitserlood, 2007) and also encoded the depicted characters' semantic roles and assessed their humanness before selecting a pivot.

The order of agent and patient was heavily influenced by the voice marking of the predicate, which determines which semantic role will canonically be positioned sentence-finally (given that pivot-final word order is the canonical word order in Tagalog). When the agent was the pivot, the canonical order is [Pred_{AV} P A_{PVT}]; when the patient is the pivot, it is [Pred_{PV} A P_{PVT}].

Perceptual accessibility neither influenced pivot selection nor the order of arguments, strongly suggesting that Tagalog speakers did not begin sentence planning with the immediate encoding of the most readily available nominal concept, in contrast to Gleitman et al. (2007).

Taken together, the influence of a character's accessibility and semantic role on structural choices supports hierarchically incremental planning, i.e., a planning process with a wide scope that takes into account information about arguments of the predicate that are only mentioned after speech onset.

2.2.3.2 *Time course of sentence formulation*

Time-course analyses for the different voice-marked sentence types also support a hierarchical account of sentence planning. At the outset of formulation (100–600 ms), speakers preferentially fixated pivot characters: In agent voice sentences, the agent was more likely to be fixated and in patient voice sentences, the patient was more likely to be fixated. However, there were no differences based on the order of agent and patient after the predicate. This suggests that these fixations reflect early encoding of the pivot argument's semantic role, which depended on the specific action in the event and which was needed to select the appropriate voice affix for the predicate.

After 600 ms, speakers shifted their gaze towards the first-mentioned character, and around speech onset they began to preferentially fixate the second-mentioned character, i.e., in sentences with [A P] word order, they gazed at the agent after 600 ms and shifted to the patient around speech onset (and vice versa for [P A] orders). This was independent of which character was selected as pivot and shows sequential lexical encoding of the arguments, guided by a structural-relational representation generated at the outset of formulation (Griffin & Bock, 2000; Konopka & Meyer, 2014; Norcliffe, Konopka, et al., 2015).

We interpret these two findings (early pivot fixations independent of word order before 600 ms and fixations contingent on the order thereafter) to reflect the purely morphosyntactic planning of the dependency between the predicate and the pivot argument in the early window, i.e., the selection of a head-marking voice affix. This is again consistent with hierarchical incrementality as it shows that structural planning is carried out early in the formulation process: At the outset of formulation, speakers selected a pivot argument and subsequently directed their visual attention to the corresponding character to plan the appropriate voice affix for the sentence-initial predicate.

2.3 EXPERIMENT 2

In Experiment 2, we investigate the production of recent perfective sentences, which do not exhibit head-marking but in which the sentence-initial predicate takes an invariant form. This allows us to assess the influence of accessibility on word order in voiceless sentences, and compared to Experiment 1, it allows us to probe the effect of head-marking morphology on the planning of predicate-initial sentences.

To investigate the time course of formulation when speakers produce sentences with different constituent orders but without the need to plan overt morphological dependencies between predicate and arguments, we compared agent-medial and agent-final recent perfective sentences (2.2). We predicted that fixation patterns would not differ early after picture onset in sentences with different orders of agent and patient because the predicate takes the same form in both sentence types.

Then, to investigate how formulation is influenced by head-marking morphology, we directly compared agent and patient fixations in recent perfective sentences from Experiment 2 to fixations in agent voice sentences from Experiment 1, which showed that when planning voice-marked sentences, speakers preferentially fixated the pivot character at the outset of formulation, reflecting extensive structural-relational encoding to select the predicate's voice affix. Thus, we predict that, for the production of recent perfective sentences in Experiment 2, speakers would distribute their attention more evenly among agent and patient characters between 100–600 ms because in these sentences no pivot function must be assigned and no voice affix must be planned.

2.3.1 *Participants*

Thirty-five participants from the same population as in Experiment 1 (12 males, mean age = 17.2 years, age range = 16–19 years) participated for payment.

2.3.2 *Materials, design and procedure*

The same stimuli as in Experiment 1[6] as well as 16 additional pictures depicting transitive events and 17 additional fillers depicting intransitive events were used (Appendix). The new stimuli were added to increase the number of observations per participant.

The procedure was identical to Experiment 1, except that participants were asked to describe the pictures as if the events had just happened. This was done to make the use of recent perfective aspect felicitous. In order to make the task clearer and to prime speakers towards using the recent perfective aspect, four additional example pictures were added to the training block at the beginning of the experiment. The example pictures were presented one at a time and were accompanied by pre-recorded descriptions in recent perfective aspect; next, participants were asked to describe the pictures themselves (using recent perfective aspect). As in Experiment 1, the experimenter provided feedback if necessary. Participants' eye movements and descriptions were recorded as in Experiment 1.

2.3.3 *Sentence scoring and analyses*

Responses on target trials were transcribed and coded as recent perfective sentences or as other constructions. Only recent perfective sentences with agents, patients and the adverb *(pa)lang* 'just' realized overtly were included in the analysis. All additional exclusion criteria used in Experiment 1 were also applied. The final dataset consisted of 1124 sentences.

The analyses of word order choice and the time course of formulation were performed as in Experiment 1. Additionally, we carried out

[6] Two stimuli from Experiment 1 were not included in Experiment 2: one stimulus picture was accidentally used as an example during the practice phase of Experiment 2, the other stimulus was removed because it only elicited other constructions that had to be excluded during sentence scoring.

a separate analysis comparing fixation patterns between sentences in which the predicate carried voice marking morphology (Experiment 1) and recent perfective sentences (Experiment 2), including only the target pictures that were presented in both experiments.

2.3.4 *Results*

2.3.4.1 *Distribution of responses*

Participants produced 1003 sentences with [Pred$_{RP}$ A P] word order (2.2a) and 121 sentences with [Pred$_{RP}$ P A] word order (2.2b). In 59% of the trials, speakers fixated the agent character first.

2.3.4.2 *Word order*

The first analysis for this experiment tested whether perceptual and conceptual accessibility influenced the order of agent and patient after the predicate (Table 2.4).

The majority of sentences were produced with the canonical [A P] word order (Figure 2.5). Human patients were more likely to be positioned sentence-finally (i.e., non-human patients were more likely to be mentioned immediately after the predicate; main effect of Patient Humanness, Table 2.4). Although the interaction between Agent Humanness and Patient Humanness was not statistically significant, non-human patients were often mentioned sentence-medially when the agent was human (Figure 2.5). Human agents were also more likely to be positioned sentence-finally (i.e., non-human agents were more likely to be mentioned immediately after the predicate; main effect of Agent Humanness, Table 2.4). First fixations (i.e., our measure of perceptual accessibility) did not significantly influence word order choices.

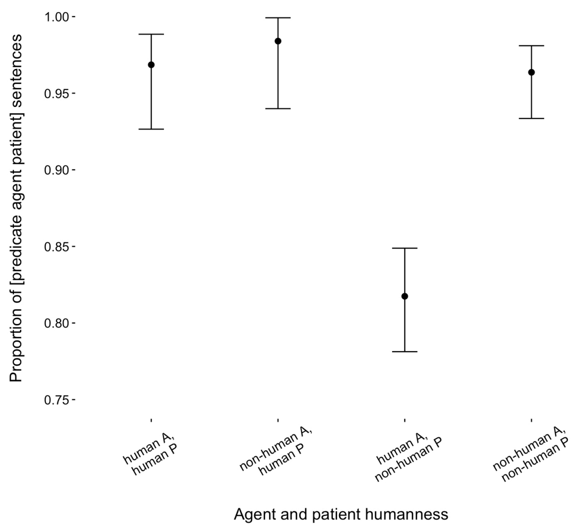

Figure 2.5: Proportions of sentences with [predicate agent patient] word order produced in Experiment 2 as a function of agent (A) and patient (P) humanness.

Table 2.4: Binomial mixed effects regression results for the effects of perceptual, conceptual and lexical accessibility on the order of agent and patient after recent perfective predicates.

	$\hat{\beta}$	$\lvert Z \rvert$	95% Wald CI	
Intercept	-2.92	9.90	[-3.50, -2.34]	***
First-fixated Character (= patient)	0.38	1.10	[-0.29, 1.04]	
Agent Humanness (= non-human)	-1.45	2.82	[-2.45, -0.44]	**
Patient Humanness (= non-human)	1.44	2.55	[0.33, 2.55]	*
Agent Humanness × Patient Humanness	-0.55	0.44	[-2.98, 1.88]	

Note: response = [Pred P A] word order; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.3.4.3 *Time course of sentence formulation*

Two sets of comparisons were carried out for recent perfective sentences, using the same time windows as in Experiment 1. The first set of comparisons contrasted recent perfective sentences with the canonical [A P] order to recent perfective sentences with [P A] order to test whether and when speakers' fixation patterns differed during production of sentences that differed only in word order (Figure 2.6). We predicted that speakers' fixation patterns would only differ between the two sentence types in the two later time windows (600–1600 ms and 1600–2600 ms) when character names are linguistically encoded. During the 0–600 ms time window, there should be no differences between sentence types because the predicate is marked in the same way and we expected speakers to primarily engage in relational planning in this time window in order to encode the predicate (Table 2.5).

The second set of comparisons assessed differences between recent perfective (Experiment 2) and agent-voice sentences (Experiment 1) to test how the processes of pivot selection and planning of voice morphology influence early formulation (Table 2.6).

RECENT PERFECTIVE SENTENCES WITH DIFFERING WORD ORDER In the first analysis time window (100–600 ms), speakers were overall more likely to fixate on the agent in both types of recent perfective sentences (Intercept; Table 2.5). In the two later time windows, fixations were contingent on word order: During the 600–1700 ms time window, speakers directed their gaze towards the agent character in [Pred$_{RP}$ A P] sentences and away from the agent in [Pred$_{RP}$ P A] sentences (interaction of Time and Word Order). In the 1700–2700 ms time window, by contrast, speakers fixated the character that was mentioned sentence-finally, i.e., the patient in

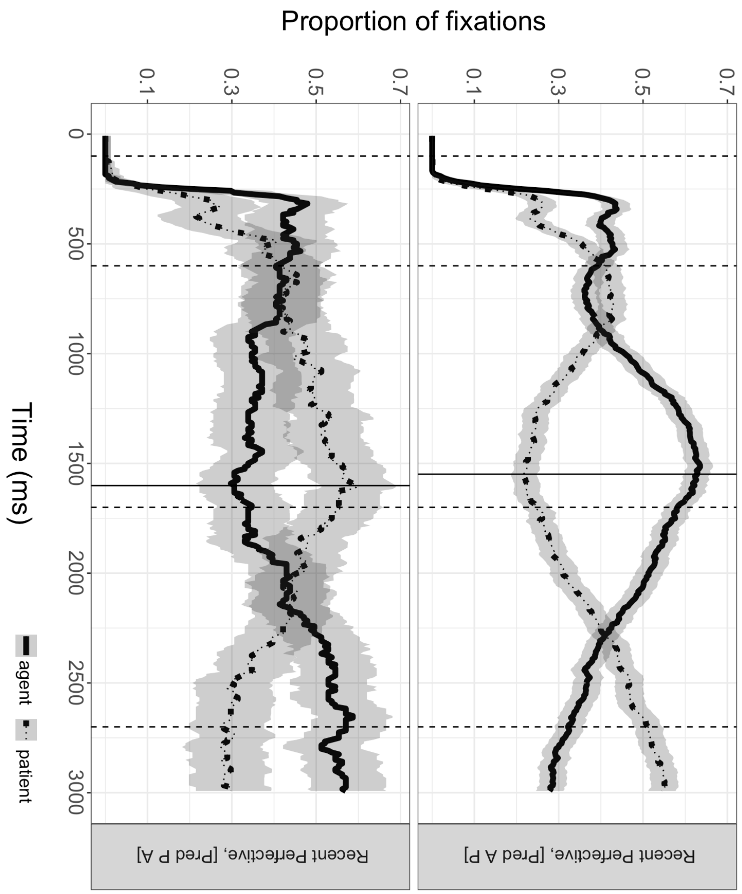

Figure 2.6: Mean proportions of fixations to agents and patients in two recent perfective sentence types with different word orders; ribbons indicate 95% confidence intervals (Villacorta, 2012); solid vertical lines indicate mean speech onsets; dashed vertical lines mark analysis time windows.

Table 2.5: Binomial mixed effects regression results modelling the influence of word order on the time course of fixations towards agent characters vs. patient characters during the formulation of recent perfective sentences.

	$\hat{\beta}$	$\|Z\|$	95% Wald CI	
100–600 ms				
Intercept	0.59	2.19	[0.06, 1.12]	*
Time	-0.52	1.85	[-1.08, 0.03]	
Word Order (= Pred P A)	$>$ -0.01	0.03	[-0.58, 0.56]	
Time × Word Order	0.03	0.38	[-0.12, 0.19]	
A Fixations in Previous Time Bin	0.22	37.13	[0.21, 0.24]	***
P Fixations in Previous Time Bin	-0.18	26.96	[-0.20, -0.17]	***
600–1700 ms				
Intercept	0.58	7.00	[0.42, 0.74]	***
Time	0.11	3.12	[0.04, 0.17]	**
Word Order (= Pred P A)	-0.62	1.59	[-1.39, 0.14]	
Time × Word Order	-0.30	2.11	[-0.52, -0.02]	*
A Fixations in Previous Time Bin	0.26	73.93	[0.25, 0.26]	***
P Fixations in Previous Time Bin	-0.27	80.10	[-0.28, -0.27]	***
1700–2700 ms				
Intercept	-0.74	6.68	[-0.96, -0.53]	***
Time	-0.08	2.75	[-0.14, -0.02]	**
Word Order (= Pred P A)	0.85	2.74	[0.24, 1.45]	**
Time × Word Order	0.26	12.77	[0.22, 0.30]	***
A Fixations in Previous Time Bin	0.35	88.45	[0.35, 0.36]	***
P Fixations in Previous Time Bin	-0.24	57.11	[-0.25, -0.24]	***

Note: response = agent fixation; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6: Binomial mixed effects regression results modeling the influence of the influence of voice marking and word order on the time course of formulating recent perfective sentences (Experiment 2) compared to agent voice sentences (Experiment 1).

	$\hat{\beta}$	$\|Z\|$	95% Wald CI	
100–600 ms				
Intercept	0.61	1.83	[-0.04, 1.27]	
Time	-0.44	1.02	[-1.28, 0.40]	
Voice (= agent voice)	1.10	3.53	[0.49, 1.71]	***
Time × Voice	0.94	1.93	[-0.01, 1.90]	
Word Order (= Pred P A)	-0.30	0.79	[-1.04, 0.44]	
Time × Word Order	-0.91	1.70	[-1.98, 0.14]	
A Fixations in Previous Time Bin	0.21	34.90	[0.20, 0.23]	***
P Fixations in Previous Time Bin	-0.19	27.23	[-0.20, -0.17]	***
600–1700 ms				
Intercept	0.69	3.38	[0.29, 1.09]	***
Time	0.16	2.10	[0.01, 0.31]	*
Voice (= agent voice)	-0.06	0.40	[-0.37, 0.25]	
Time × Voice	0.11	1.19	[-0.07, 0.30]	
Word Order (= Pred P A)	-0.87	4.63	[-1.24, -0.50]	***
Time × Word Order	-0.30	2.80	[-0.52, -0.09]	**
A Fixations in Previous Time Bin	0.27	77.74	[0.26, 0.28]	***
P Fixations in Previous Time Bin	-0.25	72.17	[-0.25, -0.24]	***
1700–2700 ms				
Intercept	-0.76	2.86	[-1.28, -0.24]	**
Time	-0.15	1.83	[-0.32, 0.01]	
Voice (= agent voice)	-0.48	1.18	[-1.28, 0.32]	
Time × Voice	-0.36	2.38	[-0.66, -0.06]	*
Word Order (= Pred P A)	1.16	2.70	[0.32, 2.00]	**
Time × Word Order	0.49	2.82	[0.15, 0.84]	**
A Fixations in Previous Time Bin	0.33	81.57	[0.32, 0.34]	***
P Fixations in Previous Time Bin	-0.25	59.73	[-0.26, -0.24]	***

Note: response = agent fixation; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

sentences with the canonical word order and the agent in sentences with [P A] word order (interaction of Time and Word Order).

COMPARING EXPERIMENT 1 AND 2: DIFFERING HEAD-MARKING There were 42 pictures that were used in both experiments. This enabled us to investigate how the presence or absence of head-marking morphology on the sentence-initial predicate influences the the time course of early sentence planning (Table 2.6). We compared sentences with both possible orders of agents and patients, i.e., agent voice sentences with [Pred_{AV} A_{PVT} P] and [Pred_{AV} P A_{PVT}] word orders from Experiment 1 (Figure 2.4a–b) and recent perfective sentences with [Pred_{RP} A P] and [Pred_{RP} P A] word orders from Experiment 2 (Figure 2.6).

Agent and patient fixations differed between agent voice and recent perfective sentences in the 100–600 ms time window. Speakers overall fixated more on agents when they were selected as pivots and thus when their semantic role was signaled on the predicate (Experiment 1) than when there was no head-marking morphology (in recent perfective sentences from Experiment 2; main effect of Voice). As in Experiment 1, in the two later time windows, the order of agent and patient determined speakers' eye movements. During the 600–1700 ms time window, speakers' fixations to the agents increased as they were mentioned immediately after the predicate in both agent voice and recent perfective sentences. Similarly, in the last time window (1700–2700 ms), speakers directed their gaze away from agents and towards patients in both sentences types (interactions of Time and Word Order). In addition, they fixated quicker on the patient in agent voice sentences in the last time window (interaction of Time and Voice); we do not have an explanation for this effect.

2.3.5 *Discussion*

2.3.5.1 *Word order*

The order of arguments after the predicate in recent perfective sentences was independently influenced by the conceptual accessibility of both agents and patients (cf. Figure 2.5). When the agent was non-human, it was more likely to be placed in the canonical sentence-medial position. One possible explanation for this effect is that speakers aimed to avoid potential semantic role ambiguities when describing pictures with the unusual configuration of showing a non-human agent. Kroeger (1993b, 111) describes a grammatical tendency to place agents in canonical position to avoid ambiguities in sentences where both arguments could be interpreted as agents, as it is potentially the case in the current experiment with sentences with non-human agents. For example, in (2.3), both the agent (*bird*) and the patient (*worm*) can be marked as non-pivots by *ng* so that case marking may not be available as a reliable cue to the semantic roles. In contrast to imperfective and perfective sentences (Experiment 1), voice marking with its associated pivot-final canonical word order is not available as a cue in recent perfective sentences where the predicate takes an invariant form.

(2.3)	ka-ka∼kain	lang	ng=ibon	ng=uod
	RP-RDP∼eat	just	NPVT=bird	NPVT=worm
	"A/the bird just ate a/the worm."			

Thus, in Experiment 2, Tagalog speakers might have strategically employed humanness to place less accessible, non-human agents in the canonical position to give a cue to their semantic role. Speakers might have traded in production ease for the goal of reducing ambiguity in their recent perfective sentences because they tended to mention the non-human agents first especially when the patient was human. This finding of early placement of non-human agents supports accounts of hierarchical incrementality. When a structural-relational utterance representation guides lexical retrieval, the conceptual accessibility of characters may not have an immediate influence on the order of mention, allowing speakers to place non-human agents sentence-medially. By contrast, in linearly incremental sentence production, it is generally assumed that incremental lexical encoding disfavors the early placement of *less* accessible concepts and favors the early placement of *highly* accessible concepts because the latter are available for linguistic encoding earlier (Branigan et al., 2008).

The early placement of non-human agents in Tagalog is similar to findings by Kurumada and Jaeger (2015) who showed that Japanese speakers were more likely to produce optional object case markers if the sentence would otherwise be ambiguous, i.e., speakers accepted an increased production effort if this ensured that the intended meaning of their utterances would thereby be easier to comprehend. We return to this point in Section 2.4.

Additionally, non-human patients were also more likely to be mentioned immediately after the predicate than human patients, especially when the agent was human (cf. Figure 2.5). In these sentences, the semantic roles of arguments were unlikely to be confusable so that speakers could have mentioned the patient in this position without risking that a potential listener would mistake it as the agent in canonical position because the predicate's meaning and the patient's non-humanness might have prevented such an interpretation.

The perceptual accessibility of characters did not influence the order of agents and patients, lending further support to the interpretation that Tagalog speakers' word order choices for recent perfective sentences were plausibly largely driven by considerations to avoid ambiguities as to which argument carries which semantic role.

In sum, the word order results from recent perfective sentences support a hierarchically incremental planning account: Speakers did not simply place highly accessible characters early in the sentence, as predicted by linear incrementality, but strategically choose word orders that minimized potential ambiguities. The early placement of *less* accessible concepts is compatible with hierarchical incrementality where linguistic encoding is guided by a structural-relational utterance representation and not by solely the conceptual accessibility of individual characters.

2.3.5.2 *Time course of sentence formulation*

Fixations to agents and patients were distributed to similar degrees from stimulus onset until approximately 600 ms in both recent perfective sentence types, independent of whether the agent or the patient were eventually mentioned sentence-medially or sentence-finally. This indicates that speakers engaged in early relational encoding of the event in order to select the predicate, supporting hierarchically incremental accounts of sentence planning that start with the generation of a structural-relational utterance representation.

Early agent and patient fixations (100–600 ms) in both recent perfective sentence types were similar because there was no head-marking morphology to plan and thus the predicate took the same form in both sentence types. This is in contrast to the findings from Experiment 1, where agent voice and patient voice sentences were associated with different fixation patterns (more agent fixations in agent voice sentences and more patient fixations in patient voice sentences), suggesting that these early fixations reflected the selection of a pivot and the preparation of the predicate's voice affix. During the planning of recent perfective sentences with invariant predicates, by contrast, no such differences were observed; this lends support to the interpretation that the head-marking morphology in the sentences in Experiment 1 influenced the time course of planning early during formulation. Further support comes from the increased agent fixations in agent voice sentences (Experiment 1) as compared to recent perfective sentences (Experiment 2), which shows that voice marking on the predicate influenced the time course of sentence formulation. We return to this point below.

Finally, in both voice-marked sentences (Experiment 1) and recent perfective sentences (Experiment 2), fixations in the time windows after 600 ms were primarily contingent on word order. This shows that in general character names were encoded largely sequentially once speakers started linguistic encoding, as is assumed by hierarchical incrementality.

2.4 GENERAL DISCUSSION

In two picture description experiments in Tagalog, we examined the influence that head-marking morphology of sentence-initial verbal predicates had on the scope of planning and the time course of sentence planning. Overall the results show that speakers engage in extensive advance planning beyond the sentence-initial predicate, with the grammatical properties of the to-be uttered sentence exerting a strong influence on the formulation process.

2.4.1 *Hierarchically incremental planning*

The results from the analyses of structural choices and eye movement data from both voice-marked sentences and voiceless recent perfective sentences broadly support accounts of hierarchical incrementality.

In Experiment 1 with voice-marked sentences, human patients were always selected as pivots (obeying a grammatical constraint), which suggests that speakers encoded the event and the semantic roles of depicted characters early during formulation in order to assign the pivot function and select a voice affix. This suggests that

Tagalog speakers assigned syntactic functions based on a representation of the entire event because the notions of agent and patient are fundamentally relational (in a transitive event, there cannot be a patient without an agent acting on it; cf. Primus, 1999).

A hierarchically incremental view of Tagalog sentence production is also supported by analyses of speakers' eye movements during the planning of both voice-marked sentences (Experiment 1) and recent perfective sentences (Experiment 2). In voice-marked sentences, speakers fixated on the pivot character in the early time window (0–600 ms) after picture onset. This suggests that they quickly encoded the depicted event as a whole (Dobel et al., 2007; Norcliffe, Konopka, et al., 2015) and the characters' semantic roles and humanness because this was necessary to select the pivot. In recent perfective sentences, there were no pivot arguments and speakers preferentially fixated on the agent character in this early time window. This suggests early relational encoding, too, because the agent fixations were independent of whether it would be mentioned immediately after the predicate or only in sentence-final position, compatible with early structural-relational encoding. Thus, the speakers in our experiments might have fixated on agent characters preferentially because they are critical for event conceptualization because agents are the instigators of the depicted actions (Bock, Irwin, & Davidson, 2004; Bornkessel-Schlesewsky & Schlesewsky, 2013a; Sauppe, 2016).

In the Introduction, we described how linearly incremental accounts that assume that planning starts with the encoding of the most accessible nominal concept (Gleitman et al., 2007) run into problems accounting for predicate-initial sentence production. We proposed that a more literal understanding of linear incrementality would be needed which assumes that planning starts with the encoding of the information that is necessary to produce the first element in a sentence. This view of linear incrementality is, however, very similar to hierarchical incrementality because planning the first word (the predicate) requires a representation of the event and the relations among event participants (Griffin & Bock, 2000). We focused on testing hypotheses derived from hierarchical incrementality here because no contrasting predictions could be derived from linear incrementality. Crucially, the prediction of hierarchical incrementality that the production of predicate-initial sentences affords early relational encoding was confirmed by the results of both reported experiments on Tagalog.

However, the question of how much relational information Tagalog speakers encode before selecting a pivot remains open for further research. While Bunger, Papafragou, and Trueswell (2013) report that speakers' linguistic choices are influenced by the early recognition of abstract event categories (e.g., a motion event or a contact event), Griffin and Bock (2000) assume that relational encoding involves a specified relation between event participants (e.g., eating, throwing, kicking). The data from the current experiments are compatible with both possibilities: either speakers only encoded what event category was depicted in any given picture or they encoded specific actions before selecting the agent or the patient as the pivot argument.

In sum, evidence from both structural choices as well as the time course of formulation support the view that Tagalog predicate-initial sentences are planned in a hierarchically incremental fashion. The findings from the current experiments on Tagalog are consistent with Norcliffe, Konopka, et al.'s (2015) findings on sentence planning in Tzeltal, suggesting that predicate-initial word order favors hierarchically incremental planning because speakers need to engage in relational encoding early to select the first word of the sentence.

2.4.2 *The influence of head-marking morphology on the time course of sentence planning*

The two Tagalog sentence types that we investigated here, voice-marked sentences (2.1) and recent perfective sentences without voice marking (2.2), most prominently differed with respect to the head-marking of the predicate, i.e., whether the head marking cross-referenced the semantic role of its pivot argument or not. Both results from word order choices and eye movement data show that Tagalog speakers employed different planning strategies depending on whether they needed to plan head-marking morphology for the predicate.

The humanness or animacy of a character influenced pivot selection in Tagalog as it did in Tzeltal (Norcliffe, Konopka, et al., 2015). These two languages are grammatically similar in this regard, since they both place the predicate first and it exhibits some head-marking morphology. However, the degree of head-marking morphology differs between the languages. In Tagalog, the predicate signals only the pivot's semantic role, whereas in Tzeltal, there is agreement with both the subject's and the object's person and number. An interesting question for future research is thus whether the same effects of conceptual accessibility would also be observed in verb-initial languages that do not exhibit head-marking morphology. Does the need to assign syntactic functions early in order to encode head-marking affixes in fact enhance the influence of conceptual accessibility on structural choices? This could be the case because animacy and humanness could be available to speakers quickly and are relatively abstract features influencing grammatical function assignment (Bock & Levelt, 1994; Branigan et al., 2008).

Tagalog speakers' word order choices for the production of recent perfective sentences (Experiment 2) in contrast to perfective and imperfective sentences (Experiment 1) suggest that not only the grammatical properties of different languages on the whole may influence the planning strategies employed (Myachykov, Garrod, & Scheepers, 2009) but that also two constructions within a single language that differ in their head-marking may lead to differences in planning. In recent perfective sentences, speakers strategically mentioned non-human (i.e., conceptually less accessible) agents immediately after the predicate, in the canonical agent position. We suggest that this could be a "hearer-oriented planning strategy" to avoid potential ambiguities for listeners. An example for such a strategy comes from the findings of Kurumada and Jaeger (2015), who show that optional case markers are highly likely to be realized in Japanese when the word order biases against the intended interpretation of the sentence. The Japanese case markers clearly indicate the syntactic functions of the arguments and thereby ensure that listeners correctly assign the semantic roles during comprehension. This account assumes that "production reflects a trade-off between production ease and the goal to be understood" (Kurumada & Jaeger, 2015, 164) and stands in contrast to accounts stating that grammatical encoding is only affected by factors relevant to the speaker, such as accessibility. For example, under V. S. Ferreira's (2008) view, speakers' choice of grammatical structures is only influenced by production efficiency while making utterances easy to understand for listeners plays no role whatsoever (cf. also V. S. Ferreira & Dell, 2000).

In Tagalog, recent perfective sentences differ from perfective and imperfective sentences in that a voice affix on the predicate signals the pivot argument's semantic role and case markers signal the different syntactic functions of pivot and non-pivot arguments in the latter. There is, on the other hand, no pivot in recent perfective sentences and case markers are not assigned based on voice marking in these sentences. It is possible that speakers aimed to adapt to the lack of both of these cues by placing non-human agents in the canonical position in order to make interpretation of these sentences easier for listeners. In voice-marked sentences, by contrast, we did not find evidence for accessibility-based effects on voice and word order that would make production easier for speakers or comprehension easier for listeners (cf. V. S. Ferreira, 2008).

In a visual world experiment, Sauppe (2016) shows that Tagalog listeners use the semantics of the predicate and the visual context to make predictions about potential agents and patients during comprehension, independently of whether the predicate is voice-marked or not (Chapter 5). This implies that it might not be necessary that speakers undertake the effort to place less prototypical and accessible agents immediately after the predicate because listeners can make good predictions about semantic role assignments on the basis of the predicate's lexical meaning alone. This would speak against an interpretation of the canonical placement of non-human agents in recent perfective sentences in terms of hearer-oriented planning processes. However, whereas agent and patient referents were easily predictable from the combination of predicate and visual display in the comprehension study of Sauppe (2016), the typicality of actions, agents and patients was not controlled for or manipulated in the current production experiments. We also do not know what role communicative goals might play in this context (cf. Barthel, Sauppe, Levinson, & Meyer, 2016, and Tanenhaus & Brown-Schmidt, 2008, for examples of tasks with a communicative component). One possible scenario would be that speakers imagined a listener to whom they described the pictures. This potential listener could have been able to also see the pictures or would not have access to this visual information. The latter, hypothetical communicative setting could have encouraged speakers to use structures that help to make semantic role assignments as easy as possible (even if this went in hand with increased production effort) because the visual context would not have helped listeners to predict agent and patient referents and to quickly decide on an interpretation of the sentence. We leave it to future research to determine how introducing an explicit communication setting to the experimental task may influence word order choices in the production of recent perfective sentences in Tagalog. In Kurumada and Jaeger's (2015) ideal speaker account, speakers' structural choices are always influenced by the aim "to produce signals that increase the probability of successfully conveying the intended message" (Kurumada & Jaeger, 2015, 154). In other words, under this account, the underlying organization of sentence formulation processes leads to a tendency for speakers to produce structures that avoid ambiguities (i.e., recent perfective sentences in Tagalog with sentence-medial non-human agents) even in the absence of a (potential) listener.

Another open question is whether and how the typicality of agents and patients or the symmetry of the event and the likelihood of false role assignments (i.e., whether both characters could be the agent or the patient, e.g., with chasing or attacking events) might influence the tendency to place non-human agents early in these sentences.

In addition to differences in word order planning, eye movement data from voice-marked sentences (Experiment 1) and recent perfective sentences (Experiment 2) show that, while speakers engaged in early relational planning in both sentence types, the head-marking morphology in voice-marked sentences influenced the visual information uptake early during formulation. In these sentences, the initial predicate signaled the pivot argument's semantic role by means of different voice affixes. Compared to the production of voiceless recent perfective sentences where the predicate does not signal any argument's semantic role, speakers preferentially fixated the pivot characters of voice-marked sentences in the earliest time window (100–600 ms). We suggest that these early fixations reflect the selection of a pivot argument and the planning of the voice affix because speakers preferentially fixated on pivot characters during this time window regardless of their semantic role or their position in the sentence. To the best of our knowledge, this is the first demonstration of how head-marking morphology influences the online planning of sentences.

In light of the findings from Tagalog, the high degree of advance planning in Tzeltal verb-initial active sentences (Norcliffe, Konopka, et al., 2015) with extensive fixations to both agent and patient characters might be explained by Tzeltal's complex head-marking morphology, which involves both subject and object agreement. Speakers of this language encoded information about both characters early in the formulation process in order to select the agreement affixes marking person and number at the verb. While third person objects are marked by a zero morpheme, speakers presumably still needed to encode number and person information for these referents in a similar fashion as they did for agent subjects marked by an overt prefix. It is also interesting to note that the time course of early agent and patient fixations in this sentence type mirrored the affix order on the Tzeltal verb: First speakers preferentially fixated on the agent character and then on the patient character. The agreement with the former is marked by a prefix and the latter by a suffix on the verb.

In sum, the predicate-initial word order of Tagalog required a hierarchically incremental planning mode and prevented speakers from beginning to produce a sentence by first encoding an accessible character. Konopka and Meyer (2014) show that speakers in subject-initial languages, on the other hand, may be flexible in the choice to give prevalence to linguistic encoding of individual characters or to relational encoding of the event depending how easy or hard it is to encode the one or the other (cf. also van de Velde et al., 2014). This provides further evidence that the grammatical properties of a language influence the planning strategies that are readily available to speakers (Myachykov et al., 2009). Additionally, we presented evidence for different planning strategies within Tagalog that are finely tuned to the grammatical properties of the constructions under production.

One further question worth exploring is how the incremental nature of sentence production and flexibility in planning strategies may influence word order patterns in language change. In the planning literature, it has been suggested that speakers' flexibility in switching between hierarchically incremental and linearly incremental planning might help to maintain fluency in discourse because it allows formulation to begin with the encoding of the most accessible information, thereby "optimizing the processing resources" (F. Ferreira & Engelhardt, 2006, 76; cf. also V. S. Ferreira & Dell, 2000). In the Introduction, we noted that languages that put verbal predicates in sentence-initial position are relatively rare among the world's languages and are vastly outnumbered by subject-initial languages (Dryer, 2013). Verb-initial word orders are also more likely to change to subject-initial word orders over time than changes from subject- to verb-initial order (Maurits & Griffiths, 2014). An explanation for these observations might thus be that linguistic systems that provide high degrees of flexibility for speakers are favored in the evolution of languages. Subject-initial languages that allow speakers to choose between planning strategies, depending on whether relational event structure or information about individual nominal concepts is easier to encode at any given moment, might be more adaptive and might have an evolutionary advantage (Christiansen & Chater, 2008) because they facilitate the speakers' ability to retain fluency. We argued that verb-initial languages, by contrast, require early relational planning and thereby restrict speakers' flexibility in choosing between planning strategies. Verb-initial word orders may thus be disfavored in language change.

2.5 CONCLUSIONS

In two experiments, we showed that the sentence planning processes in Tagalog are heavily influenced by its word order and by variations in head-marking morphology in different sentence types from the earliest stages of formulation on. We argued that linear incrementality (i.e., the proposal that planning starts with the encoding of accessible nominal concepts) cannot account for predicate-initial sentence production because the need to select a verbal predicate early on entails early relational encoding and voice marking on the predicate requires the assignment of syntactic functions long before the corresponding argument is articulated. If linear incrementality were re-conceptualized to refer to "word by word" planning in the sense that planning in predicate-initial languages starts with relational encoding of the event, it is barely distinguishable from hierarchical incrementality. The sharp differences between those two sentence production accounts that have been described in the literature are particularly suited to model the formulation of subject-initial sentences, mainly because they were developed based on data from subject-initial languages (Gleitman et al., 2007; Griffin & Bock, 2000). However, if the aim is to develop universally applicable models of sentence production, more research on under-studied languages with structures differing from, for example, English and Dutch is necessary because such studies provide evidence that is impossible to obtain through experimentation on the conventionally studied languages in our field (Jaeger & Norcliffe, 2009; Levinson, 2012).

APPENDIX

2.A TARGET EVENT PICTURES USED IN EXPERIMENT 1

Critical stimuli were the following event depictions, all of which can be described with transitive sentences in Tagalog:

- Ambulance car running over a woman
- Baker kneading bread dough
- Bird drinking water from a fountain
- Bird pulling a worm out of the ground
- Boxer beating a man
- Boy catching a frog
- Boy eating corn
- Boy kicking a football
- Boy kicking a rock
- Boy stirring a soup
- Bull attacking a girl
- Crocodile biting a man's leg
- Dog chasing a mailman
- Dog chasing a squirrel
- Dog sniffing a mandarin
- Gardener planting a tree
- Girl hanging out laundry
- Girl opening a door
- Girl playing with a jumping rope
- Girl leaving house through open door
- Man angling a fish
- Man chopping a log of wood
- Man kicking a chair
- Man leaving a hut
- Man pushing a car
- Mosquito stinging a football player
- Mouse nibbling on a chocolate bar
- Nurse washing a baby
- Old man reading a book
- Old woman climbing up stairs
- Paper boy selling newspapers
- Police officer arresting a man
- Police officer stopping a sports car
- Pupil raising his hand
- Rabbit eating a carrot
- Sailor drinking from a bottle
- Sheep eating leaves from a bush
- Soldier shooting a man
- Train colliding with school bus
- Veterinarian examining a horse's hoof
- Wind blowing off a tree's leaves
- Woman lifting a rug
- Woman inspecting a basket
- Woman walking across a bridge

2.B TARGET EVENT PICTURES USED IN EXPERIMENT 2

Critical stimuli were the following event depictions, all of which can be described with transitive sentences in Tagalog:

- Ambulance car colliding with a woman
- Baker kneading bread dough
- Bird pulling a worm out of the ground
- Boxer beating a man
- Boy breaking branch from a tree
- Boy catching a frog
- Boy eating corn
- Boy kicking a football
- Boy kicking against a rock
- Boy stirring in a soup
- Bull attacking a girl
- Cat catching a mouse
- Cat scratching a girl's knee
- Construction worker losing his hat
- Cowboy catching a bull with a lasso
- Crocodile biting into a man's leg
- Dog chasing a mailman
- Dog chasing a squirrel
- Dog sniffing on a mandarin
- Frog eating a fly
- Gardener planting a tree
- Girl hanging out laundry
- Girl opening a door
- Girl playing with a jumping rope
- Girl pushing a boy
- Girl running towards an open door
- Girl tripping a construction worker
- Lion eating a dead zebra
- Man angling a fish
- Man breaking a piece of wood with a hammer
- Man chopping a log of wood
- Man cutting wood
- Man kicking against a chair
- Man leaving a hut
- Man pushing a car
- Man throwing a baby up in the air
- Monkey holding a crab in its hand
- Mosquito stinging a football player

- Mouse nibbling on a chocolate bar
- Nurse washing a baby
- Old man opening a window
- Old man reading a book
- Old woman climbing up stairs
- Owl carrying a bag
- Paper boy selling newspapers
- Police officer arresting a man
- Police officer stopping a sports car
- Police officer stopping a walker-by
- Pupil raising his hand
- Rabbit eating a carrot
- Sailor drinking from a bottle
- Sheep eating leaves from a bush
- Soldier shooting a man
- Train colliding with school bus
- Veterinarian examining a horse's hoof
- Woman lifting a rug
- Woman looking inside a basket
- Woman walking across a bridge

3

WORD ORDER AND VOICE INFLUENCE THE TIMING OF VERB PLANNING IN GERMAN SENTENCE PRODUCTION

Sauppe, S. (2017). Word order and voice influence the timing of verb planning in German sentence production. Manuscript under review.

ABSTRACT

Theories of incremental sentence production make different assumptions about when speakers encode information about described events and when verbs are selected, accordingly. An eye tracking experiment on German testing the predictions from linear and hierarchical incrementality about the timing of event encoding and verb planning is reported. In the experiment, participants described depictions of two-participant events with sentences that differed in voice and word order. Verb-medial active sentences and actives and passives with sentence-final verbs were compared. Linear incrementality predicts that sentences with verbs placed early differ from verb-final sentences because verbs are assumed to only be planned shortly before they are articulated. By contrast, hierarchical incrementality assumes that speakers start planning with relational encoding of the event. A weak version of hierarchical incrementality assumes that only the action is encoded at the outset of formulation and selection of lexical verbs only occurs shortly before they are articulated, leading to the prediction of different fixation patterns for verb-medial and verb-final sentences. A strong version of hierarchical incrementality predicts no differences between verb-medial and verb-final sentences because it assumes that verbs are always lexically selected early in the formulation process. Based on growth curve analyses of fixations to agent and patient characters in the described pictures, and the influence of character humanness and the lack of an influence of the visual salience of characters on speakers' choice of active or passive voice, the current results suggest that while verb planning does not necessarily occur early during formulation, speakers of German always generate an event representation early.

3.1 INTRODUCTION

When speakers plan and formulate a sentence, they have to generate a message and transform it into a linearly ordered series of words. This process is generally believed to be incremental, such that speakers can start to articulate an utterance before it is planned in its entirety (Bock & Levelt, 1994; V. S. Ferreira & Slevc, 2007, inter alia). There are multiple views of incremental sentence planning that differ in their assumptions about when speakers engage in conceptual encoding and when grammatical structures are built. These views thus also differ in their expectations of when speakers encode information about the event that they are about to describe (i.e., the relations between agents and patients) and when they plan a sentence's verb, expressing the carried out action.

Theoretical accounts of sentence production can be grouped into linearly (or lexically) incremental approaches and hierarchically (or structurally) incremental approaches.

Accounts of linear incrementality assume that sentences are planned "word-by-word" (or phrase-by-phrase) and that speakers begin with the encoding of a nominal message element. The accessibility of message elements is assumed to influence which one is selected first by speakers who begin their sentences by encoding and articulating the most accessible message element's name first. Accessibility is influenced by many factors, two of the most notable are animacy (Branigan et al., 2008) and visual salience (Gleitman et al., 2007; Myachykov, Tomlin, & Posner, 2005). The encoding of other message elements and the event relations between them may be deferred until after speech onset (Gleitman et al., 2007; cf. also Kempen & Hoenkamp, 1987). Linear incrementality thus assumes that, during the piecemeal formulation of utterances, verbs are planned only shortly before they are uttered.

For example, in a series of picture-word interference experiments, Schriefers et al. (1998) found semantic interference effects in German for verbs that were produced early in transitive sentences (in verb-second position) but no interference for sentence-final verbs. This suggests that participants in this study planned verbs before speech onset only when they were positioned near the sentence beginning. Similarly, Allum and Wheeldon (2007) propose that only the first phrase of an utterance is planned before speaking starts. In subject-initial languages this means that the verb may only be planned after articulation of the subject already started.

In contrast, hierarchical incrementality assumes that speakers always generate a representation of the utterance at the outset of formulation. This representation provides "a linguistic action plan that provides information about where to start and how to continue an utterance" (Kuchinsky et al., 2011, 749).[1] The weak version of hierarchical incrementality assumes that planning starts with the encoding of the relationship between agent and patient, i.e., that speakers encode the thematic structure of the event that they are about to describe and what kind of action is carried out. This relational event representation allows speakers then to select a starting point, i.e., to decide which character to mention first (cf. Bock et al.,

[1] The idea of hierarchically incremental sentence planning goes back to Wilhelm Wundt's *Die Sprache* (Wundt, 1900), where he proposed that speakers begin sentence formulation with a *Gesamtvorstellung* ("total image") of an event, which is then dissected into its parts — entities, properties, and states —, allowing them to be lexicalized and uttered (Levelt, 2013).

2004). This view of sentence production was prominently proposed by Griffin and Bock (2000). They report results from a picture description experiment in which English speakers distributed their fixations between agents and patients in an early time window of up to 400 ms after the stimulus pictures appeared on the screen. This initial, pre-linguistic event apprehension phase is assumed to subserve the extraction of "a coarse understanding of the event as a whole" (Griffin & Bock, 2000, 274). This process leads to the construction of a conceptual representation of the utterance, which then guides the linguistic encoding of the sentence's individual words. The results from an eye tracking experiment by Bock et al. (2003) where participants were presented with different clock displays and their task was to tell the time in different formats in Dutch and English also support the hierarchically incremental accounts. Within a few hundred milliseconds after stimulus onset, speakers had parsed the clocks shown and directed their gaze towards parts in the displays that were relevant for planning the first part of their utterance, supporting the view that linguistic encoding is guided by an utterance representation generated at the outset of formulation (cf. also Kuchinsky et al., 2011). Thus, weakly hierarchically incremental accounts assume that speakers always engage in planning to express the event early by encoding relational information to determine either a rather specific action that is carried out (e.g., *kicking* or *shooting*; Griffin & Bock, 2000) or at least the kind of action or event class (e.g., physical contact event; Bunger et al., 2013).

The strong version of hierarchical incrementality additionally assumes that speakers must engage in some verb planning before articulation of the first words of a sentence can be initiated. Verb planning always entails that the event structure was encoded earlier because speakers need to know about the action and the relations among the participants of a to-be-described event. For example, in Bock and Levelt's (1994) model of sentence production, verbs play a central role in the planning process by controlling the assignment of syntactic functions, which is served by information contained in the verb lemma about the arguments that each verb takes.

Early evidence for advance verb planning comes from Lindsley (1975) who showed that English speakers needed more time to initiate subject-verb (SV) sentences than to name just the subject when describing pictures; reaction times were also longer for verb-only utterances compared to SV sentences in which the participants already knew the subject. Lindsley argues that these results suggest that verbs are at least partly planned before speech onset. In another picture description study, Kempen and Huijbers (1983) extended Lindsley's account by showing that speakers engage in lexical planning of verbs before speech onset when producing SV sentences.

F. Ferreira (2000) presents a sentence production model based on Tree-Adjoining Grammar. This model explicitly assumes that lexical selection of verbs is necessary before speakers can plan the first nominal elements of sentences. A verb must be selected before speakers can assign the subject function to one nominal lemma and articulate it sentence-initially because syntactic functions and case marking can only be assigned by verbs. These are assumed to be the syntactic heads of sentences, introducing the necessary structure that is needed to build a syntactic tree which allows the grammatical encoding of subject and object arguments. In support of this account, F. Ferreira (1994) found that the choice between active and passive voice, i.e., whether the agent or the patient of an event is encoded as the subject, depends on properties of the sentence's verb.

Thus, strongly hierarchically incremental accounts of sentence production assume that speakers always engage in relational encoding *and* select verbs early during formulation because syntactic function assignment depends on verb lemmas.

Two recent studies have dealt with the advance planning of verbs in verb-final sentences. The first study is Momma, Slevc, and Phillips

(2016), showing that in Japanese verbs are planned before objects. In a picture-word interference experiment, Japanese speakers described pictures with SV or OV sentences while seeing distractor words (superimposed on the pictures) that were either semantically related or unrelated to the verb. Longer speech onset latencies were found for related distractors when speakers produced OV sentences as compared to SV sentences. Momma et al. interpret this as evidence that verb selection occurs before objects but not before subjects because verbs and objects may be more closely associated syntactically than verbs and subjects (cf. Kratzer, 1996).

The second study, Hwang and Kaiser (2014), investigated whether verbs are planned earlier in English (SVO word order) than in Korean (SOV word order) in a combined picture-word interference and eye tracking experiment. Speakers of English and speakers of Korean described pictures of transitive events while hearing auditory distractors that were either unrelated, semantically related to the verb or semantically related to the object (patient). Hwang and Kaiser report longer speech onsets when the distractor was related to the verb in English but not in Korean. There were also differences in the fixation patterns for the two languages: In English, where the verb immediately followed the subject, speakers fixated early (400–600 ms after stimulus onset) on the "action" region, i.e., the part of the picture where agent and patient interacted or were in physical contact and which therefore "provide[s] crucial information about what action is being depicted" (Hwang & Kaiser, 2014, 1365). The authors interpret this to be consistent with strongly hierarchically incremental production (called the *lexicalist hypothesis* by Hwang and Kaiser), where verbs are selected already when an utterance plan is generated at the outset of formulation. In Korean, on the other hand, speakers' fixations to the "action" region only increased towards the end of sentences, when the verb was mentioned. This is consistent with weakly hierarchical incrementality which assumes that relational encoding always occurs early but that verbs are planned only shortly before they are articulated; it is, however, also consistent with linearly incremental word-by-word conceptual and linguistic encoding of the sentences. In essence, Hwang and Kaiser interpret their findings as evidence that a language's word order influences when speakers plan the lexical verb during sentence formulation.

The experiment reported here aims to test the centrality of verb planning in sentence production in German by looking at whether voice word order variations within in a single language may influence when speakers engage in relational encoding and verb planning.

Taking a somewhat different approach than most previous studies, the timing of relational encoding and verb planning during the formulation of German sentences with lexical main verbs placed in different positions is investigated in a simple picture description paradigm. Participants' eye movements were recorded while they spontaneously described pictures of transitive events and the temporal development of fixation preferences before and after speech onset is analyzed.[2]

In German independent declarative sentences, the inflected verb is placed in the second position, i.e., sentence-medially (V-medial). In present or past tense the lexical verb thus immediately follows the first constituent, which is the subject in the current context. However, when the lexical verb occurs as an infinitive or a participle, it is placed sentence-finally (V-final) because the second position is then occupied by another inflected verbal element, e.g., an auxiliary in perfect sentences (*haben* 'to have'), a matrix verb (e.g., *versuchen* 'to try') or modal verb (e.g., *sollen* 'to be supposed to'). Similarly, the lexical verb is also in final position in passive sentences because the second position is occupied by an auxiliary. Also, agents are not arguments but obliques introduced by prepositions in these sentences. These different sentence types are illustrated in Table 3.1.

[2] Pupillometric data from this experiment are reported in Chapter 4.

Table 3.1: German sentence types relevant for the current experiment

Active, V-medial	*Der*	*Junge*	*tritt*	*den*	*Ball.*		
	The	boy	kicks	the	ball		
	"The boy kicks the ball."						
Active, V-final	*Der*	*Junge*	*hat*	*den*	*Ball*	*getreten.*	
	The	boy	has	the	ball	kicked	
	"The boy has kicked the ball."						
	Der	*Junge*	*versucht*	*den*	*Ball*	*zu*	*treten.*
	The	boy	tries	the	ball	to	kick
	"The boy tries to kick the ball."						
Passive	*Der*	*Ball*	*wird*	*vom*	*Jungen*	*getreten.*	
	The	ball	is.being	by.the	boy	kicked	
	"The ball is being kicked by the boy."						

In the current experiment, German speakers described pictures showing transitive, two-participant events using active and passive sentences with V-medial and V-final word orders. Unlike in Hwang and Kaiser (2014), "action" regions are not used here. The reason for this is that it is generally difficult to define regions in the stimulus pictures that exclusively "belong" to the action. Instead, following Norcliffe, Konopka, et al. (2015), relational encoding and verb selection is assumed to manifest itself in the patterns of agent and patient fixations. In an eye-tracked picture description experiment on the Mayan language Tzeltal Norcliffe, Konopka, et al. showed that speakers distributed their fixations extensively between agents and patients before speech onset when planning verb-initial (verb-patient-agent) sentences. On the other hand, when Tzeltal speakers planned subject-initial (agent-verb-patient) sentences, they preferentially fixated the agent in a time window from the onset of the stimulus picture until speech onset. Norcliffe, Konopka, et al. took these differences to be indicative of early verb planning in verb-initial sentences in Tzeltal. It is thus assumed that the early planning of verbs goes in hand with extensively distributed fixations to agents and patients. Planning verbs requires the encoding of the event relations between referents and it is thus assumed that information about the "action" is mainly distributed between them. A similar approach is also taken by Griffin and Bock (2000) and Konopka and Meyer (2014) who interpret distributed gazes between agents and patients as being indicative of event encoding. Fixation preferences for only one referent before speech onset, or more evenly distributed gazes over both referents, are thus taken to signal different degrees of relational encoding effort and (in the latter case) as being indicative of verb selection.[3]

Given these points, the accounts of hierarchical and linear incrementality make different predictions about when speakers engage in relational event encoding and verb planning in the three different German sentence types in Table 3.1.

Linearly incremental production accounts assume that speakers do not necessarily start sentence planning by encoding the relations in the described event but may instead start to immediately encode the most accessible nominal concept, which is to be mentioned first (Gleitman et al., 2007). Thus, linear incrementality predicts that the conceptual acceptability or the visual salience of depicted characters influences speakers' choice between producing an active or a passive sentence. Animate or human characters are conceptually more accessible than inanimate or non-human characters and are predicted to be more likely to be chosen as subjects and produced first under this account (Branigan et al., 2008). Visually salient characters are also predicted to be more likely to be selected as subjects because speakers will fixate on them first due to their prominence in the pictures, leading to the encoding of the corresponding character names (Gleitman et al., 2007).

To ensure fluency in sentences with V-medial word order in which the lexical verb is mentioned immediately after the subject, speakers

[3]Note that hypotheses on speakers' eye movement behaviors refer to the patterns that can be observed after aggregating over trials.

should engage in extensive relational encoding before speech onset under a linearly incremental account as there might otherwise not be enough time to select and retrieve the verb while uttering the subject (Griffin, 2003). For the production of V-final sentences speakers may postpone relational and verb encoding until after speech onset because the lexical verb does not have to be ready for articulation so early. It is thus further predicted by linear incrementality that speakers distribute their attention more between agent and patient before speech onset when producing V-medial active sentences than when producing V-final actives and passives, where verb-processing can be delayed. Compared to V-medial sentences, more fixations on the subject (the agent in actives and the patient in passives) are expected for V-final sentences because speakers should prioritize lexical encoding of the first referent as they may postpone relational encoding until later.

By contrast, both the weak and the strong version of hierarchical incrementality assume that speakers always encode the event and the relations between referents early in the formulation process in order to generate a conceptual representation of the utterance that guides linguistic encoding. Both versions of hierarchical incrementality predict that the choice between active and passive sentences, and therefore which concept is lexicalized first, may also be influenced by the structure of the event and not only by the conceptual accessibility or visual salience of depicted characters. With respect to verb planning, weakly hierarchical incrementality assumes that speakers encode the action in the depicted event but do not necessarily select a verb early (Griffin & Bock, 2000; Konopka & Meyer, 2014). From this assumption the prediction follows that V-medial and V-final sentences should be associated with different fixation patterns. In order to ensure fluency in the former sentence type (Griffin, 2003), speakers would need to select a verb lemma and begin to linguistically encode the verb before speech onset because it is mentioned immediately after the subject. When formulating V-final sentences, however, speakers may postpone lexical verb selection until after speech onset. Thus, weakly hierarchical incrementality predicts that speakers distribute their visual attention more between agent and patient referents before speech onset during the planning of V-medial as compared to V-final sentences, indicating earlier verb selection in the former. The prediction of differing fixation patterns during the formulation of sentences with different verb positions in weakly hierarchically incremental planning is similar to what is predicted by linear incrementality. However, hierarchical incrementality makes a distinct prediction about the influence of accessibility on grammatical structure choices because this account assumes that the action in the event is encoded early during formulation. Thus, event relations among referents may also be taken into account when speakers choose between producing active or passive sentences.

The strong version of hierarchical incrementality states that relational encoding of the event also always includes or is immediately followed by the selection of a verb lemma, which guides syntactic function assignment (F. Ferreira, 2000). Therefore, this account predicts that the planning of V-medial and V-final actives as well as passive sentences should all be associated with similar eye movement patterns in which speakers distribute their fixations between agent and patient characters in the pictures, which is assumed to reflect the lexical selection of a verb.

The current experiment tests the predictions of the three different accounts of sentence production by analyzing the influence of agent and patient characters' conceptual accessibility and visual saliency on speakers' choice of active or passive voice when describing pictures of transitive events and by analyzing the patterns of fixations to the characters before and after speech onset using growth curve modeling (Mirman, 2014).

Figure 3.1: Example stimulus picture

3.2 METHODS

3.2.1 *Participants*

Thirty-three native speakers of German (mean age = 25 years, 10 male) recruited among the (PhD) students of Radboud University and HAN University of Applied Sciences in Nijmegen participated in the experiment. All participants were unaware of the hypotheses of the experiment.

The reported experiment conforms to the American Psychological Association's ethical principle of psychologists and code of conduct (as declared by the ombudsman of the Max Planck Institute for Psycholinguistics). At the time of data collection, ethical approval was not legally required for this kind of study. Written informed consent was obtained from participants at the beginning of the experiment session.

3.2.2 *Materials*

Colored line drawings of transitive and intransitive events were used as stimuli, including events with both human and non-human agents and patients (cf. Figure 3.1 for an example). The stimuli are overlapping with the stimuli set of Norcliffe, Konopka, et al. (2015). Target pictures were drawings of 58 transitive events[4], which were interspersed among 93 filler pictures of intransitive events. Two versions of each target were created by mirror-reversing the pictures. Stimuli were arranged in four lists created by randomizing the order and counterbalancing the two mirror-reversed versions of the targets; at least one filler intervened between any two targets.

3.2.3 *Apparatus and procedure*

Stimuli were presented on a Tobii T120 eye tracker (resolution = 1024 × 768 pixels, sampling rate = 60 Hz, distance to participant ≈ 58–60 cm). Eye data and vocal responses were recorded together by the Tobii Studio software. Fixations and saccades were defined by the Tobii I-VT filter (Olsen, 2012). Areas of interest (AOIs) were defined covering the agent and patient characters and a slim margin around them in the stimulus pictures (Holmqvist et al., 2011). Testing took place in a dimly lit and soundproof booth.

The participants' task was to describe the pictures in one sentence as quickly and accurately as possible, naming all depicted characters. Each stimulus was preceded by a fixation dot appearing in one out of five random positions at the top of the screen. When participants fixated on the dot, the experimenter (who monitored their gaze on the computer controlling the eye tracker) initiated the next trial, making sure that their gaze did not fall on the agent or patient when the stimulus appeared. The experiment started with a practice phase in which 15 example pictures with accompanying pre-recorded descriptions were presented to participants in order to familiarize

[4] Only pictures of transitive events were included as target items to ensure that all critical trials exhibited similar "relational complexities" in the sense of involving two participants, agent and patient. Momma et al. (2016, 820) acknowledge that the difference in transitivity between their SV (intransitive) and OV (transitive) conditions might have also contributed to planning differences.

them with the task. Next, the same pictures were presented again one at a time and participants were asked to describe them themselves. The testing phase followed in which stimuli were presented in three blocks lasting 8–10 minutes each. The eye tracker was calibrated before each block.

3.2.4 *Sentence scoring and data selection*

Utterances produced on target trials were transcribed and scored as active or passive sentences or as other constructions (e.g., existentials). For each target trial, the onset and offset of each word were annotated manually in Praat (Boersma, 2001). Only actives and passives with both referents realized overtly were included in the analyses. Trials in which participants corrected themselves were excluded. However, when the response contained disfluencies (like "uh") or pauses it was still included because these are normal phenomena occurring during spontaneous speaking. Trials were also excluded when the first fixation fell on the agent or patient instead of the fixation dot, if the first fixation to agent or patient occurred later than 600 ms after picture onset, and where two consecutive fixations were longer than 600 ms apart (indicating track loss). Additionally, trials where excluded if speech onset was longer than 6500 ms or longer than three standard deviations away from the grand mean. Two stimulus pictures did not elicit any responses conforming to these criteria and were thus excluded from the analyses. Two trials with semantic role reversals, i.e., where participants conceptualized the event so that the intended patient served as the agent and the intended agent as the patient, were included and recoded accordingly. The final data set consisted of 1207 sentences (active, V-medial: 922 sentences, 6 with disfluencies; active, V-final: 180 sentences [including 23 sentences with matrix verbs, cf. Table 3.1]; passive: 105 sentences, 1 with a disfluency).

3.2.5 *Analyses*

A generalized linear mixed effects regression model was employed to analyze which variables influenced voice choice. This model included as predictors the humanness of agent and patient (conceptual accessibility) and which referent was looked at first (visual saliency). Humanness was chosen over animacy as predictor because only few sentences with inanimate agents were included in the dataset.

To examine the time course of relational encoding and verb selection, the likelihoods of fixations to first and second referents (subjects and objects/obliques) were analyzed with logistic growth curve regression (Donnelly & Verkuilen, 2017; Mirman, 2014). Growth curve analysis is a variety of linear mixed effects regression that uses orthogonalized polynomial time terms as predictors to describe the major aspects of the observed fixation curve shapes.

The overall development of gaze patterns was assessed in two analysis time windows. Within each time window, fixations were aggregated into 200 ms bins for each trial to reduce statistical non-independencies in eye tracking data (temporal autocorrelation; cf. Barr, 2008). These non-independencies result from the fact that eye gaze cannot change location instantaneously but must rather "travel through time and space" (Barr, 2008, 464), making participants' eye movement behavior at one time step highly correlated with that at the next time step. Aggregation of fixations into trial-wise time bins helps to filter "out the eye-movement based dependencies" (Barr, 2008, 464). Additionally, for each time bin in each trial, the number of samples with fixations to the agent or the patient in the previous time bin was included as a nuisance variable (cf. Sassenhagen & Alday, 2016) with the aim of further reducing temporal autocorrelation which is due to the fact that eye movements are relatively slow as compared to the sampling rate of the eye tracker (Barr, 2008; Duchowski, 2007). Fixation likelihoods were calculated on the basis of all fixations to agent and patient AOIs as well as fixations to "whitespace" (Holmqvist et al., 2011) outside of these AOIs.

The first analysis time window spanned the time between presentation of the stimulus and grand mean speech onset (100–1700 ms, grand mean speech onset = 1712 ms). Fixations during the first 100 ms were not included because eye movements in response to stimulus presentation are unlikely to occur this early (Duchowski, 2007). The second time window reached from grand mean speech onset until 200 ms before grand mean speech offset by which most planning should have been finished (1700–3500 ms; grand mean speech offset = 3732 ms).

Sentence Type and fourth-order orthogonalized polynomial time terms were included as predictors in all regressions (Mirman, 2014; Mirman, Dixon, & Magnuson, 2008). Each of the included polynomial time terms describes a different aspect of the eye movement data (cf. Kalénine, Mirman, Middleton, & Buxbaum, 2012). Linear time ($Time^1$) describes the angle or the slope of fixation curves, with more positive predictor estimates indicating a steeper increase over the course of the analysis time window. Quadratic time ($Time^2$) describes the rate of increases and decreases in the form of a parabolic curve, where more positive estimates indicating a more "U-shaped" curve and more negative estimates describing curves with "inverted U-shapes". Cubic time ($Time^3$) describes earlier or later increases or decreases of the fixation curves, i.e., how "S-shaped" the fixation curves are. Finally, quartic time ($Time^4$) describes secondary peaks in the curves' tails, at the beginning or end of the analysis time window. A main effect of Sentence Type in the growth curve regression models means that one sentence type exhibited overall higher or lower fixation likelihoods to a character in the analysis time window than the other sentence type. Interaction effects between Sentence Type and the polynomial time terms mean that fixation likelihoods changed differently over the course of the analysis time window in different sentence types, i.e., that the fixation curves exhibit different shapes.

Two further nuisance variables and their interactions with the time terms were included in order to control for their effects on speakers' likelihood to fixate on subject and object characters during sentence formulation. First, speech onset latencies were included to account for variations in fixation patterns that are reducible to sentence formulation processes with different timing across trials, independently of the produced sentence type. For example, if the speech onset in one trial is earlier than the grand mean of 1712 ms, the speaker might look away from the subject character already earlier than in trials with longer speech onsets, leading to potentially different fixation curves.

Second, event codability was also included as a nuisance variable. This variable models to what degree speakers use the same verbs to describe an event, which is related to the difficulty of recognizing the depicted action (Kuchinsky, 2009). It is characterized as reflecting "consensus […] about the conceptual structure of an event" by van de Velde et al. (2014, 125). In other words, the codability describes how "easy" or "hard" it was for speakers to find an appropriate verb to name the depicted event. By manipulating event codability in a picture description experiment, van de Velde et al. show that speakers distributed their visual attention more between agent and patient if the event was highly codable (i.e., it is "easy" to name the verb) but focused more in the first-mentioned character when event codability was low. The aim of the current experiment is to test differences in eye movement patterns in order to examine whether planning time courses differences between sentence types with different voice marking and word order. Including event codability as nuisance variable in the regression models accounts for potential differences in fixation patterns that might be explained solely by different planning strategies employed for highly and lowly codable events. To assess the event codability of each stimulus picture, the Shannon entropy

H (Shannon, 1948) was computed, describing the variability in which verbs participants used in their responses. If it was difficult to recognize the event depicted in a picture and participants therefore produced many different verbs, that picture's H was larger (low codability) than that of a picture where participants highly agreed on which verb to choose (high codability).

Speech onset latencies were analyzed using a linear mixed effects regression model predicting log-transformed reaction times. In addition to the predictors used in the analysis of voice choice (humanness of agent and patient characters and the identity of the first-fixated character), sentence type was a further predictor in this model. Event codability was also included as a nuisance variable in order to control for effects on speech onset latencies caused the relative ease of finding a verb to describe the event.

Significance of fixed effects was assessed with Wald Z tests in generalized linear regression models (Agresti, 2007; Jaeger, 2008) and with Type II Wald F-tests with Kenward-Roger approximation of degrees of freedom in the linear regression model for speech onsets (Halekoh & Højsgaard, 2014; Kenward & Roger, 1997). Where these variables were included, sentence type was Helmert-coded and all other categorical variables were contrast-coded. The maximal random effects structure justified by design (that allowed convergence) was used for all models (Barr, 2013; Barr, Levy, Scheepers, & Tily, 2013). Categorical predictors were contrast-coded (Cohen, Cohen, West, & Aiken, 2003). All models were computed using the *lme4* package in R (Bates et al., 2015; R Core Team, 2015). Graphs were produced using *ggplot2* (Wickham, 2009).

3.3 RESULTS

Figure 3.2 shows the proportion of active sentences produced as a function of the humanness of agent and patient characters. In general,

Table 3.2: Results from binomial generalized linear mixed effects regression model predicting voice choice

	β	$\lvert Z \rvert$	p	
Intercept	-5.69	6.61	< 0.001	***
Agent Humanness (= non-human)	0.19	0.24	0.81	
Patient Humanness (= non-human)	-7.92	5.88	< 0.001	***
Agent Humanness × Patient Humanness	-6.04	3.57	< 0.001	***
First-Fixated Character (= patient)	0.73	1.54	0.12	

participants were more likely to produce passives when the patient was human (main effect of Patient Humanness in Table 3.2). In addition, passives were more likely to be produced in the descriptions of stimulus pictures with non-human agents and human patients. The identity of the first-fixated character in the pictures had no statistically significant influence on voice choice.

Figure 3.3 shows the distribution of speech onset latencies for each sentence type. Speech onset latencies were numerically shorter for V-medial actives (mean = 1662 ms, *SD* = 458 ms) than for V-final actives (mean = 1859 ms, *SD* = 504 ms) and passives (mean = 1912 ms, *SD* = 722 ms). However, there were no statistically significant effects on speech onset latencies (all *p*s > 0.14, Table 3.3). Sentence types did also not differ with respect to the onset of the verb or auxiliary after the subject or the onset of the second NP (all *p*s > 0.23 in a linear mixed effects regressions with Sentence Type as predictor, cf. Tables 3.A.1 and 3.A.2 in the Supplementary Materials). This indicates that the occurrence of pauses was not systematically related to voice or word order in the current dataset. Thus, it is justified to compare participants' eye movements during the production of V-medial actives, V-final actives and passive sentences in the same analysis time windows because speech onset times, phrase durations

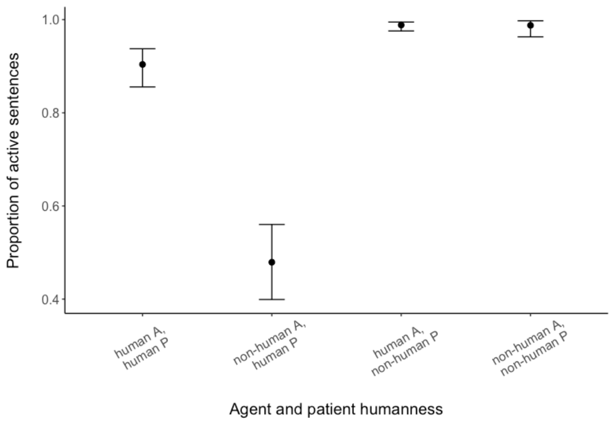

Figure 3.2: Proportions of active sentences as a function of agent (A) and patient (P) humanness. Bars indicate 95% confidence intervals (Agresti & Coull, 1998).

and the distribution of pauses and disfluencies was similar across sentence types.

In all sentence types (actives with V-medial and V-final word order and passives), speakers concentrated their gazes on the first mentioned referent until shortly before speech onset when they switched to the second mentioned referent (Figure 3.4; cf. also Figure 3.C.1). Towards the end of the sentences, proportions of fixations to the two referents approximated each other (cf., e.g., Griffin & Bock, 2000, Konopka & Meyer, 2014 and Norcliffe, Konopka, et al., 2015 for similar fixation patterns). Despite these overall similarities, speakers' gaze behavior differed between sentence types.

When planning and producing passive sentences, speakers distributed their visual attention more evenly between characters then when planning active sentences in general. Before speech onset, fixations to the subject character (agent in actives, patient in passives) showed an earlier and steeper increase and decrease before speech onset in passives (interactions of Sentence Type and Time^2 and

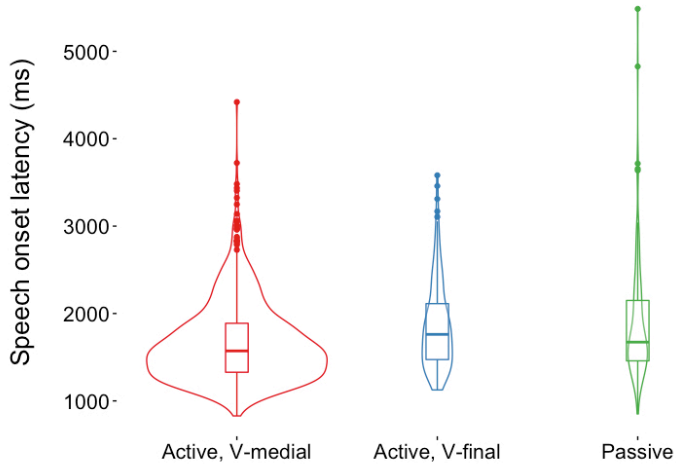

Figure 3.3: Densities and box plots of speech onset latencies (relative to stimulus picture onset) for three German sentence types (Hintze & Nelson, 1998); width of the violins is proportional to the number of underlying data points.

Table 3.3: Results from linear mixed effects regression model predicting log-transformed speech onset latencies

	β	$\|t\|$	F statistic	p
Intercept	7.45	230.54		
Agent Humanness (= non-human)	> -0.01	0.07	$F(1, 41) = 0.03$	0.87
Patient Humanness (= non-human)	-0.03	1.00	$F(1, 57) = 0.69$	0.41
Agent Humanness × Patient Humanness	> -0.02	0.29	$F(1, 30) = 0.06$	0.81
Actives vs. Passives	-0.04	0.97	$F(2, 27) = 2.08$	0.14
V-final Actives vs. V-medial Actives	0.05	1.93		
First-Fixated Character (= patient)	-0.02	1.34	$F(1, 31) = 1.49$	0.23
Event codability (= standardised)	0.01	1.03	$F(1, 33) = 0.69$	0.41

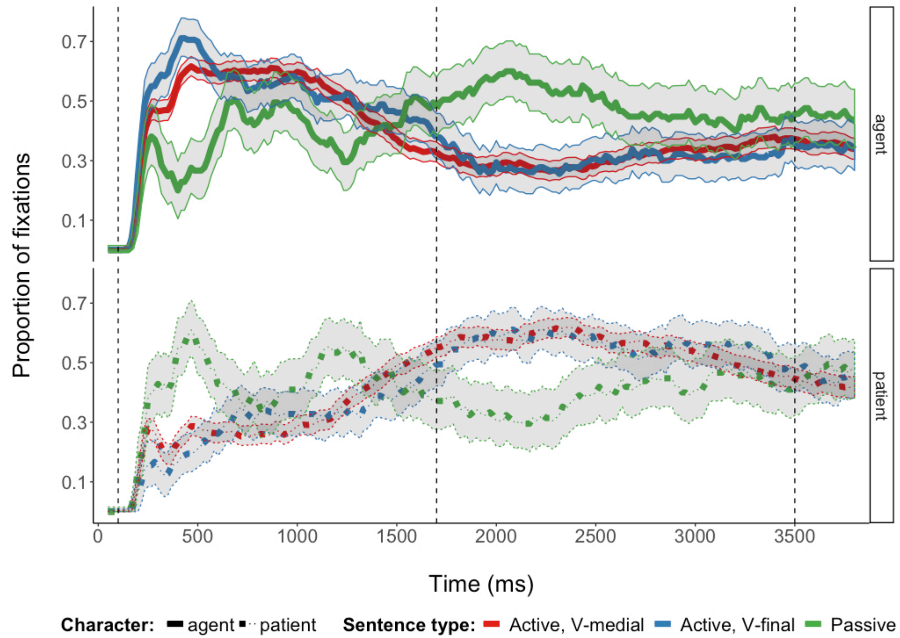

Figure 3.4: Proportions of fixations to agents and patients during the production of three German sentence types. Proportions are based on fixations to agent and patient AOIs and on "whitespace" (Holmqvist et al., 2011) not covered by these AOIs. Ribbons indicate 95% multinomial confidence intervals (Sison & Glaz, 1995; Villacorta, 2012); vertical lines indicate analysis time windows.

Table 3.4: Results from binomial generalized linear mixed effects regression models predicting subject and object/oblique fixations in V-medial actives, V-final actives and passive sentences

	Subject fixations				Object/oblique fixations			
100–1700 ms								
Intercept	-0.59	4.09	< 0.001	***	-1.89	10.32	< 0.001	***
Time^1	-1.21	2.33	0.02	*	2.38	4.39	< 0.001	***
Time^2	-2.98	5.09	< 0.001	***	-1.11	2.54	0.01	*
Time^3	-1.27	2.31	0.02	*	3.06	6.29	< 0.001	***
Time^4	-1.23	3.92	< 0.001	***	-0.58	1.54	0.12	
Actives vs. Passives	-1.19	3.69	< 0.001	***	1.06	3.06	< 0.01	**
V-final Actives vs. V-medial Actives	-0.21	0.79	0.43		-1.43	4.18	< 0.001	***
Time^1 × Actives vs. Passives	0.69	0.61	0.55		1.90	1.32	0.18	
Time^1 × V-final Actives vs. V-medial Act.	-1.58	1.64	0.10		1.86	2.08	0.04	*
Time^2 × Actives vs. Passives	-2.73	2.22	0.03	*	1.61	1.18	0.24	
Time^2 × V-final Actives vs. V-medial Actives	-2.21	1.97	< 0.05	*	-1.76	1.72	0.09	
Time^3 × Actives vs. Passives	-3.47	2.47	0.01	*	2.75	2.62	< 0.01	**
Time^3 × V-final Actives vs. V-medial Actives	-1.14	1.79	0.07		2.18	2.29	0.02	*
Time^4 × Actives vs. Passives	-0.02	0.04	0.96		1.27	1.46	0.14	
Time^4 × V-final Actives vs. V-medial Actives	-1.54	2.80	< 0.01	**	-0.25	0.35	0.72	
Fixations to AOI in previous time bin	0.19	101.74	< 0.001	***	0.21	97.39	< 0.001	***
Speech Onset Latency (standardized)	0.28	27.49	< 0.001	***	-0.25	19.46	< 0.001	***
Time^1 × Speech Onset Latency	0.88	26.80	< 0.001	***	-0.77	20.49	< 0.001	***
Time^2 × Speech Onset Latency	0.31	9.46	< 0.001	***	-0.41	11.14	< 0.001	***
Time^3 × Speech Onset Latency	-0.49	14.94	< 0.001	***	0.49	13.49	< 0.001	***
Time^4 × Speech Onset Latency	-0.41	13.05	< 0.001	***	0.39	11.05	< 0.001	***
Event Codability (standardized)	-0.06	1.19	0.23		-0.08	1.42	0.16	
Time^1 × Event Codability	-0.28	2.72	< 0.01	**	0.46	2.67	< 0.01	**
Time^2 × Event Codability	0.03	0.39	0.70		-0.37	1.90	0.06	
Time^3 × Event Codability	-0.07	0.58	0.56		0.33	2.32	0.02	*
Time^4 × Event Codability	0.44	4.60	< 0.001	***	-0.56	4.72	< 0.001	***

Table 3.4: (continued) Results from binomial generalized linear mixed effects regression models predicting subject and object/oblique fixations in V-medial actives, V-final actives and passive sentences

1700–3500 ms	Subject fixations				Object/oblique fixations			
Intercept	-1.52	6.35	< 0.001	***	0.11	0.44	0.66	
$Time^1$	1.61	2.50	0.01	*	-1.59	2.80	< 0.01	**
$Time^2$	-1.44	2.20	0.03	*	1.57	2.36	0.02	*
$Time^3$	1.88	3.19	< 0.01	**	-0.73	1.30	0.19	
$Time^4$	1.48	3.34	< 0.001	***	0.82	2.08	0.04	*
Actives vs. Passives	0.33	0.51	0.61		-1.35	2.09	0.04	*
V-final Actives vs. V-medial Actives	-0.92	2.97	< 0.01	**	0.52	1.06	0.29	
$Time^1$ × Actives vs. Passives	3.50	1.83	0.07		-3.38	2.26	0.02	*
$Time^1$ × V-final Actives vs. V-medial Act.	0.07	0.08	0.94		0.36	0.39	0.69	
$Time^2$ × Actives vs. Passives	-2.53	1.34	0.18		3.52	2.02	0.04	*
$Time^2$ × V-final Actives vs. V-medial Actives	-1.09	1.96	0.05		0.99	1.00	0.32	
$Time^3$ × Actives vs. Passives	5.22	3.11	< 0.01	**	-0.97	0.68	0.50	
$Time^3$ × V-final Actives vs. V-medial Actives	0.38	0.49	0.63		-0.81	0.76	0.45	
$Time^4$ × Actives vs. Passives	3.00	2.59	< 0.01	**	-1.81	1.79	0.07	
$Time^4$ × V-final Actives vs. V-medial Actives	0.57	0.98	0.33		-0.55	1.00	0.32	
Fixations to AOI in previous time bin	0.30	164.74	< 0.001	***	0.29	173.84	< 0.001	***
Speech Onset Latency (standardized)	-0.04	3.25	< 0.01	**	0.11	8.88	< 0.001	***
$Time^1$ × Speech Onset Latency	-0.49	12.84	< 0.001	***	0.60	16.93	< 0.001	***
$Time^2$ × Speech Onset Latency	0.32	8.63	< 0.001	***	-0.26	7.28	< 0.001	***
$Time^3$ × Speech Onset Latency	-0.05	1.45	0.15		0.07	2.30	0.02	*
$Time^4$ × Speech Onset Latency	-0.01	0.28	0.78		< 0.01	0.03	0.98	
Event Codability (standardized)	0.15	2.69	< 0.01	**	-0.10	1.43	0.15	
$Time^1$ × Event Codability	0.12	0.99	0.32		-0.26	2.58	< 0.01	**
$Time^2$ × Event Codability	0.11	1.30	0.19		-0.03	0.34	0.73	
$Time^3$ × Event Codability	0.16	1.36	0.17		0.15	1.04	0.30	
$Time^4$ × Event Codability	-0.11	1.21	0.23		0.20	1.91	0.06	

Time^3 in 100–1700 ms time window, Table 3.4). Speakers' fixations to object/oblique characters (patient in actives, agent in passives) also increased earlier in passive sentences than in actives (interaction of Sentence Type and Time^3 in 100–1700 ms time window). After speech onset, the more even distribution of visual attention between subject and object/oblique characters is revealed by overall fewer and more quickly declining looks to the agent in passives as compared to the patient in actives (main effect of Sentence Type and interactions between Sentence Type and Time^1 and Time^2 in 1700–3500 ms time window, regression model for object/oblique fixations in Table 3.4). Fixations to the subject character increased earlier again after speech onset in passives as compared to actives (interactions between Sentence Type and Time^3 and Time^4 in 1700–3500 ms time window).

Crucially, there were also differences in fixation patterns for the planning and formulation of active sentences with different word orders. During the production of V-medial actives, where the verb was mentioned immediately after the subject, speakers looked at the subject (agent) with a steeper increase and decrease before speech onset than in V-final actives (interaction of Sentence Type and Time^2 in 100–1700 ms time window, Table 3.4). In addition, there were slightly more subject fixations at the beginning of the analysis time window in V-medial actives, which is captured by the significant interaction between Sentence Type and Time^4. Visual inspection of the fitted values from the regression model supports this interpretation of the quadratic and quartic differences between V-medial and V-final actives (cf. Figure 3.D.1). After speech onset, there were overall more fixations directed towards the subject character when speakers produced V-final sentences. Actives with V-medial and V-final word orders differed also with respect to fixations to the object (patient) characters before speech onset. Object fixations increased earlier but with a flatter curve shape for V-medial actives than for V-final actives before speech onset (interactions of Sentence Type and Time^2 and

Time[3] in 100–1700 ms time window, Table 3.4). Active sentence types did not differ in object fixations after speech onset.

3.4 DISCUSSION AND CONCLUSIONS

The current experiment yielded three main findings: Firstly, there were differences in gaze patterns for active sentences with V-medial and V-final word order. Secondly, German speakers' fixation behavior before speech onset also differed for the production of active and passive sentences. Thirdly, the choice of passives was largely determined by patient humanness.

The fixation differences that were found between V-medial and V-final actives suggest that verb planning was influenced by the position of the verb in the sentence. When the verb was in an early position and thus would be articulated immediately after the subject in V-medial actives, German speakers distributed their attention more between agent (subject) and patient (object) before speech onset. Put differently, speakers looked more at the subject and less at the object before speech onset in V-final actives as compared to V-medial actives, suggesting that they concentrated more on lexical encoding of the subject before speech onset in verb-final sentences.[5] This finding

[5]Based on the results of the growth curve analysis (interaction between sentence type and the quadratic and quartic time terms) and the proportions of fixations (Figure 3.4) it may be speculated that the (arguably small) differences between V-medial and V-final actives are manifested at the beginning and the end of the 100–1700 ms analysis time window, defined by the presentation of the stimulus picture and grand mean speech onset (cf. also Figures 3.C.1 and 3.D.1 in the Appendix). One possible explanation for this is that the more evenly distributed fixations among agents and patients for V-medial actives shortly after stimulus presentation reflect that speakers primarily engaged in relational encoding and verb selection to prepare the verb early in the production of these sentences, deciding very quickly on what kind of sentence they would produce. After these initial differences (presumably resulting from earlier verb planning), speakers might have concentrated on the encoding of the subject character in both active sentence types. The fixation proportions in Figure 3.4 suggest that this was followed again by slightly more distributed agent (subject) and patient (object) fixations in V-medial sentences; this can be interpreted as a consequence of speakers' need to begin preparing the word forms of verbs already before speech onset when they

is incompatible with the predictions made by strongly hierarchical incrementality (Bock & Levelt, 1994; F. Ferreira, 2000). If verb selection was a prerequisite to assign syntactic functions and prepare the sentence's subject, speakers should have distributed their visual attention among agents and patients in similar ways in all sentence types, independently of the position of the verb. The differences in fixation patterns between V-medial and V-final actives are compatible with linear incrementality, which predicted that sentence planning proceeds word-by-word and thus that verbs are only planned shortly before they will be articulated (Gleitman et al., 2007). They are also compatible with weakly hierarchical incrementality, which predicted that first a structural-relational utterance representation is generated that guides linguistic encoding, which is carried out in the order in which words will be uttered. This leads sentence-medial verbs to be prepared earlier than sentence-final verbs (Griffin & Bock, 2000; Kuchinsky et al., 2011).

When formulating passives, speakers distributed their visual attention even more between agent and patient before speech onset than when formulating active sentences. Figure 3.4 shows that speakers first primarily fixated on the patient referent (the subject of these sentences) before fixating on the agent (realised as oblique); towards the end of the 100–1700 ms time window the patient was again fixated primarily by the participants. This suggests that planning passives required more relational encoding than active planning and the reason for this could be that passives often describe non-prototypical animacy configurations in which a human is acted upon, that they are less frequent, and that the planning operations involved potentially differ from those of actives (cf. Chapter 4). Potentially different planning operations between actives and passives might also account for why speakers still distributed their visual attention more evenly

were mentioned immediately after the subject to avoid disfluencies (cf. Griffin, 2001, 2003), whereas in V-final sentences this was not necessary before speech onset.

between characters in passives also after speech onset, maybe because the affordances of preparing passivized verb forms.

Additionally, the humanness of patient characters as well as the patient and the agent characters' relative humanness influenced the choice between the production of active and passive voice marking. This shows that speakers did not simply assign the subject function to conceptually highly accessible human referents but analyzed the depicted event and the semantic roles and humanness of referents early in the formulation process in order to guide their structural choices (cf. Dobel et al., 2007; Hafri et al., 2013). In a picture description experiment on German, van Nice and Dietrich (2003) also found that agent and patient animacy influenced the choice between active and passive sentences. The current effect of humanness, which is contingent on semantic roles, supports hierarchically incremental accounts of sentence production (Kuchinsky et al., 2011), which propose that planning starts with the generation of an utterance plan that includes the relations among event participants. The effect is thus incompatible with linearly incremental accounts (cf. Branigan et al., 2008).

Altogether, the current results support weakly hierarchically incremental accounts of sentence production (Griffin & Bock, 2000; Konopka & Meyer, 2014). The differences in fixation patterns between V-medial and V-final active sentences suggest that verbs are only selected early when they were mentioned immediately after the subject and the finding that the choice of active versus passive was primarily driven by the humanness of the patient (and not just by humanness in general) indicates that speakers always encoded the event early to assess the semantic roles of the depicted characters. Thus, while speaking may start without the selection of a verb lemma (Iwasaki, 2011), the formulation of sentences still appears to depend on a representation of the described event instead of being solely driven by the retrieval of individual words (Norcliffe, Konopka, et al., 2015). These representations may be sufficient to assign syntactic functions, without the need to completely encode the verb first (in the case of verb-final active sentences). In general, differences in fixation patterns between active and passive sentences arose shortly after the stimulus pictures were presented, indicating that participants decided on the structure of the to-be-produced sentence early in the formulation process (Griffin & Bock, 2000).

The differences in the timing of verb planning between V-medial and V-final active sentences in German are similar to the differences that Hwang and Kaiser (2014) found between English and Korean. A weakly hierarchically incremental production account can explain the finding that speakers generated a representation of the event early but at the same time only engaged in additional extensive relational encoding and verb planning early when the verb was positioned sentence-medially.

It is an open question whether the event representations that German speakers appeared to have generated at the outset of sentence formulation to choose between active and passive sentence structures are "impoverished" or whether they are homologous to the utterance plans that are assumed to be generated at the beginning of the planning process in accounts of hierarchical incrementality (Griffin & Bock, 2000). The event representations must minimally contain information about the semantic roles of event participants and about their humanness. Utterance plans, however, are also assumed to contain more detailed information about the type of the event and a structural representation of the sentence under production (Bock & Ferreira, 2014).

In sum, the current experiment provides a temporally fine-grained view of verb planning in unscripted German sentence production, suggesting that the generation of an event representation is a necessary pre-requisite to start speaking, but not the retrieval of a verb, especially when it is positioned sentence-finally. The results reported here are consistent with the findings of Schriefers et al. (1998) on German and Momma et al. (2016) on Japanese, who demonstrated that speakers do not have to select sentence-final verbs before they can initiate the articulation of subjects.

The scope and time course of sentence planning may be influenced by many factors, ranging from time pressure (F. Ferreira & Swets, 2002) and speakers' working memory capacity (Swets et al., 2014) to prior knowledge about the event and ease of event encoding (Konopka & Kuchinsky, 2015; van de Velde et al., 2014). Here, it was shown that just as differences in grammars may lead to different planning preferences across languages (Norcliffe & Konopka, 2015), word order and voice variations can also influence the timing of relational encoding and verb planning within a language.

APPENDIX

3.A ANALYSES OF ONSET LATENCIES OF VERBS/AUXILIARIES FOLLOWING THE SUBJECT AND OF SECOND NPS

Table 3.A.1: Results from linear mixed effects regression model predicting log-transformed onset latencies of verbs (V-medial actives) or auxiliaries (V-final actives, passives) following the subject

	β	$\|t\|$	F statistic	p
Intercept	7.87	242.51		
Actives vs. passives	0.03	1.01	$F(2, 24)$ =	0.23
V-final actives vs. V-medial actives	-0.04	1.46	1.55	

Table 3.A.2: Results from linear mixed effects regression model predicting log-transformed onset latencies of second NP (patient in actives and agent in passives)

	β	$\|t\|$	F statistic	p
Intercept	7.75	238.86		
Actives vs. passives	-0.02	0.92	$F(2, 23) =$ 0.49	0.62
V-final actives vs. V-medial actives	0.02	0.60		

3.B Target stimulus pictures

* – primarily elicited V-medial active sentences; ¶ – primarily elicited V-final active sentences; † – primarily elicited passive sentences; stimulus pictures in parentheses were excluded from analyses

1 Ambulance car colliding with a woman†
2 Baker kneading bread dough*
3 Bird pulling a worm out of the ground*
4 Boxer beating a man*
5 Boy breaking branch from a tree*
6 Boy catching a frog*
7 Boy eating corn*
8 Boy kicking a football*
9 Boy kicking against a rock*
10 Boy stirring in a soup*
11 Bull attacking a girl*
12 Cat catching a mouse*
13 Cat scratching a girl's knee*
14 Construction worker losing his hat*
15 Cowboy catching a bull with a lasso*
16 Crocodile biting into a man's leg*
17 Dog chasing a mailman†
18 Dog chasing a squirrel*
19 Dog sniffing on a mandarin*
20 Frog eating a fly*
21 Gardener planting a tree*
22 Girl hanging out laundry*
23 Girl opening a door*
24 (Girl playing with a jumping rope)
25 Girl pushing a boy*

26 (Girl running towards an open door)
27 Girl tripping a construction worker*
28 Lion eating a dead zebra¶
29 Man angling a fish¶
30 Man breaking a piece of wood with a hammer¶
31 Man chopping a log of wood¶
32 Man cutting wood*
33 Man kicking against a chair*
34 Man leaving a hut*
35 Man pushing a car*
36 Man throwing a baby up in the air*
37 Monkey holding a crab in its hand*
38 Mosquito stinging a football player†
39 Mouse nibbling on a chocolate bar*
40 Nurse washing a baby*
41 Old man opening a window*
42 Old man reading a book*
43 Old woman climbing up the stairs*
44 Owl carrying a bag*
45 Paper boy selling newspapers*
46 Police officer arresting a man*
47 Police officer stopping a sports car*
48 Police officer stopping a walker-by*
49 Pupil raising his hand*
50 Rabbit eating a carrot*
51 Sailor drinking from a bottle*
52 Sheep eating leaves from a bush*
53 Soldier shooting a man*
54 Train colliding with a bus*
55 Veterinarian examining a horse's hoof*
56 Woman lifting a rug*
57 Woman looking inside a basket*
58 Woman walking across a bridge*

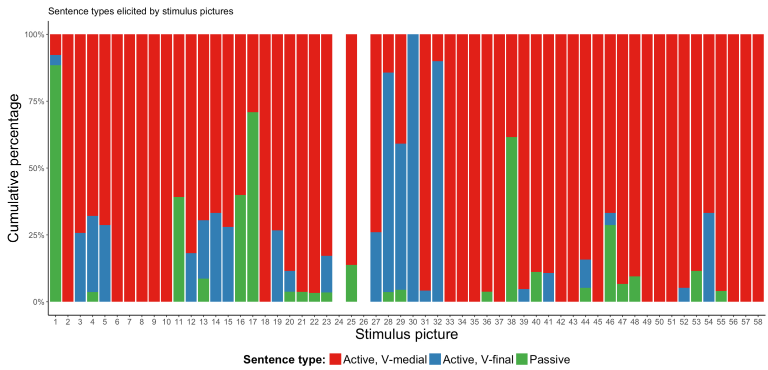

Figure 3.B.1: Proportions of sentence types elicited by stimulus pictures.

3.C PROPORTIONS OF FIXATIONS

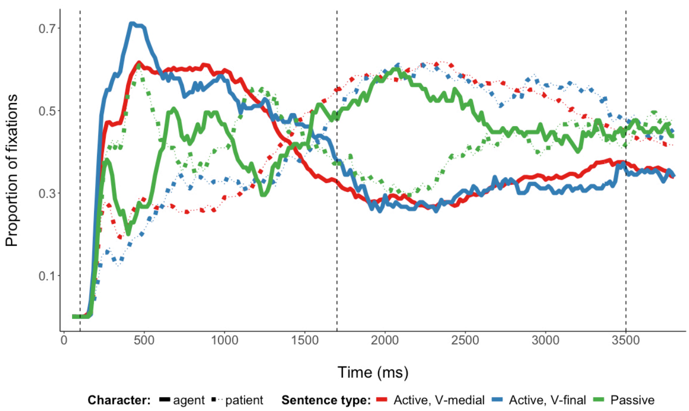

Figure 3.C.1: Proportion of fixations to agents and patients during the production of three German sentence types. Vertical lines indicate analysis time windows.

3.D MODEL FITS

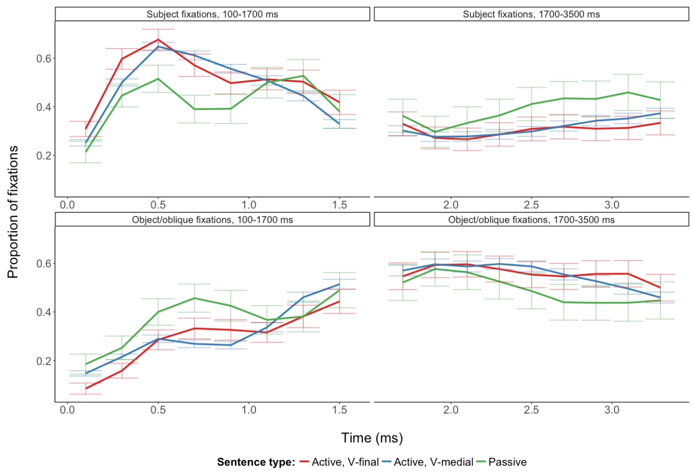

Figure 3.D.1: Mean fitted values from logistic mixed effects regression models predicting fixations on subject and object/oblique characters in three German sentence types. Error bars indicate one standard error of the mean fitted values.

4

SYMMETRICAL AND ASYMMETRICAL VOICE SYSTEMS AND PROCESSING LOAD: PUPILLOMETRIC EVIDENCE FROM SENTENCE PRODUCTION IN TAGALOG AND GERMAN

Sauppe, S. (2017). Symmetrical and asymmetrical voice systems and processing load: Pupillometric evidence from sentence production in Tagalog and German. *Language*, 93(2):288–313. doi: 10.1353/lan.2017.0015

ABSTRACT

The voice system of Tagalog has been proposed to be symmetrical in the sense that there are no morphologically unmarked voice forms. This stands in contrast to asymmetrical voice systems which exhibit unmarked and marked voices (e.g., active and passive in German). This paper investigates the psycholinguistic processing consequences of the symmetrical and asymmetrical nature of the Tagalog and German voice systems by analyzing changes in cognitive load during sentence production. Tagalog and German native speakers' pupil diameters were recorded while they produced sentences with different voice markings. Growth curve analyses of the shape of task-evoked pupillary responses revealed that processing load changes were similar for different voices in the symmetrical voice system of Tagalog. By contrast, actives and passives in the asymmetrical voice system of German exhibited different patterns of processing load changes during sentence production. This is interpreted as supporting

the notion of symmetry in the Tagalog voice system. Mental effort during sentence planning changes in different ways in the two languages because the grammatical architecture of their voice systems is different. Additionally, an anti-Patient bias in sentence production was found in Tagalog: cognitive load increased at the same time and at the same rate but was maintained for a longer time when the patient argument was the subject, as compared to agent subjects. This indicates that while both voices in Tagalog afford similar planning operations, linking patients to the subject function is more effortful. This anti-Patient bias in production adds converging evidence to "subject preferences" reported in the sentence comprehension literature.

4.1 INTRODUCTION

The grammatical voice systems found in the world's languages are often categorized into nominative-accusative, ergative-absolutive and active-inactive systems, among others. Some Austronesian languages have long been known to defy a classification into one of the commonly found systems, which led to the proposal to extend the typology of voice systems by introducing a distinction between "asymmetrical" voice systems, which exhibit a distinction between unmarked and marked voices, and "symmetrical" voice systems, in which all voices are equally marked (Foley, 2008; Riesberg, 2014b). In this paper I will first introduce key properties of these two kinds of voice systems and give a brief review of the arguments for the postulation of symmetrical systems. I will then present two sentence production experiments that investigate the psycholinguistic processing consequences of the grammatical architectures of asymmetrical and symmetrical voice systems. The first experiment focuses on Tagalog (Austronesian) as a representative of the symmetrical voice system type, the second experiment focuses on German (Indo-European) as a representative of the asymmetrical type. Changes in processing load over time during the planning and production of sentences will be used to investigate whether the grammatical distinction between asymmetrical and symmetrical voice systems is also reflected in psycholinguistic processes.

4.1.1 *Extending the typology of voice systems*

When discussing the position of "Philippine-type" languages in the typology of voice systems, Foley (2008) suggests extending it by introducing a distinction between asymmetrical and symmetrical systems (cf. Himmelmann, 2005a). This distinction is mainly based on the observation that the more familiar nominative-accusative and ergative-absolutive voice systems exhibit an unmarked voice that is the default choice when describing transitive events and a marked voice that is syntactically less transitive and involves additional overt marking to signal that a non-default voice was chosen. Foley observes that in Philippine-type languages, on the contrary, there is no unmarked voice and all voices are equally marked morphologically (e.g., by affixes carried by the verbal predicate). Below, I briefly discuss each of these systems in turn.

4.1.1.1 *German as an example of asymmetrical voice systems*

In nominative-accusative voice systems as found in, e.g., Germanic languages, the unmarked voice to describe a transitive event is the active. In this voice, the agent argument is the syntactic subject and carries nominative case, and the patient argument functions as syntactic object and is assigned accusative case, as in the German example in (4.1)[1]. The mapping between semantic roles and syntactic

[1]The following abbreviations are used in glosses: AV = agent voice, BV = beneficiary voice, CORE = core non-pivot argument, IPFV = imperfective aspect, IRR = irrealis mood, IV = instrumental voice, LV = locative voice, PTCP = participle, PFV = perfective aspect, PV = patient voice, PVT = pivot argument, RLS = realis mood, SG = singular. Glosses of example sentences cited from other sources are sometimes adapted to make them coherent throughout the paper.

functions is different in passives, where the patient argument is mapped to the subject function.

(4.1) a.

Der	Mann	fängt	den	Fisch.
the.NOM	man	catches	the.ACC	fish

"The man catches the fish."

b.

Der	Fisch	wird	von	dem	Mann
the.NOM	fish	AUX	by	the.DAT	man
gefangen.					
catch:PTCP					

"The fish is being caught by the man."

Voice alternations in nominative-accusative systems usually involve a transitivity alternation. When a marked mapping between semantic roles and syntactic functions is expressed in nominative-accusative systems, the verb is de-transitivized so that the patient can be the syntactic subject. Additional marking is also often required to indicate this non-default mapping between semantic roles and syntactic functions. In German this is achieved by placing an auxiliary in V2 position and using the past participle of the main verb, and demoting the agent argument to oblique status so that it has to be introduced by a preposition.

Sentences in asymmetrical voice systems — which also include, e.g., ergative-absolute systems — are thus formally marked for their syntactic (in)transitivity, i.e. the number of direct core arguments. It is this formal marking that is at the center of interest in the current paper.

4.1.1.2 *Tagalog as an example of symmetrical voice systems*

Austronesian is an often discussed language family when it comes to the issue of voice systems, because it includes a number of languages that seem to defy categorization into the more familiar kinds of voice systems (Himmelmann, 2005a; Paul & Travis, 2006). Tagalog, a Western Austronesian language spoken in the Philippines, is one of these languages. The nature of its voice system has been debated for over a century (e.g. Blake, 1906; Kroeger, 1993a, 1993b; Schachter, 1976, 1995a) and it has been variously categorized as a nominative-accusative system (Bloomfield, 1917), a nominative-accusative-like system (Kroeger, 1993b), an ergative-absolutive system (Aldridge, 2012; Payne, 1982), or a "Philippine-type" symmetrical voice system (Foley, 2008).

In basic transitive sentences in Tagalog, one of the predicate's arguments is singled out and marked by the proclitic *ang* and its semantic role is reflected on the predicate[2] by means of voice affixes.[3] Apart from the morphologically overt dependency with the predicate, the *ang*-marked argument is privileged in a number of constructions, such as quantifier float where a floated *lahat* 'all' is always interpreted as quantifying this argument (cf. Kroeger, 1993b and Schachter, 1995a for further constructions that privilege *ang*-marked arguments).

Following Foley (2008), the *ang*-marked argument will be referred to as the *pivot argument* henceforth in this paper (cf. also Foley & Van Valin, 1984); depending on the specific analysis this argument is also referred to as the nominative or absolutive argument or trigger in the literature (cf. Kroeger, 1993b, Aldridge, 2012, and Schachter, 1995b, respectively). Other arguments which do not have their semantic role reflected on the predicate are marked by proclitic *ng* and are referred to as *core arguments* here (also called genitive arguments, cf. Kroeger, 1993b). Oblique arguments are marked by *sa* (often referred to as dative arguments in the literature).

[2] It is a matter of debate whether Tagalog exhibits a noun/verb distinction (Himmelmann, 2008). To circumvent this discussion, words that carry affixes that cross-reference the semantic role of one of the arguments in the clause will simply be referred to as predicates.

[3] There are constructions in which there are two *ang*-marked NPs in a sentence, such as possessor ascension and contrastive fronting (cf., e.g., Latrouite, 2011). However, those will not be dealt with in this paper.

The sentences in example (4.2) illustrate how the semantic role of the pivot is reflected on the predicate (pivot arguments and voice markers are in boldface).

(4.2) a. h<**um**>uli ng=isda sa=lawa **ang=lalaki**
<AV>catch CORE=fish OBL=lake PVT=man
"The man caught fish in the lake."

b. hu∼hulih-**in** ng=lalaki sa=lawa **ang=isda**
IRR∼catch-PV CORE=man OBL=lake PVT=fish
"The man will catch fish in the lake."

c. hu∼hulih-**an** ng=lalaki ng=isda **ang=lawa**
IRR∼catch-LV CORE=man CORE=fish PVT=lake
"The man will catch fish in the lake."

d. **ipang**-hu∼huli ng=lalaki ng=isda
IV-IRR∼catch CORE=man CORE=fish
ang=pamingwit
PVT=fishing.pole
"The man will catch fish with the fishing pole."

e. **i**-hu∼huli ng=lalaki ng=isda **ang=bata**
BV-IRR∼catch CORE=man CORE=fish PVT=child
"The man will catch fish for the child."

In (4.2a) the agent in the event is the pivot and the predicate takes the voice affix *<um>* signaling the semantic role of agent for the pivot. In (4.2b) on the other hand, the pivot is the patient of the event so the predicate takes a different voice affix (-*in*). Example sentences (4.2c–e) illustrate locative, instrumental and benefactive pivots, respectively. The sentences in (4.2) demonstrate that the morphological marking on the predicate is indeed a voice phenomenon in the sense that changes in the morphology "regularly

[... correspond] to a change in alignment between semantic role and syntactic function" (Himmelmann, 2002, 12). In the following, the current paper focuses on agent voice (4.2a) and patient voice (4.2b) because these are the two most frequent voice types in Tagalog and they involve the same semantic roles as active and passive sentences in German.

Both agent voice sentences and patient voice sentences are syntactically transitive in the sense that agents and patients are both direct core arguments of the predicate (Riesberg, 2014a). This is supported by the fact that *ng*-marked patients in agent voice sentences and *ng*-marked agents in patient voice sentences cannot undergo adjunct fronting which "implies that they are [... (core) arguments] since any [... non-argument] can appear in initial position in this construction" (Kroeger, 1993b, 47).[4]

In brief, there are no formally unmarked voice forms in Tagalog because both mappings of agents and patients to the pivot function are marked by voice affixes and de-transitivisation is not involved.

Taking into account that the Tagalog voice system seems to work differently from more familiar voice systems, Foley (2008) proposes to distinguish asymmetrical from symmetrical voice systems. The characteristic of symmetrical systems is that there is no default voice, i.e. no unmarked mapping between semantic roles and syntactic functions, and that all voice oppositions are equally formally marked (cf. also Himmelmann, 2004, 2005a; Riesberg, 2014a, 2014b). Foley (2008, 42) classifies the Tagalog voice system as symmetrical (and also the voice systems of other Austronesian languages have been

[4] Tanangkingsing and Huang (2007) proposed that the voice system of the closely related language Cebuano employs actives, passives and inverse sentences (Lawrence Reid, p.c.). In inverse sentences, the patient is more topical than the agent, which is reduced in prominence but not demoted. Here, it cannot be excluded that a similar analysis may apply to Tagalog agent voice and patient voice sentences. Crucially, however, the central finding that both voice types are equally transitive and equally morphologically marked would remain — preserving the main characteristics of the symmetrical voice system approach.

described as being symmetrical to varying degrees, cf. Cole, Hermon, & Yanti, 2008; Donohue, 2008; Riesberg, 2014b, inter alia). In these voice systems "[n]o one NP type is preferred for pivot choice [...]; regardless of which choice is made, all are signaled by some overt verbal voice morpheme [...]". Symmetrical voice systems are also characterized by having more than one basic transitive construction and that pivot and core arguments behave equally in the different voices. Maclachlan (1996) suggests treating both agent voice sentences and patient voice sentences as basic, because both are syntactically transitive to the same degree and also Riesberg (2014b) argues that no voice in Tagalog "is more basic than the other one(s)" (cf. also Kroeger, 1993b; Riesberg, 2014a). The characteristic of asymmetrical voice systems, on the other hand, is that they have "a marked preference [...] as to which NP should function as the pivot" (Foley, 2008, 42).[5]

In short, voice systems that have an unmarked mapping of syntactic functions to semantic roles and that involve de-transitivisation processes and additional overt marking when this mapping is to be altered are asymmetrical. Voice systems in which all mappings between syntactic functions and semantic roles are equally marked and no de-transitivisation takes place are symmetrical. Do these differences the architectures of asymmetrical and symmetrical voice systems lead to different processing signatures of the two systems?

4.1.2 *Potential consequences for sentence production*

Different predictions can be made regarding how sentence production may be influenced by the grammatical architectures of asymmetrical and symmetrical voice systems. In the following, I will give a brief overview of psycholinguistic models of sentence production and

[5]Latrouite (2011, 86) observes some degree of "patient-orientedness" with respect to certain constructions in Tagalog. This, however, does not influence its categorization as having a symmetrical voice system based on the distribution of case markers and voice affixes and the transitivity of agent voice and patient voice sentences.

describe how asymmetrical and symmetrical voice systems differ in sentence planning processes by hypothesis. Subsequently, I present two experiments that explore the processing differences between these two kinds of voice systems.

Sentence production is generally assumed to proceed in three stages (Bock, 1995; Bock & Levelt, 1994; V. S. Ferreira, 2010; Levelt, 1989). In the message encoding stage, speakers form a conceptual representation of what they want to convey. In the grammatical encoding stage, speakers construct linguistic structures and retrieve words that can be used to express the intended message. The final stage encompasses articulation of the planned material. Sentence production is also generally assumed to proceed incrementally, i.e. processing in the following stage can start on the still incomplete output of the current stage (Levelt, 1999, 88), interleaving the processing in the three stages.

The processes carried out during the grammatical encoding stage are the most relevant for the current purpose because asymmetrical and symmetrical voice systems should differ with respect to their requirements for these processes. Grammatical encoding is often described as consisting of several subprocesses that proceed partly in parallel (e.g., V. S. Ferreira, 2010; V. S. Ferreira & Slevc, 2007). One subprocess is structure building, where syntactic functions are assigned (functional processing) and syntactic structures are planned (constituent assembly). The other subprocess is content processing (lexical selection and retrieval), during which the words to be used are determined and morphological processes are carried out. Finally, during phonological encoding, phonological words are created to be passed on to the articulation stage.

By hypothesis, asymmetrical and symmetrical voice systems differ in their requirements for the functional processing part of grammatical encoding in which syntactic functions are assigned and constituent structure and word order are planned.

In the situation of planning a sentence in an asymmetrical voice system, the speaker would have to decide whether to produce a sentence with the verb in the unmarked voice or in marked voice. In the case of German, the decision to produce the marked voice (passive) means additional planning in comparison to what also has to be planned in actives: an additional auxiliary in V2 position and sentence-final placement of the past participle form of the main verb have to be planned. Thus, by hypothesis, depending on which voice is chosen in German, different planning operations have to be carried out.

In Tagalog with its symmetrical voice system, by contrast, all voices are equally marked morphologically and exhibit the same possible word orders. Thus, by hypothesis, Tagalog speakers always have to perform similar planning operations, regardless of whether they produce an agent voice or patient voice sentence. In both cases, they have to select a predicate and choose one argument to function as the pivot and encode an appropriate voice affix (Sauppe, Norcliffe, Konopka, Van Valin, & Levinson, 2013).

Two sentence production experiments tested for effects of potential differences in sentence planning in asymmetrical and symmetrical systems. While speakers produced sentences with different voice markings, the size of their pupils was measured. Pupil size changes are associated with attention allocation and mental effort. It is assumed that different operations during sentence production lead to differences in timing and amount of cognitive resource allocation, in turn leading to different patterns of pupil size changes. The experiments described below aim to investigate whether mental effort varies for speakers as they produce sentences with different voices.

Before turning to the description of the experiments, the use of pupil size measurements as an index of cognitive processing load will be briefly reviewed.

4.1.3 *Task-evoked pupillary responses*

The dilation and constriction of human pupils is controlled by the locus coeruleus (LC), a subcortical structure near the brain stem that emits the neurotransmitter norepinephrine (Laeng, Sirois, & Gredebäck, 2012). Among other functions, LC activity has been linked to attention allocation (Sara, 2009). Since the LC is a region that also controls the muscles of the iris (Samuels & Szabadi, 2008; Sirois & Brisson, 2014), there is a tight link between pupil diameter and activity in the LC.

Changes in pupil diameter in relation to experimental tasks have been used as an indirect measure of LC activity to study cognitive effort or processing load in psychology during the last 60 or so years (Hess & Polt, 1964; Laeng et al., 2012). When a task-relevant stimulus in an experiment is processed, the pupils dilate in response to the occurrence of the stimulus. The time course of the pupillary response (most commonly measured in the form of peak amplitude and peak latency) is related to the cognitive effort that is necessary to process the stimulus. Changes in pupil diameter in response to the presentation and processing of experimental stimuli are called task-evoked pupillary responses (TEPRs).

TEPRs are an index of activity in the LC-norepinephric system in the so-called phasic mode of activity where neurons fire rapidly to optimize performance during a specific task and thereby to focus attention (Laeng et al., 2012; Sirois & Brisson, 2014).

The experimental relevance of the relation between TEPRs and attention and the allocation of cognitive resources has been demonstrated by many studies (Ahern & Beatty, 1979; Gabay, Pertzov, & Henik, 2011; Kahneman & Beatty, 1966; Laeng, Ørbo, Holmlund, & Miozzo, 2011; Murphy, Robertson, Balsters, & O'Connell, 2011; Richer & Beatty, 1985; Zylberberg, Oliva, & Sigman, 2012, i.a.). Beatty (1982, 291) concludes that the "task-evoked pupillary response [...]

provides a reliable and sensitive indication of [...] variations in processing load", thereby making it a very useful method to investigate even potentially small effects of differential cognitive effort exerted by "qualitatively different mental operations" in experimental tasks (Beatty, 1982, 290).

The pupillary response can also be used to investigate language processing. Just and Carpenter (1993) were able to show that in English, reading sentences with greater syntactic complexity demands more cognitive resources than reading sentences with less complex syntax (e.g., object relative clauses vs. subject relative clauses), participants' pupillary response was larger when reading more complex sentences (cf. also Piquado, Isaacowitz, & Wingfield, 2010). TEPRs have also been shown to be sensitive to other aspects of language processing, such as mismatches between prosody and syntax (Engelhardt, Ferreira, & Patsenko, 2010), intelligibility of the signal (Zekveld & Kramer, 2014; Zekveld, Kramer, & Festen, 2010), speech rate (Koch & Janse, 2016), simultaneous interpretation (Hyönä, Tommola, & Alaja, 1995), frequency effects in lexical decision (Kuchinke, Võ, Hofmann, & Jacobs, 2007), prosody in discourse processing (Zellin, Pannekamp, Toepel, & van der Meer, 2011), and word retrieval in second language processing (Schmidtke, 2014).

Most studies that use pupillometric measures have investigated comprehension processes. Papesh and Goldinger (2012) present one of the few production studies measuring TEPRs (Schluroff et al., 1986 is another example). Papesh and Goldinger showed that pupillary responses during speech planning are sensitive to word frequency and that frequency effects emerge after lexical access. They conclude that examining pupil size changes during language tasks can "potentially [... reveal] differences in cognitive demands, even in cases with equivalent overt performance" (Papesh & Goldinger, 2012, 760).

4.1.4 *Current experiments*

The experiments reported in this paper employed task-evoked pupillary responses to investigate whether asymmetrical and symmetrical voice systems differ in the cognitive resources that speakers have to allocate during the sentence production process. Cognitively induced changes in the size of speakers' pupils were measured to assess differences in production between voices and between voice systems. Specifically, it was tested whether processing load develops in ways that are predicted from the voice systems' supposedly different demands on the functional processing stage of sentence production.

As we have seen, the voice system of Tagalog is symmetrical. It is hypothesized from the voice system's architecture that the planning processes that a speaker has to carry out are similar for all voices. It is predicted that the production of agent voice and patient voice sentences will elicit similar task-evoked pupillary responses. This was tested in Experiment 1 on Tagalog.

Conversely, it is hypothesized from the architecture of the asymmetrical voice system of German that the planning processes differ between voices. It is predicted that producing actives and passives taxes speakers in contrasting ways, derived from the assumption that processing load is distributed unevenly among active and passive because additional material has to be planned and produced in the latter. Therefore, distinct TEPRs are expected to be found.

The analyses focus on the temporal unfolding of cognitive load while speakers plan and produce sentences. Especially, experimental trials were not reduced to a single number or two by only analyzing peak dilation amplitudes or peak dilation. Instead, the entire time course of the pupillary response was analyzed. Growth curve analysis was used to model changes in the shape of pupillary responses. This statistical technique is able to capture non-linear changes in curve shapes and is thus well-suited to investigate task-evoked pupillary responses and to obtain a detailed picture of pupil size changes over the time course of sentence planning. The reasoning behind employing growth curve analysis is that it increases the sensitivity of pupillometric analyses (Kuchinsky et al., 2013, 31) by assessing pupil size changes over time. It is assumed that if the production of different voice types affords different planning operations this will correlate with distinct patterns of attention allocation and LC activity which are reflected in differential TEPR curve shapes.

4.2 EXPERIMENTS

4.2.1 *Experiment 1: Tagalog*

The Tagalog data reported here were collected during an eye tracking experiment, reported in (Sauppe et al., 2013). The purpose of this experiment was to investigate the time course of argument planning during Tagalog sentence production using the picture description paradigm (Griffin & Bock, 2000). In this paradigm, participants see line drawings of simple transitive events and are asked to describe them in one sentence while their gaze and speech are recorded. Additionally, participants' pupil size is measured by the eye tracker. An advantage of this paradigm is that there are usually very few restrictions on what form the speakers' responses should take. Thus, the elicited utterances are relatively spontaneous and natural.

4.2.1.1 *Participants*

Fifty-three native speakers of Tagalog (mean age = 17 years, 13 male) were recruited from De La Salle University, Manila, to participate in the experiment for payment.[6] All of them reported speaking Tagalog at least five hours per day and to at least one of their parents.

4.2.1.2 *Materials*

Target pictures were 44 colored line drawings of transitive events (cf. Figure 4.1), interspersed among 76 filler pictures of intransitive events, with at least one filler separating any two target pictures. Two versions of each target picture were created by mirror-reversing the picture.[7] Pictures were then arranged in four lists created by pseudo-randomizing the order of the target and filler pictures so that every two targets were separated by at least one filler. The two mirror-reversed versions of each target picture were counterbalanced across lists. Each target and filler picture was preceded by a black fixation dot randomly appearing in one out of five positions on the upper part of the computer screen against a white background.

Target pictures were normed by twenty different Tagalog speakers from De La Salle university, who did not participate in the experiment and who provided written descriptions of the pictures in a questionnaire. The pictures were then scored with respect to their tendency towards being described using agent voice or patient voice or exhibiting no tendency. Pictures with tendencies towards agent voice descriptions and towards patient voice descriptions had an equal share among target items (19 pictures with agent voice tendency, 19 pictures with patient voice tendency, 5 pictures with no tendency). The purpose of the norming was to ensure that the set of target pictures included both pictures that were likely to elicit agent

[6] Data from two additional participants had to be excluded due to technical problems with the recording equipment.

[7] Mirror-reversed versions of the target pictures were used in order to ensure that the left-to-right order of agent and patient in the pictures would not influence the participants' responses, which was especially important in the light of the hypotheses of the eye-tracking experiment for which the data were collected (Sauppe et al., 2013).

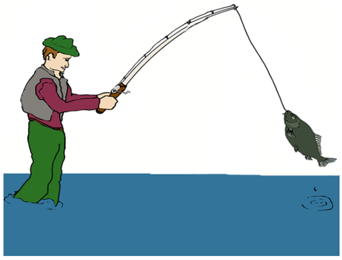

Figure 4.1: Example stimulus picture

voice sentences and pictures that were likely to elicit patient voice sentences.

4.2.1.3 *Procedure*

Before the testing, participants read the instructions for the experiment and completed a questionnaire on their linguistic background, both in Tagalog. The instructions were repeated orally again in Tagalog to make sure that participants fully understood them. Participants were asked to describe the events shown in the pictures with one predicate-initial sentence that named all the depicted characters taking part in that event as accurately and as quickly as possible. There was a practice phase at the beginning of the experiment in which participants saw eleven example pictures and simultaneously heard example sentences illustrating how they could be described (these were seven intransitive sentences and four transitive sentences, two in which the agent was the pivot argument and two in which the patient in the depicted event was the pivot argument). Participants then described the example pictures themselves and the experimenter provided feedback if they started speaking very late after stimulus onset, or if they did not name all characters or used non-predicate-initial structures. Then the experiment started and participants described the target and filler pictures presented to them. Experimental sessions lasted approximately 40 minutes.

4.2.1.4 *Apparatus*

Pupil size was recorded with a Tobii T120 remote eye tracker controlled by the Tobii Studio software. Both eyes were sampled at 120 Hz. Stimuli were presented with a resolution of 1280×1024 pixels. Participants' vocal responses were recorded with a microphone and saved and time-stamped together with the eye tracking data by Tobii Studio. All participants were tested in the same windowless room; illumination conditions were identical.

4.2.1.5 *Data pre-processing*

Pupil diameters measured with low validity (validity value ≥ 1 as coded by the Tobii Studio software) were coded as missing values (cf. Schmidtke, 2014) as were physiologically unlikely pupil diameters (smaller than 2 mm or greater than 7 mm) and pupil diameters farther away than 2.5 standard deviations from the trial mean (cf. Alnæs et al., 2014). If the absolute change in pupil diameter from one time step to the next exceeded 0.1 mm it was also coded as missing value (cf. Schmidtke, 2014) in order to remove probable measurement artifacts. Missing values were then linearly interpolated (Zeileis & Grothendieck, 2005) separately for each eye. Remaining missing values for one eye were replaced by the value from the other eye when available. Pupil diameters from the left and right eyes highly correlated (ρ = 0.94); to reduce noise, pupil diameter measurements of both eyes were averaged (cf. Schmidtke, 2014).

To reduce computational cost given the large amount of data resulting from the eye tracker output, data were down-sampled (Signal Developers, 2013) from 120 Hz to 30 Hz, resulting in one sample every $\sim$33 ms. For each trial the mean diameter during the last 1000 ms of the presentation of the fixation dot was taken as the baseline pupil diameter and subtracted from all pupil diameter values of a trial to account for differences in pupil diameter between trials. Finally, data were smoothed by local polynomial regression fitting (degree = 2, span = 0.1, cf. Alnæs et al., 2014).

4.2.1.6 *Data selection*

Participants' picture descriptions were transcribed by a native Tagalog speaker and annotated by the author with respect to which voice marking (and word order) was used and which words were chosen to describe the event.

Trials with more than 50% missing data points for the left or the right eye before linear interpolation were excluded from analyses, as were trials where participants started to speak later than 6500 ms after stimulus onset or where the speech onset was more than three standard deviations longer than the mean speech onset latency. Trials were also excluded if the description did not contain overt agent and patient arguments, if it was not predicate-initial, or if speakers corrected themselves. This left 481 agent voice sentences and 780 patient voice sentences for analysis. The distribution of voice types in the responses reflects the general frequency distribution of voice types in Tagalog (McFarland, 1984).[8]

[8] Additionally, participants produced 31 locative voice sentences (4.2c) and nine benefactive voice sentences (4.2e). However, although all non-agent voices (4.2b–e) are often subsumed under the cover term *undergoer voice* because they share several semantic and formal characteristics (Himmelmann, 2005b, 363), analyses were restricted to patient voice sentences because potential TEPR differences between patient voice and the other non-agent voices could not be assessed with these small numbers non-agent voice sentences. Such potential differences could arise due to the different semantic roles of the pivot arguments or because locative voice, benefactive voice, and instrumental voice are much less frequent than patient voice.

4.2.1.7 *Analyses*

Statistical analyses were performed in R (R Core Team, 2015) using the package *lme4* (Bates et al., 2015). *P*-values for effects in regression models were calculated with the *car* package (Fox & Weisberg, 2011).

Linear mixed effects regression analyses were performed to model the temporal dynamics of pupil diameter changes by growth curve analysis (cf. Kalénine et al., 2012; Kuchinsky et al., 2013; Mirman et al., 2008). Growth curve analysis is a type of multilevel regression to model variations in curve shapes over time by using orthogonal polynomial time terms as explanatory variables (Mirman, 2014).

For the current analyses, linear, quadratic, cubic and quartic time terms were employed to model the shape of the pupil diameter changes while Tagalog speakers produced sentences of different voice types.

Fourth-order orthogonal polynomials were chosen after visual inspection of the grand mean of the pupillary response in order to accommodate the number of inflection points of the curve. Each of the polynomial time terms reflects a separate component of the TEPR curve. The linear time term reflects the overall slope of the pupillary response (greater estimates meaning a larger pupil size at the end of the analysis time window). The quadratic term describes the primary inflection point of the curve (smaller estimates meaning an overall flatter distribution), the cubic term describes secondary inflection points (positive estimates meaning that peaks occur earlier whereas negative estimates mean that peaks occur later). The quartic term describes tertiary inflection points in the tails of the pupillary response curve (Kuchinsky et al., 2013, 27; Mirman, 2014, 49f.). For computational reasons, only interactions between voice type and the linear, quadratic and cubic time terms were included in the model. The interaction between voice type and the quartic time term was waived to reduce model complexity and also because it is of less theoretical interest as it mainly describes differences in the tails of the curves. Nevertheless, this time term was included as predictor to account for quartic components in pupillary response curves across voices.

Changes in pupil diameter were analyzed in a time window between 0 and 4250 ms. The to-be-described picture stimulus appeared on the screen at 0 ms and the grand mean length of the produced utterances was 4242 ms.

Linear mixed effects regression models were fit with a random effects structure that came close to maximal random effects structure justified by design (Barr, 2013; Barr et al., 2013). A random intercept and correlated random slopes for the orthogonal polynomial time terms of interest as well as for voice and their interactions were included for subjects. For items, a random intercept as well as correlated random slopes for the time terms were included.

The two levels of the categorical independent variables were coded as -0.5 and 0.5, respectively (Cohen et al., 2003).

The effects under investigation may have been influenced by other factors than voice. Speech onset latency, the mean greyscale value of each item, and the codability of the predicate were added to the model as control variables. They were allowed to interact with the time terms of interest in order to account for the influence that they might have had on the TEPR curves. However, no random slopes for these control variables were included in the regression models (Barr et al., 2013, 275). The inclusion of the control variables ensures that variance in pupil size changes that can be explained by these variables alone is not (falsely) attributed to an influence of voice on TEPR curve shapes (cf. Wurm & Fisicaro, 2014).

The first control variable was the individual trials' speech onset latency as a behavioral measure of formulation difficulty. Speakers are expected to take longer to initiate articulation of an utterance, e.g., if they find it difficult to conceptualize the depicted event or need more time to retrieve words with lower frequencies. Longer speech onset latencies are expected to go hand in hand with linear increases in pupil size reflecting the elevated processing load. Additionally, this variable is also expected to have an influence on the cubic components of TEPRs because latencies might reflect whether speakers perform certain planning steps and allocate cognitive resources earlier or later.

The second control variable was luminance of the display, which changed when the fixation dot disappeared from the screen and the to-be-described picture stimulus appeared. This elicited a constriction of the pupil (pupillary light reflex, Bergamin & Kardon, 2003) that can be sensitive to cognitive resource allocation (Verney, Granholm, & Dionisio, 2001, cf. also Mathôt, van der Linden, Grainger, & Vitu, 2013). Including the mean greyscale values of the presented pictures in the model accounts for these differences in luminance that could have had an influence on participants' pupil size.

Finally, the codability of the predicate was included as control variable. Codability reflects how easy or hard the depicted event was to recognize and encode for speakers (cf. van de Velde et al., 2014). It was determined by calculating the Shannon entropy H (Shannon, 1948) of the predicate separately for each item and then categorizing this item as highly codable if the respective H was smaller or equal than the median H of all items and as lowly codable if the H was larger. Put simply, if all speakers used the same words to describe an event, codability was high; if individual speakers tended to use different words to describe an event, codability was low.

4.2.1.8 *Results*

The upper panel of Figure 4.2 shows the time course of pupil size changes during the production of agent voice and patient voice sentences in Tagalog. The picture stimulus appeared on the screen at time = 0 ms. Before picture presentation a fixation dot was displayed; the mean pupil diameter of the last 1000 ms of the fixation dot presentation were taken as a baseline period. The TEPR curve shows the shape of an initial constriction (as a response to the increased luminance of the display when the picture stimulus is presented) followed by pupil dilation under cognitive load (cf. Verney et al., 2001, 78), which is the typical shape of pupillary response curves.

Agent voice and patient voice sentences showed similar temporal pupillary response dynamics. Pupil size decreased and increased at the same time in both voice types.

Table 4.1 shows the results of the growth curve analysis for Tagalog. There is a significant interaction between voice type and the linear time term, indicating that pupil size was larger at the end of the analysis window when speakers produced patient voice sentences than when they produced sentences with agent voice marking. This effect may be due to a slower decrease of the pupil diameter after having reached its peak for patient voice trials. The quadratic and cubic time terms did not significantly interact with voice type. This indicates that overall the pupillary response curves for agent voice and patient voice sentences had similar shapes. Specifically, pupil size increased with the same steepness and around the same time during the production of both voices.

The control variables had significant effects on pupil sizes. With longer speech onset latencies, speakers' pupil size dilated more (speech onset latency $\times$ time^1), faster (speech onset latency $\times$ time^2) and earlier (speech onset latency $\times$ time^3). In addition, the relative luminance of pictures (mean greyscale value) significantly influenced the shape of pupillary response curves. Speakers' pupils also dilated more in trials in which the predicate codability was low, i.e. where it was harder to identify the depicted event and where speakers thus agreed less as to which words to use to describe the event.

The linear mixed effects regression model in Table 4.1 ignores word order differences between sentences and only includes voice type as critical predictor variable. The basic word order in Tagalog is predicate-initial, as the sentences in (4.2) illustrate. However, the order of arguments following the predicate is relatively free. Example (4.3) demonstrates that different constituent orders are equally grammatical. Agent voice and patient voice sentences can thus also exhibit internal variation as to whether the pivot argument is realized sentence-finally or sentence-medially. The canonical word order, however, is one where the *ang*-marked pivot argument is sentence-final (Kroeger, 1993b; Schachter & Otanes, 1972), as in (4.3a,c).

(4.3) a. h<**um**>ili ng=isda **ang=lalaki**
<AV>catch CORE=fish PVT=man
b. h<**um**>ili **ang=lalaki** ng=isda
"The man caught fish."
c. hu~hulih-**in** ng=lalaki **ang=isda**
IRR~catch-PV CORE=man PVT=fish
d. hu~hulih-**in ang=isda** ng=lalaki
"The man will catch the fish."

To rule out the possibility that the differences in TEPRs for agent voice and patient voice were just due to word order differences, a model was constructed in which word order was the critical predictor variable. This model compared TEPR curve shapes for sentences in which the *ang*-marked argument was final (as in (4.3a,c)) to sentences in which it was non-final (as in (4.3b,d)). The order of core argument and pivot argument after the predicate did not significantly influence pupil size (all p-values > 0.14).

Agent voice and patient voice sentences also differed with respect to speech onset latencies. Speakers were able to start speaking earlier when planning sentences with agent pivots (mean speech onset = 1579 ms, measured from the moment the to-be-described picture appeared on the screen, SD = 454 ms) than when the patient was the pivot argument (mean = 1684 ms, SD = 474 ms). Linear mixed effects regression models predicting log-transformed latencies confirm this difference (β = 0.05, $|t|$ = 2.24, $\chi^2(1)$ = 5.01, $p < 0.03$). Pivot-final and pivot-medial sentences (as in ex. (4.3a–b)), however, did not differ in their speech onset latencies (β = 0.03, $|t|$ = 1.02, $\chi^2(1)$ = 1.05, $p \approx 0.31$). Inclusion of predicate codability as control variable did not influence the pattern of results for speech onset latencies.

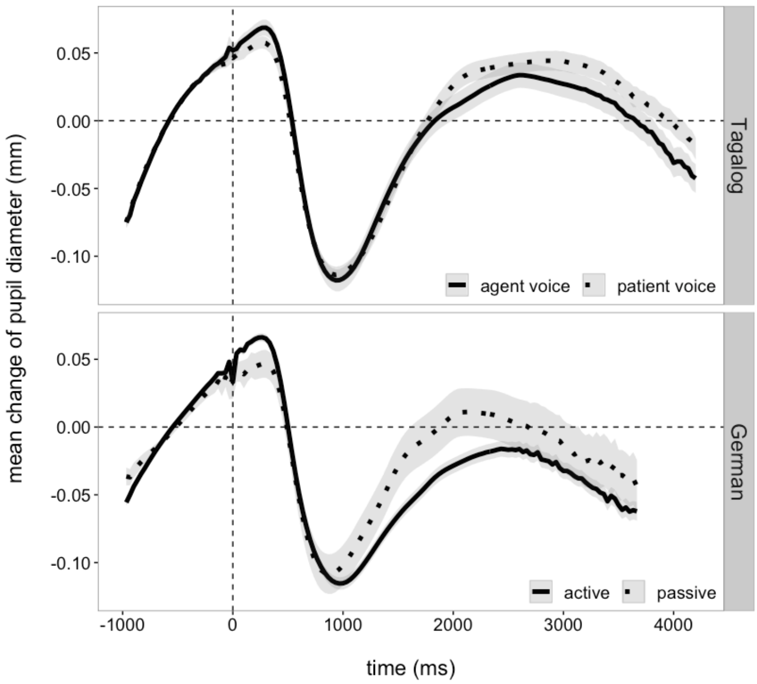

Figure 4.2: Time course of pupil diameter changes relative to baseline during sentence production in Tagalog (Experiment 1, upper panel) and German (Experiment 2, lower panel); ribbons indicate one standard error of the mean; dashed vertical line indicates the moment when the to-be-described picture appeared on the screen; dashed horizontal line indicates the baseline pupil diameter.

Table 4.1: Linear mixed effects regression results (Tagalog)

	β	$\|t\|$	χ^2	
Intercept	5.57×10^{-3}	0.37		
time^1	1.68×10^{-1}	2.40	8.57	**
time^2	3.72×10^{-2}	0.73	0.58	
time^3	-4.69×10^{-1}	14.23	207.42	***
time^4	1.91×10^{-1}	47.29	2236.17	***
voice (= PV)	1.25×10^{-2}	1.52	0.27	
voice × time^1	1.38×10^{-1}	3.98	15.85	***
voice × time^2	-3.87×10^{-2}	0.98	0.96	
voice × time^3	1.30×10^{-2}	0.54	0.29	
speech onset latency	1.88×10^{-2}	36.84	1329.96	***
speech onset latency × time^1	1.77×10^{-1}	30.84	951.18	***
speech onset latency × time^2	7.14×10^{-2}	12.44	154.69	***
speech onset latency × time^3	-3.42×10^{-2}	6.05	36.55	***
mean greyscale value	-2.98×10^{-2}	3.54	0.75	
mean greyscale value × time^1	-8.53×10^{-2}	2.25	5.06	*
mean greyscale value × time^2	8.54×10^{-2}	2.74	7.53	**
mean greyscale value × time^3	-5.93×10^{-2}	3.82	14.59	***
predicate codability (low)	2.42×10^{-2}	1.46	0.22	
predicate codability × time^1	1.39×10^{-1}	1.85	3.43	.
predicate codability × time^2	-9.98×10^{-3}	0.16	0.02	
predicate codability × time^3	-2.75×10^{-3}	0.09	< 0.01	

Note: Dependent variable: baselined pupil diameter (mm); . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, p-values from Type II Wald χ^2 tests with df = 1 (Fox & Weisberg, 2011); condition number $\kappa = 1.82$ (Cohen et al., 2003)

4.2.1.9 *Discussion*

The results of the growth curve analysis of task-evoked pupillary responses elicited by the production of agent voice and patient voice sentences in Tagalog suggest that speakers carried out similar processes when planning both voice types.

Notably, the *shape* of TEPRs did not differ between voices. The quadratic and cubic polynomial time terms, which describe the inflection points and thus the shape of the pupillary response curves, did not interact with voice type (both p-values > 0.33). In other words, Tagalog speakers' pupils started to increase at the same time and with the same steepness in both agent voice and patient voice. Cognitive processing effort changed in similar ways over the course of sentence production. This is taken to indicate that the same planning operations were carried out in both voices.

However, there was a significant interaction between voice type and the linear time term. This means that speakers' pupils were larger towards the end of the analysis time window for patient voice sentences. The TEPR curves in Figure 4.2 suggest that speakers' pupils constricted more slowly after having reached their peak dilation during the production of these sentences. This indicates that processing load was maintained for a longer time for the planning of patient voice as compared to agent voice sentences. Thus, the two voice types in Tagalog shared the same processing load time course (no interactions of voice and quadratic or cubic time terms) but there were more cognitive resources allocated to patient voice sentences towards the end of the analysis time window. It would not have been possible to detect this effect with an analysis focusing only on peak dilations because the reduction to a single measurement point per trial could not have captured the prolonged maintenance of processing load over the time course of patient voice production.

The analysis of speech onset latencies provides additional evidence for the interpretation that the planning of patient voice sentences demands more cognitive effort than the planning of agent voice sentences. Speakers needed significantly more time to initiate articulation of patient voice sentences — despite the result from the growth curve analysis that the time course of processing load changes is similar for both voices.

On the whole, however, the results from the current experiment are in line with Foley's (2008) analysis of the Tagalog voice system as being symmetrical (cf. also Riesberg, 2014b). Symmetry in the Tagalog voice system means that all voices are equally marked morphologically. This also entails that the same operations have to be carried out for agent voice and patient voice during sentence formulation when semantic roles are encoded and syntactic functions are assigned to arguments. Speakers selected one discourse entity to become the pivot argument and planned equal amounts of marking (voice affixes and case markers) in both voices (cf. Sauppe et al., 2013). Yet, by the same token, processing load was maintained for longer time when pivot arguments were patients.

4.2.2 *Experiment 2: German*

An analogous sentence production experiment was carried out on German. The data were collected during an eye tracking experiment, reported in Sauppe (2017).

German exhibits an asymmetrical voice system in which active is the unmarked voice and passives are marked. Given that speakers have to carry out planning of additional morphological material during the production of passive sentences, differential TEPR curve shapes are predicted for the production of actives and passives. More precisely, it is predicted that the voice type will interact with the quadratic or cubic time terms that describe the shape of the task-evoked pupillary response curves. The different or additional planning operations may be carried out at different times for actives and passives, leading to earlier or steeper increases of cognitive processing load for passives.

The comparison of possible TEPR differences between actives and passives in German with the pattern of results from Tagalog will furthermore make it possible to investigate whether changes in cognitive load over time are more similar for different voices in symmetrical systems than for different voices in asymmetrical systems. In other words, are there are potential "processing profiles" for the planning of sentences in asymmetrical or symmetrical voice systems?

4.2.2.1 *Participants*

Thirty-three native speakers of German (mean age = 25 years, 10 male) were recruited from Radboud University Nijmegen, HAN University of Applied Sciences in Nijmegen and among the PhD students of the Max Planck Institute of Psycholinguistics to participate in this experiment. Students received payment for their participation. All participants were unaware of the hypotheses of the experiment.

4.2.2.2 *Materials, procedure, and apparatus*

The picture stimuli of this experiment consisted of the pictures used in the Tagalog experiment, as well as 16 additional transitive and 17 additional filler pictures in order to elicit more picture descriptions from each participant (cf. Sauppe, 2017).

The experimental procedure was identical to the procedure in the Tagalog experiment. Participants first read the instructions and completed a questionnaire on their linguistic background. After an oral repetition of the instructions, they entered the practice phase consisting of 15 example pictures depicting transitive and intransitive events.

Pupil size was recorded with a Tobii T120 remote eye tracker running Tobii Studio, sampling both eyes with 60 Hz. Participants' vocal responses were recorded with a microphone and time-stamped and saved by Tobii Studio. All participants were tested in the same experimental booth.

4.2.2.3 *Data pre-processing and data selection*

Data pre-processing was performed as for the Tagalog experiment. Participants' picture descriptions were transcribed by student assistants and checked and annotated for voice marking and words used in the descriptions by the author.

Trials were excluded from analyses if they had more than 50% missing data points for the left or right eye before linear interpolation, if speech onset was later than 6500 ms after stimulus onset or more than three standard deviations longer than the mean speech onset latency. Trials were also excluded if the description was neither a transitive active sentence or a passive sentence overtly including the agent (e.g., existentials) or if speakers corrected themselves. This left 1172 active sentences and 105 passive sentences for analysis.

4.2.2.4 *Analyses and results*

A linear mixed effects regression model was fit to the task-evoked pupillary response curves from sentence production in German using growth curve analysis. The fixed effects structure and the random effects structure were identical to the structure of the models in Experiment 1. Changes in pupil diameter were analyzed in a time window between 0 and 3700 ms. The to-be-described picture stimulus appeared on the screen at 0 ms and the grand mean length of the produced utterances was 3693 ms.

The time course of pupil size changes during the production of active and passive sentences in German is shown in the lower panel of Figure 4.2. The overall shape of the pupillary response curves resembles the curve shapes obtained in Experiment 1 on Tagalog.

It is noticeable, however, that the time course of the pupillary response differs between active and passive sentences. After the initial constriction, speakers' pupil diameter started to increase earlier during the production of passives than during the production of actives.

Table 4.2 shows the regression results for this experiment. The interaction of voice type with the cubic time term was significant. This indicates that pupil diameter increased earlier during the planning of passive sentences.

As in Tagalog, speech onset latency and mean greyscale value of the picture significantly influenced the time course of pupil size changes. This was a highly expected result since the effect of the latter is largely attributable to the pupillary light reflex and the effect of the former is a general indicator of production difficulty, i.e. pupillary responses were influenced by how much planning time speakers needed before they could initiate articulation, indicating that differences in planning difficulty went in hand with variations in mental effort. The codability of the predicate also significantly influenced TEPR curve shapes. Pupil size changes differed between trials in which speakers concurred in naming the event and trials in which they were divided over how to name the event.

To make the analysis of the German data more comparable to the data from Experiment 1 on Tagalog, a linear mixed effects regression model was built that only included those pictures that were also presented in the Tagalog experiment. Very similar effects were found in this model (voice $\times$ time^1: $\beta = 8.50 \times 10^{-2}$, $|t| = 1.35$, $\chi^2(1) = 1.82$, $p = 0.18$; voice $\times$ time^2: $\beta = -2.68 \times 10^{-3}$, $|t| = 0.06$, $\chi^2(1) < 0.01$, $p = 0.96$; voice $\times$ time^3: $\beta = 5.37 \times 10^{-2}$, $|t| = 1.97$, $\chi^2(1) = 3.90$, $p < 0.05$).

German speakers did not need significantly more time to initiate the production of passive sentences (mean = 1778 ms, *SD* = 498 ms) as compared to active sentences (mean = 1699 ms, *SD* = 475 ms).

Table 4.2: Linear mixed effects regression results (German)

	β	$\|t\|$	χ^2	
Intercept	-3.73×10^{-2}	2.13		
time^1	-7.89×10^{-2}	0.83	1.88	
time^2	2.10×10^{-1}	3.24	12.87	***
time^3	-3.69×10^{-1}	9.97	145.77	***
time^4	8.84×10^{-2}	24.15	583.25	***
voice (passive)	8.70×10^{-4}	0.06	0.31	
voice × time^1	4.53×10^{-2}	0.76	0.57	
voice × time^2	-1.08×10^{-2}	0.25	0.06	
voice × time^3	7.12×10^{-2}	2.68	7.20	**
speech onset latency	-5.97×10^{-4}	1.29	4.71	*
speech onset latency × time^1	1.09×10^{-1}	22.19	492.55	***
speech onset latency × time^2	1.03×10^{-1}	21.06	443.68	***
speech onset latency × time^3	2.17×10^{-3}	0.45	0.20	
mean greyscale value	-3.23×10^{-2}	4.78	1.47	
mean greyscale value × time^1	-1.15×10^{-1}	2.99	8.95	**
mean greyscale value × time^2	6.29×10^{-2}	2.71	7.34	**
mean greyscale value × time^3	-3.42×10^{-2}	2.23	4.99	*
predicate codability (low)	2.90×10^{-2}	2.15	1.13	
predicate codability × time^1	1.71×10^{-1}	2.25	5.05	*
predicate codability × time^2	-3.59×10^{-2}	0.78	0.61	
predicate codability × time^3	-4.47×10^{-2}	1.47	2.17	

Note: Dependent variable: baselined pupil diameter (mm); * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, p-values from Type II Wald χ^2 tests with df = 1 (Fox & Weisberg, 2011); condition number $\kappa = 3.06$ (Cohen et al., 2003)

4.2.2.5 *Discussion*

Participants' pupil size changed in different ways for actives and passives over the course of sentence production in German: Processing load increased earlier during the production of the latter. This is indicated by the significant interaction of voice and the cubic time term.

There was no significant interaction of voice type and the linear time term, in contrast to what was found in Tagalog. This suggests that potentially higher mental effort during the production of passives had leveled out by the end of the analysis time window. There were also no differences in the steepness of pupil dilations (voice $\times$ time^2).

The differential pupillary response curve shapes for active and passive sentences suggest that there were different processing dynamics involved. As outlined earlier, the planning of passives supposedly required cognitive operations that might be qualitatively different from the operations employed during active sentence production.

One possible interpretation of the earlier increase of pupil diameter in the production of passives is that speakers had to manipulate the argument structure of verbs. Lemmas of verbs may only specify the unmarked mapping of agent and patient to the subject and object syntactic functions, which corresponds to the mapping in actives. When speakers produced passives, they would have had to compute the marked mapping in which the patient argument was the syntactic subject and the agent was demoted to oblique status (Levelt, 1989).

An alternative interpretation would be that speakers did not have to perform mapping computations online but that the different curve shapes for actives and passives resulted from the fact that more material had to be planned in passives in various positions in the sentence. It is not possible with the data at hand to decide between these explanations. However, it appears that they do not differ notably with respect to their consequences; both entail that the differential pupillary response curve shapes are due to the execution of qualitatively different planning processes for actives and passives.

Differences in TEPR curve shapes between German actives and passives could also have been caused by increased cognitive effort when speakers assessed the felicity of using a passive given the depicted event. However, Tagalog speakers also had to assess the properties of the event in order to decide which voice to use (there is, e.g., a grammatical constraint favoring human patients to be pivots, cf. Latrouite, 2011). Thus, distinct curve shapes would also have been expected for Tagalog if they reflected the evaluation of felicity conditions in order to select an appropriate linguistic form.

The German pupillary response pattern stands in contrast to the TEPR pattern from Tagalog. In Experiment 1, growth curve analysis suggested that there were no differences in the curve shapes between the two Tagalog voices, providing evidence that speakers performed the same planning processes for both voice types.

It is to be noted that especially the analysis of the time course of pupil size changes and the shape of pupillary response curves revealed different patterns in Tagalog and German. If speech onset latencies alone had been analyzed, the interpretation of the results would have been that different voice types in Tagalog employ different planning processes while no differences could be detected in German — whereas the observed pupil size changes tell a different story. The similar TEPR curve shapes in Tagalog and the contrasting curve shapes in German suggest that active and passive are distinct from each other in a different way than agent voice and patient voice are distinct from each other. Specifically, voices in Tagalog differ by the cognitive effort that is required to link the agent or patient semantic roles to the highest syntactic function, but the similar TEPR curve shapes suggest that the same planning operations are performed to do this. The two German voice types, by contrast, appear to employ different operations as indexed by diverging pupillary response curve shapes. Analyzing pupillometric data in addition to speech onset latencies made it possible to "distinguish mental effort from behavioral performance" (Karatekin, Couperus, & Marcus, 2004, 184), showing that patterns of cognitive resource allocation can give different insights than patterns of speech onset latencies. These two kinds of data are thus best considered synergistically.

4.3 CONCLUSIONS

To sum up, the experiments on Tagalog and German sentence production revealed different patterns of pupil size changes for voice forms in the two voice systems.

In Tagalog, cognitive load dynamics did not differ between agent voice and patient voice sentences, both of them exhibiting similar pupillary response curve shapes. This suggests that speakers carried out the same operations during planning of either voice type, namely choosing one event participant to function as syntactic pivot and encoding an appropriate voice affix and the relevant case markers (cf. Sauppe et al., 2013). However, processing load was maintained for a longer time during the production of patient voice sentences.

In German, the pupillary response curves for active and passive sentences differed in their shape but not their overall slope. This suggests that speakers had to carry out qualitatively different planning operations, which were potentially distributed over the whole time course of production because additional material had to be planned in various sentence positions for passives.

The distinct patterns of differences in task-evoked pupillary response curves for different voice types in Tagalog and German — similar TEPR curve shapes in Tagalog and differential curve shapes in German — indicate that there are psycholinguistic processing consequences of the (a)symmetry of voice systems. Looking at how voices differ within languages makes it possible to identify what might be the "processing profiles" of voice systems because most factors, such as whether the first constituent is a verbal predicate or a noun phrase, are held constant.

The results of the current experiments support the idea of categorizing languages as exhibiting asymmetrical or symmetrical voice systems as proposed by Foley (2008). It is possible to detect differences between these types of voice systems in how their grammatical properties influence the allocation of cognitive resources during sentence planning.

It can be deduced that the differences in TEPR curve shapes that were observed across and within languages originated during the grammatical encoding phase. It is, however, not possible to localize a specific point in time at which the exact source of the effects is to be found. Neurophysiological response latencies for the locus coeruleus are approximately 150–250 ms (Laeng et al., 2011) and reliable effects can occur at least 200–300 ms after a cognitive event (Beatty, 1982). However, Wierda, van Rijn, Taatgen, and Martens (2012) suggest that pupil diameter changes result from attentional pulses. This suggests that there is no single neural event that is the source of the TEPR curves that were observed for individual trials in the current experiments. Thus, pupil size changes over the course of sentence production are interpreted here as a "summed index of brain activity during cognitive events" (Goldinger & Papesh, 2012, 91). This also acknowledges the fact that sentence production is very complex, involving many subprocesses (cf., e.g., V. S. Ferreira, 2010) and that it is to some degree also a temporal blackbox for which it is hard to say which process has been carried out at which point in time exactly.

Another finding of Experiment 1 is that patient voice sentences are more effortful to plan than agent voice sentences, although the general time course of resource allocation was the same. Speakers' pupil size decreased more slowly after having reached its peak dilation when sentences that carried patient voice marking were produced. Although they are as morphologically marked as agent voice sentences, patient voice sentences seem to be slightly disfavored by the processing system in the sense of demanding more cognitive resources to be allocated to their planning and production. However, the difference in cognitive load between the production of the two voice types is likely to be rather small because speakers appear to have no difficulties producing patient voice and it is in fact the more frequent voice type in Tagalog (Latrouite, 2011; McFarland, 1984).

Interestingly, Riesberg and Primus (2015) show that although there are no grammatical preferences for linking agents to the pivot function in symmetrical voice languages, there is still some degree of agent prominence (e.g., in parts of the paradigms of voice affixes). Additionally, Schachter (1976, 1995a) argues that syntactic privileges for different constructions are divided between the pivot and the agent argument; agents are, for example, binders of reflexives irrespective whether they are the pivot or not (Schachter, 1977).

That processing load is maintained for longer during the production of patient voice can be interpreted as an instance of an anti-P(atient) bias in sentence production: It is more effortful for speakers to plan sentences in which a patient (P) argument is mapped to the highest syntactic function compared to sentences where this function is fulfilled by an agent. This interpretation is also supported by the speech onset latency analyses that showed that speakers started speaking later for patient voice.

By the same token, however, this effect is compatible with the notion of symmetry in the Tagalog voice system. Cognitive load decreased more slowly in patient voice after having increased at the same time and with the same steepness as in agent voice sentences. This indicates that speakers performed the same or similar planning operations with equal timing for both voices but that these operations were more effortful to complete for patient pivots.

A similar anti-P bias might be operating in German, too. It is, however, not possible to disentangle mental effort derived from mapping the patient argument to the subject function and having to plan additional marking because these two factors are intrinsically connected in the German voice system. Passives also come with more pragmatic restrictions on their use than actives.

Moreover, there is also evidence from the sentence comprehension literature for anti-P effects. Listeners follow strategies that allow them to identify the agent as quickly and unambiguously as possible (Alday, Schlesewsky, & Bornkessel-Schlesewsky, 2014; Bornkessel-Schlesewsky & Schlesewsky, 2009, 2013b). There is ample evidence for a "subject-first" preference in many (European) languages in which ambiguous noun phrases tend to be interpreted as syntactic subjects (amounting to an agent interpretation in transitive sentences). This has been shown for Dutch (Frazier, 1987), German (Schriefers, Friederici, & Kühn, 1995), English (Traxler, Morris, & Seely, 2002), among other languages. A similar effect has also been demonstrated in Chinese (Wang, Schlesewsky, Bickel, & Bornkessel-Schlesewsky, 2009), Hindi (Bickel, Witzlack-Makarevich, Choudhary, Schlesewsky, & Bornkessel-Schlesewsky, 2015; Choudhary, Schlesewsky, Roehm, & Bornkessel-Schlesewsky, 2009), and Turkish (Demiral, Schlesewsky, & Bornkessel-Schlesewsky, 2008).

This bias may help listeners to keep the structures that they build more minimal when they first try to interpret a role-ambiguous noun phrase as agent (Wang et al., 2009). Agents can be causally and existentially independent (Primus, 1999), i.e. they can launch actions without patients (as in "*Mary was working all day*"); this does not hold true for patients which are affected by (causal) actions that must be instigated by an agent.

Together with the evidence from Experiment 1 on Tagalog which suggests that cognitive load was maintained for a longer time during the production of patient voice sentences, it may be concluded that there is a general bias against structures in which a patient is mapped to the highest syntactic function, thus causing more effort to produce and comprehend these structures.

The existence of an anti-P bias in production and comprehension supports approaches that include interfaces between the two modes of language processing (e.g., Dell & Chang, 2014; Kempen, Olsthoorn, & Sprenger, 2012; MacDonald, 2013; Pickering & Garrod, 2013). This effect might be due to the special status of agents in cognition which makes them easier to be mapped to the highest syntactic function — and in turn disfavors mappings of non-agents to this function — because linguistic agents overlap with instigators of goal-directed actions in the world in many of their features (Bornkessel-Schlesewsky & Schlesewsky, 2013a). Agents are general cognitive attractors which may be related to seeing the self as an acting agent capable of voluntary control (Haggard, 2008); (awareness of) agency also plays an important role in the conceptualization of the self and the distinction between self and other (David et al., 2006; Decety & Sommerville, 2003; Frith & Frith, 2010). Additionally, parts of Broca's area, a brain region which is also involved in language processing, are involved in the representation of actions and goal-directed human movements (Clerget, Winderickx, Fadiga, & Olivier, 2009; Fazio et al., 2009). In his review of the literature on the connection between representations of syntactically transitive sentences and motor aspects of goal-directed actions Kemmerer (2012, 60) concludes that these neural mechanisms are "biased toward clauses with canonical mappings between syntax and semantics", i.e. where agent arguments are mapped to the highest syntactic function, in their capturing of the hierarchical (and sequential) organization of actions.

Evans and Levinson (2009, 446) suggest that the diversity of human languages and the distribution of typological features in the world's languages involves functional and cognitive attractors. Agents could operate as cognitive attractors causing a general anti-P bias in language processing that is detectable in production and comprehension and may also influence the distribution of voice systems among the world's languages. Languages tend to disfavor voice systems in which a mapping of patient-like arguments to the highest syntactic function is unmarked: symmetrical voice systems are only found in some Austronesian languages (cf. Riesberg, 2014b). Since language change may be influenced by both cognitive biases (Bickel et al., 2015) and lineage-specific tendencies (Dunn, Greenhill, Levinson, & Gray, 2011), amongst other things, it is an open question how different factors may have jointly contributed to the genesis and retention of symmetrical voice systems in the evolution of Austronesian (Himmelmann, 2005a; Ross, 2002).

To conclude, the pupillometric data presented in this paper supports the notion of symmetrical and asymmetrical voice systems by showing that the distinction has processing consequences during sentence production. Furthermore, the Tagalog data support the notion of an anti-P bias operating in sentence production, for which there is converging evidence from the sentence comprehension literature.

The current paper contributes to the literature on under-studied languages (Jaeger & Norcliffe, 2009; Norcliffe, Harris, & Jaeger, 2015) by investigating Tagalog in comparison to German, extending our understanding of the interplay of a language's grammatical properties and general psycholinguistic mechanisms during sentence production.

Part III

VOICE AND WORD ORDER IN SENTENCE COMPREHENSION

5

VERBAL SEMANTICS DRIVES EARLY ANTICIPATORY EYE MOVEMENTS DURING THE COMPREHENSION OF VERB-INITIAL SENTENCES

Sauppe, S. (2016). Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology*, 7(95). doi: 10.3389/fpsyg.2016.00095

ABSTRACT

Studies on anticipatory processes during sentence comprehension often focus on the prediction of postverbal direct objects. In subject-initial languages (the target of most studies so far), however, the position in the sentence, the syntactic function, and the semantic role of arguments are often conflated. For example, in the sentence "The frog will eat the fly" the syntactic object ("fly") is at the same time also the last word and the patient argument of the verb. It is therefore not apparent which kind of information listeners orient to for predictive processing during sentence comprehension. A visual world eye tracking study on the verb-initial language Tagalog (Austronesian) tested what kind of information listeners use to anticipate upcoming postverbal linguistic input. The grammatical structure of Tagalog allows to test whether listeners' anticipatory gaze behavior is guided by predictions of the linear order of words, by syntactic functions (e.g., subject/object), or by semantic roles (agent/patient). Participants heard sentences of the type "Eat frog fly" or "Eat fly frog" (both meaning "The frog will eat the fly") while looking at displays containing an agent referent ("frog"), a patient referent ("fly") and a distractor. The verb carried morphological marking that allowed the order and syntactic function of agent and patient to be inferred. After having heard the verb, listeners fixated on the agent irrespective of its syntactic function or position in the sentence. While hearing the first-mentioned argument, listeners fixated on the corresponding referent in the display accordingly and then initiated saccades to the last-mentioned referent before it was encountered. The results indicate that listeners used verbal semantics to identify referents and their semantic roles early; information about word order or syntactic functions did not influence anticipatory gaze behavior directly after the verb was heard. In this verb-initial language, event semantics takes early precedence during the comprehension of sentences, while arguments are anticipated temporally more local to when they are encountered. The current experiment thus helps to better understand anticipation during language processing by employing linguistic structures not available in previously studied subject-initial languages.

5.1 INTRODUCTION

Anticipation, the prediction of upcoming events, is an important property of human cognition and it has been argued recently that brains are essentially "prediction machines" (Clark, 2013, cf. also Bubic, von Cramon, & Schubotz, 2010). Predictive processes are found, for example, in interaction between individuals when people predict the outcome of actions performed by others (Sebanz & Knoblich, 2009) and even their movements (Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004).

Anticipation is also involved in language processing. During the comprehension of spoken or written sentences, language users build predictions about the upcoming linguistic input. Words are, for example, read faster when they are predictable from the context as compared to unpredictable words (Ehrlich & Rayner, 1981). Language users may even predict the phonological form of an upcoming word: DeLong et al. (2005) found differential EEG responses when listeners encountered a determiner (*a/an*) that did not fit with the noun that they assumed will follow (*"The day was breezy so the boy went outside to fly... a kite* vs. *an airplane"*). Anticipatory processes are also found in conversation where listeners predict the end of their interlocutor's turn, in order to be able to take their own turn in a timely manner (Magyari, Bastiaansen, de Ruiter, & Levinson, 2014; Magyari & de Ruiter, 2012).

The visual world paradigm has been used extensively to investigate predictive processes during language comprehension. In this experimental paradigm, participants see a display and hear an accompanying sentence while their eye movements are recorded (cf. Huettig, Rommers, & Meyer, 2011 for a review). In a seminal visual world study, Altmann and Kamide (1999) showed that in English the lexical semantics of verbs is used to anticipate the syntactic object of a sentence by incrementally narrowing down the set of potential referents. Participants saw displays showing, e.g., a boy, a ball, a toy train, a toy car, and a cake, and heard sentences of the form *"The boy will move/eat the..."*. The verb of the sentence could either take any of the depicted things (*move*) or only one of them (*eat*) as its syntactic object. Listeners used the verb's selectional restrictions and fixated on the corresponding element in the display already before it was mentioned when the verb only allowed one object referent in this position (*eat* and *cake* in this case).

Further visual world studies substantiate the idea that sentence comprehension is highly predictive and that listeners use various kinds of information to form their expectations. Kamide, Scheepers, and Altmann (2003) showed that case marking information can be combined with semantic information from the verb in German to anticipate syntactic objects. Kamide, Altmann, and Haywood (2003) showed that information from several constituents can be combined to predict upcoming elements in English ditransitive sentences and in verb-final Japanese sentences. Boland (2005) showed that arguments are more likely to be anticipated than adjuncts in English. Knoeferle et al. (2005) showed that listeners rapidly integrate visual information that is provided to them and that this information is used to anticipate object referents in German, even when the sentences accompanying a display describe unusual situations and therefore run counter to listeners' world knowledge.

All of these studies have in common that they investigated how information provided by sentential and visual context are integrated to predict elements that occur at the end of sentences. The already encountered input restricts language users' attention to the anticipation of the only remaining element of the sentence. Transitive verbs, such as *eat*, take two arguments and in languages with subject-initial word order (e.g., English and German), listeners already have heard one of the arguments when they encounter the verb — the point from which anticipatory eye movements are measured in most studies. Thus, listeners already have information about one argument, including its referential identity and its semantic role (in the case of Kamide, Altmann, & Haywood, 2003 even about two arguments of ditransitives). Put differently, in previous studies on subject-initial languages the anticipation target has always been a single element at the end of a sentence, conflating syntactic function, word order, and semantic role.

There is thus still an open question regarding what kind of information listeners orient to for predictive processes during sentence comprehension. Do they try to anticipate referents based on syntactic function (e.g., direct object)? Alternatively, are their expectations based solely on what they expect to follow next? Or do listeners rather exploit semantic information to form expectations about the event and therefore anticipate referents carrying certain semantic roles (e.g., patient or goal)? Unfortunately, studies of subject-initial languages are not suited to answer these questions because the three different types of anticipation targets are conflated on the last noun phrase position that is usually employed to test prediction processes. Taking Altmann and Kamide's (1999) sentences, *cake* is the direct object, the patient and the word directly following the verb. Examining the prediction of this element cannot differentiate between these three types of information as the anticipation target.

Verb-initial languages offer a possibility to disentangle these various theoretical possibilities. In these languages, the verb is the first word of a sentence and information about the described event and selectional restrictions are provided upfront, potentially enabling listeners to identify referents and the semantic roles that they play. Importantly, the early position of the verb may enable listeners to anticipate upcoming arguments before any of them is mentioned. This means that all three anticipation target types are still available — prediction based on semantic roles, on syntactic functions, or on word order. In subject-initial languages, on the other hand, one argument is always mentioned before the verb.

In the following, a visual world eye tracking experiment on Tagalog will be reported. Tagalog is an Austronesian language primarily spoken in the Philippines. The experiment was devised to test what kind of information listeners anticipate in verb-initial languages upon having heard the verb.

5.1.1 *Current Experiment*

In the experiment described below, participants looked at visual displays depicting three potential referents (cf. Figure 5.1) while hearing verb-initial Tagalog sentences. Two elements in the display corresponded to the agent and to the patient of the sentences, the third element was an unrelated distractor. Participants' eye movements were recorded in order to analyze their looks to the elements as the sentences unfolded. The experiment was designed to investigate what kind of information listeners orient toward upon hearing a

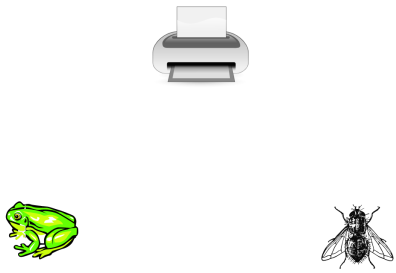

Figure 5.1: Example stimulus

sentence-initial verb and what it is that they anticipate, especially when there are more possible anticipation targets than just the last word of the sentence. There are different sentence types in Tagalog that can be used to test the three possible anticipation targets; these sentence types are described in the following.

Basic word order in Tagalog is verb-initial and the verb carries voice affixes that cross-reference the semantic role of one of its arguments. This argument is marked by *ang* and will be referred to as the pivot argument. Non-pivot arguments that do not have their semantic role cross-referenced are marked by *ng*. Canonically and most frequently, the non-pivot argument immediately follows the verb and the pivot argument is realized sentence-finally (cf., e.g, Himmelmann, 2005b, 357).

In (5.1a)[1] the agent in the event (*frog*) is marked by *ang* and the verb exhibits voice morphology that signals that the semantic role of this pivot argument is agent (AV). In (5.1b) the patient (*fly*) is marked by

[1]The following abbreviations are used in the current paper: A = agent, AV = agent voice, NPVT = non-pivot argument, P = patient, PV = patient voice, PVT = pivot argument, RP = recent perfective aspect.

ang and the verb signals that the pivot argument's semantic role is patient by means of different voice morphology (PV).[2]

(5.1) a. Kakain sa umaga ng=langaw ang=palaka
eat:AV in the morning NPVT=fly (**P**) PVT=frog (**A**)
"The frog will eat a fly in the morning."[3]

b. Kakainin sa umaga ng=palaka
eat:PV in the morning NPVT=frog (**A**)
ang=langaw
PVT=fly (**P**)
"A/the frog will eat the fly in the morning."

Importantly, both sentences are equally transitive. Kroeger (1993b, 40–48) shows with a number of syntactic tests that *ng*-marked patients in agent pivot sentences (5.1a) and *ng*-marked agents in patient pivot sentences (5.1b) are arguments of the verb. Tagalog can thus be described as exhibiting a so-called symmetrical voice system (Foley, 2008; Riesberg, 2014b). This is in contrast to English where passive sentences are intransitive and the agent may only be realized as oblique.

Therefore, in sentences like (5.1), the initial verb provides language users with semantic information about the described event. In the context of a visual world eye tracking experiment, this might allow them to identify which referents in the visual display could sensibly be involved in the described event (e.g., a frog as the agent and a fly as the patient in sentences like in 5.1 or a boy and a cake as in Altmann & Kamide, 1999). Additionally, the voice marking carried by the verb provides information about the canonical order of agent and patient in the unfolding sentence. When the verb signals that the agent is

---

[2] Tagalog also exhibits a variety of other voice forms where, e.g., the instrument, the beneficent or the location of an event is the pivot and has its semantic role cross-referenced at the verb (e.g. Himmelmann, 2005b; Schachter & Otanes, 1972).

[3] Differences in the definiteness of agent and patient in the translations arise due to constraints on interpreting the *ang*-marked argument as specific (Adams & Manaster-Ramer, 1988, cf. also Latrouite, 2015).

the pivot (example 5.1a), listeners know that it will be canonically and most frequently realized sentence-finally, i.e., that the canonical order is [patient agent]. When the verb marks a patient pivot (example 5.1b), listeners know that the canonical order is [agent patient]. Thus, the sentence-initial verb provides listeners with information about the event from which agent and patient referents in the display can be inferred and it provides them with information about the canonical and most frequent order in which these referents will be mentioned.

Tagalog also exhibits a construction that differs from the sentences in (5.1) in an interesting way. Sentences in the recent perfective aspect describe events that recently happened. In these sentences the verb is not marked for voice but carries an invariant aspect marker. Thus, there is no pivot in recent perfective sentences (5.2) and the canonical order of arguments is [agent patient].

(5.2)	Kakakain	pa lang	ng=palaka	sa=langaw
	eat:RP	just	NPVT=frog (**A**)	NPVT=fly (**P**)
	"A/the frog just ate the fly."			

Taken together, sentences with agent pivots, patient pivots and recent perfective sentences provide a way of investigating what kind of information language users anticipate after having heard a sentence-initial verb. The three sentence types contrast in their verb marking, i.e., whether the semantic role of a pivot argument is reflected on the verb (5.1) or not (5.2) — and if there is a pivot argument, whether it is the agent or the patient of the sentence. Additionally, the three sentence types also differ in the canonical order of the agent and patient arguments ([patient agent] for agent pivot sentences, 5.1a, and [agent patient] for patient pivot and recent perfective sentences, 5.1b and 5.2). Whether Tagalog listeners anticipate upcoming linguistic input on the basis of semantic or syntactic information can be investigated by comparing the comprehension of these three sentence types. It is possible to formulate differential hypotheses for each possible kind of information that may be used in anticipatory processing based on listeners' behavior during sentence comprehension. These hypotheses will be laid out in more detail in the following.

If Tagalog listeners primarily orient toward syntactic information in anticipation, they could use the semantic and morphosyntactic information provided by the verb to identify agent and patient referents and assign syntactic functions (pivot, non-pivot) to them.

A strong form of syntactically based anticipation would be the prediction of pivot arguments, i.e., that listeners anticipate the sentence-final pivot NP by already fixating on the corresponding referent in the display while or shortly after hearing the sentence-initial verb. When the verb signals that the agent is the pivot (5.1a), listeners should look toward the agent more after having heard the verb than when the patient is signaled to be the pivot (5.1b) — in which case listeners should direct their gaze toward the patient. Sauppe et al. (2013) found that in Tagalog sentence production the pivot argument plays an important role early in the planning of sentences: Tagalog speakers select a pivot at the outset of formulation in order to be able to retrieve an appropriate voice affix. If the role of the pivot argument is mirrored in anticipatory processing during sentence comprehension, fixation preferences for the agent in (5.1a) or the patient in (5.1b) are expected shortly after listeners encountered the verb.

Another syntactically based process would be the anticipation of the first-mentioned argument upon hearing the verb. Under this scenario, listeners use verbal information to identify referents and their canonical order to determine whether agent or patient will be mentioned first and will subsequently direct their gaze toward them. After having heard a verb that signals an agent pivot, listeners should direct their gaze toward the patient element in the display because the canonical word order for these sentences is [patient agent]. After having heard a verb with patient pivot or recent perfective marking,

Table 5.1: Overview of sentence types; pivot arguments underlined

Region		Verb	Adverb	NP1	NP2
agent pivot	(5.1a)	eat$_{AV}$	in the morning	fly (**P**)	frog (**A**)
patient pivot	(5.1b)	eat$_{PV}$	in the morning	frog (**A**)	fly (**P**)
recent perfective	(5.2)	eat$_{RP}$	just	frog (**A**)	fly (**P**)

listeners should direct their gaze toward the agent referent (cf. Table 5.1).

Finally, if Tagalog listeners directed their attention toward semantic roles and therefore toward the structure of the event, they should fixate on the agent in all three sentence types after having heard the verb. Agents play a prominent role in communication in general because they are initiators of events. Cohn and Paczynski (2013) propose that agents are centrally involved in building representations of events and may take early precedence during the cognition of events since they are the "heads of causal chains that affect patients" (Kemmerer, 2012). Agents are also attended to more than patients by infants (Robertson & Suci, 1980) and play a highlighted role in many grammatical hierarchies (Aissen, 1999; Lockwood & Macaulay, 2012). Given these points, it seems justified to assume that agents are the target of anticipatory processes in Tagalog if prediction was guided by semantic roles.

In the grammatical literature it has also been proposed that Tagalog exhibits a "patient primacy," partly because sentences in which the patient is the pivot are more frequent than agent pivot sentences (cf. Latrouite, 2011 for a discussion). Theoretically, the patient could thus also be fixated preferentially after the verb was heard. However, on the hypothesis that the anticipation of semantic roles would mainly serve to construct an event representation, it seems a priori more likely that agents would be targeted for this purpose.

5.2 EXPERIMENT

5.2.1 *Participants*

Forty-nine students of the University of the Philippines, Diliman, participated in the experiment for payment (mean age = 18.8 years, 22 male). All of them reported being native speakers of Tagalog and speaking the language with at least one of their parents. All participants had normal or corrected-to-normal vision.

The reported experiment conforms to the American Psychological Association's ethical principle of psychologists and code of conduct (as declared by the ombudsman of the Max Planck Institute for Psycholinguistics). Written informed consent was obtained from participants at the beginning of the experiment session.

5.2.2 *Materials and Methods*

5.2.2.1 *Materials*

In the experiment, participants looked at stimulus displays while hearing pre-recorded sentences. Stimulus displays consisted of three colored line drawings that were arranged in a triangular shape (Figure 5.1). Line drawings either represented the agent or patient of the event described in the accompanying sentence or were distractors which were not mentioned. The position of agent, patient and distractor was counterbalanced across displays.

Displays were paired with sentences that were either intransitive or transitive. All intransitive sentences were fillers. Transitive sentences described a range of animacy scenarios in which agent and patient were humans, animals, or inanimate entities. However, scenarios in which both agents and patients were inanimate were not included.

Verbs and arguments were semantically associated to varying degrees (ranging from *police car chases thief* to *owl carries bag*).

In all sentences the initial verb was followed by an adverb (*sa umaga* 'in the morning', *sa tanghali* 'at noon', or *sa hapon* 'in the afternoon' for sentences as in 5.1 and *pa lang* 'just' for recent perfective sentences as in 5.2). The adverb was included to increase the time between hearing the verb and the first noun phrase[4] in order to give participants time to parse the verb and direct their gaze toward the anticipation target (cf., e.g., Kamide, Scheepers, & Altmann, 2003; Mishra, Singh, Pandey, & Huettig, 2012 for similar stimulus sentence structures).

Sentences were recorded by a female native speaker of Tagalog and had a neutral intonation contour so that none of the words was particularly highlighted.

Fifty-one critical displays were paired with transitive sentences which exhibited either marking of agent voice, patient voice, or recent perfective on the sentence-initial verb; agent and patient were depicted together with a distractor element semantically unrelated to the two arguments and the verb. In these displays only one element could be the agent referent and only one could be the patient referent. Seventy-nine filler displays depicted only one argument of the accompanying sentence and two distractors. The sentences were either intransitive and thus included only one argument (49 sentence-display pairs) or transitive (30 sentence-display pairs). In the latter case, one argument was mentioned but not depicted as an element in the display or two elements were possible agents or patients of the verb. Three pseudo-randomized lists were created so that each critical display occurred with one of the three sentence types in each list and at least one filler intervened between any two

[4]Strictly speaking, the arguments are expressed by determiner phrases headed by the markers *ng*, *ang* and *sa*, which define the referential meaning of the phrases. Content words are not sub-classified for syntactic categories in Tagalog and therefore there are no noun and verb classes (Himmelmann, 2008). For the sake of simplicity, however, the term *NP* will be used in this paper, following Himmelmann (2005b).

critical displays. For sentences describing scenarios where humans were acted on, either undergoer voice or recent perfective was used in two lists as there is a grammatical constraint against agent voice when the patient is human (Latrouite, 2011).

5.2.2.2 *Procedure*

Participants were seated in front of a 17" laptop computer with a screen resolution of 1024 × 768 pixels. Eye movements were recorded with 120 Hz sampling rate by a SMI RED-m eye tracker attached below the computer's screen. Auditory stimuli were presented via headphones.

Trials began with the presentation of a fixation cross in the middle of the screen that triggered the presentation of the experimental display after participants looked at the cross for 700 ms. The auditory presentation of sentences started 1000 ms after the onset of the display, which stayed visible until after the end of the sentence.

After a quarter of the trials participants were asked to indicate whether all the referents mentioned in the sentence were also depicted; this was always true for the critical transitive sentences and sometimes true and sometimes false for filler sentences. Five practice trials were included at the beginning of the experiment.

The judgment task that participants had to carry out was similar to the task employed in Altmann and Kamide (1999) where participants had to indicate whether the event could apply to the picture, which was the case when all relevant referents were depicted. This kind of "look and listen" task was also employed in other visual world eye tracking studies investigating anticipatory processes (e.g. Huettig, Singh, & Mishra, 2011). Huettig, Rommers, and Meyer (2011, 154) conclude that "the listeners' eye movements during a trial of a visual world experiment reflect the direction of their visual attention, which depends both on the visual and auditory input," i.e., listeners look at the elements in the display as they are mentioned and become relevant (Huettig, Rommers, & Meyer, 2011, 153). The linking hypothesis employed in the current paper is thus that listeners' gaze is a reliable reflection of their attention allocation during sentence comprehension.

Before testing, participants read instructions for the experiment in Tagalog and completed a questionnaire on their linguistic background. The whole session lasted approximately 35 min.

5.2.2.3 *Analyses*

To test the hypotheses regarding possible anticipation targets outlined above, the time course of participants' fixations to agent and patient referents in experimental displays during the comprehension of the three different sentence types was analyzed.

Likelihoods of agent and patient fixations were analyzed with quasi-logistic linear mixed effects regression models (Barr, 2008; Bates et al., 2015; Pinheiro & Bates, 2000; R Core Team, 2015) in three time windows. The first time window encompassed the sentence-initial verb and the immediately following adverb (*Verb + Adverb* region, duration: mean = 1183 ms, *SD* = 96 ms), the second time window spanned the period during which the first argument was presented (*NP1* region, duration: mean = 703 ms, *SD* = 187 ms), finally the third time window covered the presentation of the second argument (*NP2* region, duration: mean = 815 ms, *SD* = 201 ms). To account for variations in the duration of regions across stimuli due to differing word lengths, the duration of each time window was normalized. For every stimulus, the onset of the respective region for each analysis time window corresponded to *time = 0* and the region's offset corresponded to *time = 1*. In this way, only fixations that occurred during the presentation of any given sentence region of each item were included in the corresponding analysis time windows. Fixations were aggregated into empirical logits over five consecutive bins for each analysis time window.

Time and sentence type were included as predictors in all regression models and the maximal random effects structure justified by design (that allowed the models to converge) was used (Barr, 2013; Barr et al., 2013). Significance of fixed effects was assessed using Type II Wald F-tests with Kenward-Roger approximation of denominator degrees of freedom (Fox & Weisberg, 2011; Halekoh & Højsgaard, 2014; Kenward & Roger, 1997). Sentence type as categorical predictor was coded with Helmert contrasts.

Trials were excluded from analyses if track-loss occurred, defined as the eye tracker having lost the participant's eyes for more than 650 ms (236 trials, 9.4%), or due to technical problems with the recording equipment (15 trials, 0.6%). Trials were also excluded if the question after a given trial was answered incorrectly; six participants that answered less than 80% of questions correctly were excluded entirely from the analyses (296 trials, 11.8%). One item was excluded from analyses because it was accidentally in the same condition in all lists. In one list, the trials from one critical display were excluded because it accidentally was presented together with a filler sentence. Three combinations of display and recent perfective sentence were discarded because they were rated as only marginally acceptable in a post-hoc internet-based acceptability rating study conducted with 50 Tagalog speakers from the Philippines (51 trials, 2%). Nine stimuli were excluded because the accuracy of agent recognition (given the display and the voiceless and aspect-less gerund form of the verb) was less than 10% above chance in a *post-hoc* internet-based rating study with 29 Tagalog speakers from the Philippines (322 trials, 13%). In total, 1568 trials were included in the analyses.

5.2.3 *Results*

The time course of listeners' fixations to agents and patients during the auditory presentation of the three different sentence types is

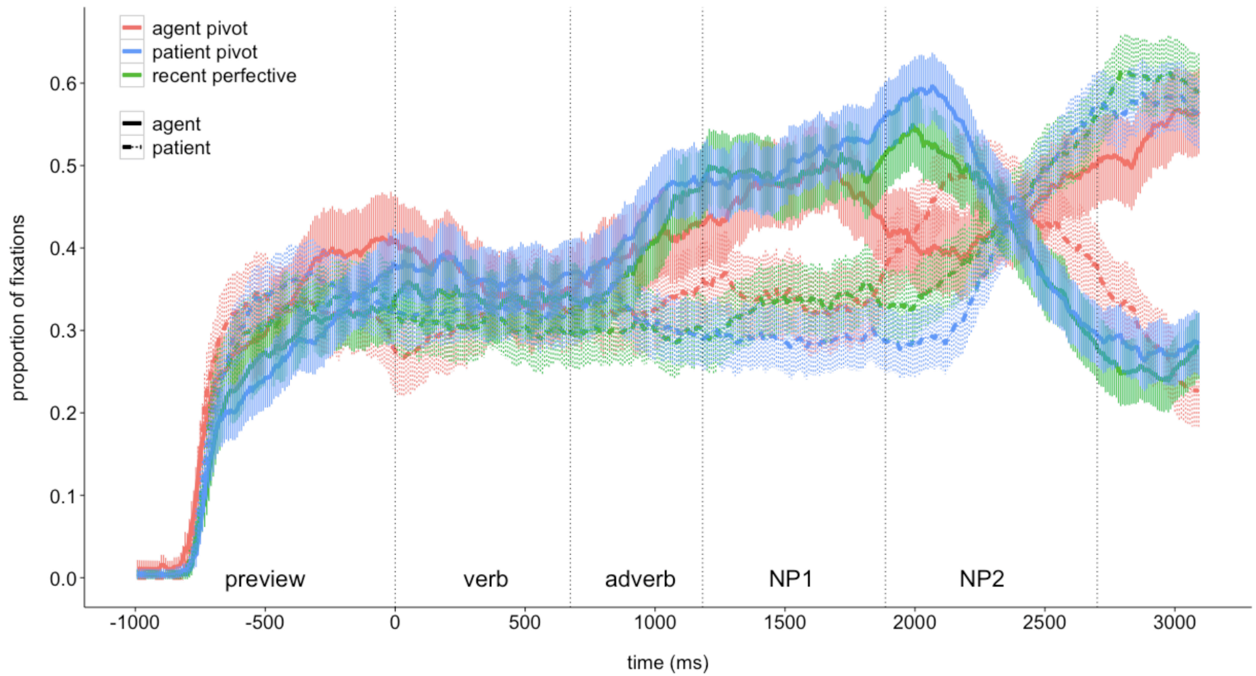

Figure 5.2: Fixations to depictions of agents and patients during auditory presentation of three different sentence types; ribbons indicate 95% confidence intervals; vertical dotted lines indicate mean onset and offset of regions in the critical sentences (cf. Table 5.1)

shown in Figure 5.2. Visual inspection of the graph suggests that agent fixations increased during the *Verb + Adverb* region in all three sentence types after listeners encountered the verb. Agent fixations then continued to increase in sentences with patient voice (5.1b) and recent perfective marking (5.2) until the agent was mentioned. For sentences with agent voice marking (5.1a), participants' agent fixations decreased during the *NP1* region where the patient was mentioned and increased again later when the agent was mentioned during the *NP2* region. In contrast, fixations to the patient did not increase during the *Verb + Adverb* region in any of the sentence types. In sentences where the patient was encountered after the adverb (5.1a), participants' fixations to that referent started increasing toward the end of the *NP1* region and decreased during the *NP2* region in which the agent was mentioned. In sentences with patient pivots or recent perfective marking the patient was mentioned only sentence-finally. In these sentences, participants' fixations to the patient started to increase only toward the end of the *NP1* region and during the *NP2* region where it was mentioned.

Table 5.2 shows the results of the quasi-logistic linear mixed effects regression models for fixations to the agent in the three analysis time windows. During the *Verb + Adverb* region, only time is a significant predictor. This means that during this time window, the likelihood of agent fixations increased over time and it did so to a similar degree in all sentence types; in other words, the slope does not vary with verb marking.

During the *NP1* region, there was a steeper increase in agent fixations by-subjects in sentences where it was mentioned first, i.e., sentences with a sentence-final patient pivot (5.1b) or recent perfective marking (5.2). The fixation patterns associated with these two sentence types were highly similar but differed from fixation patterns observed when listeners heard sentences with an agent pivot (i.e., where the agent was heard first, 5.1a). This difference arose because

Table 5.2: Quasi-logistic linear mixed effects regression results predicting empirical logits of fixations to the agent referent in three different sentence types

	by subject β	by subject *SE*	by subject 95% Wald CI	by subject *F* statistic		by item β	by item *SE*	by item 95% Wald CI	by item *F* statistic	
Verb + Adverb										
Intercept	-0.50	0.05	-0.59, -0.40			-0.48	0.08	-0.63, -0.32		
Time	0.36	0.11	0.14, 0.57	$F(1,10842)$ = 14.91	***	0.38	0.16	0.07, 0.70	$F(1,4241)$ = 6.28	*
{PV, RP} vs. AV	<0.01	0.09	-0.17, 0.17	$F(2,5754)$		0.03	0.07	-0.12, 0.17	$F(2,18430)$	
PV vs. RP	0.13	0.09	-0.05, 0.32	= 1.17		0.12	0.07	>-0.01, 0.25	= 2.23	
Time × {PV, RP} vs. AV	0.29	0.29	-0.28, 0.85	$F(2,10546)$		0.17	0.18	-0.19, 0.53	$F(2,43723)$	
Time × PV vs. RP	0.15	0.19	-0.21, 0.52	= 0.67		0.09	0.25	-0.40, 0.57	= 0.51	
NP1										
Intercept	-0.05	0.05	-0.15, -0.05			-0.06	0.07	-0.19, 0.08		
Time	0.15	0.10	-0.04, 0.34	$F(1,3879)$ = 3.41		0.23	0.09	0.04, 0.41	$F(1,3471)$ = 4.86	*
{PV, RP} vs. AV	0.20	0.10	0.01, 0.39	$F(2,1257)$	**	0.21	0.10	<0.01, 0.41	$F(2,686)$	*
PV vs. RP	0.10	0.08	-0.06, 0.26	= 4.72		0.11	0.12	-0.12, 0.34	= 3.18	
Time × {PV, RP} vs. AV	0.57	0.26	0.07, 1.08	$F(2,2443)$	*	0.43	0.26	-0.08, 0.93	$F(2,1581)$	
Time × PV vs. RP	0.28	0.12	-0.17, 0.74	= 3.18		0.26	0.27	-0.27, 0.78	= 1.84	
NP2										
Intercept	-0.22	0.05	-0.32, -0.13			-0.22	0.08	-0.39, -0.06		
Time	-0.59	0.10	-0.79, -0.39	$F(1,4618)$ = 61.64	***	-0.62	0.13	-0.87, -0.36	$F(1,3317)$ = 9.91	**
{PV, RP} vs. AV	-0.13	0.16	-0.44, 0.17	$F(2,517)$		0.02	0.14	-0.26, 0.30	$F(2,1011)$	***
PV vs. RP	-0.06	0.10	-0.24, 0.13	= 0.99		0.01	0.14	-0.25, 0.28	= 8.48	
Time × {PV, RP} vs. AV	-2.22	0.22	-2.65, -1.79	$F(2,4010)$	***	-2.05	0.32	-2.67, -1.43	$F(2,1898)$	***
Time × PV vs. RP	-0.55	0.22	-0.99, -0.11	= 51.58		-0.40	0.28	-0.94, 0.14	= 23.79	

agent fixations decreased toward the end of this time window in sentences with sentence-final agent pivots but not in the other two sentence types. By-items, the interaction of time and sentence type did not reach statistical significance. There is, however, a significant main effect of sentence type meaning that there were more fixations to the agent for sentences in which it was mentioned first, i.e., sentences with patient pivot (5.1b) or recent perfective marking (5.2), as compared to agent pivot sentences.

During the *NP2* region, agent fixations decreased in sentences with patient pivots and recent perfective marking, in which the patient was mentioned in sentence-final position, as compared to agent pivot sentences with the agent in final position. In fact, fixations to the agent in the latter sentence type increased during this time window. Additionally, there was a steeper decrease in agent fixations for sentences where the patient was the pivot argument (5.1b) as compared to pivot-less recent perfective sentences in the by-subjects regression model. However, this effect was not detectable in the by-items model.

Table 5.3 shows the results of the quasi-logistic linear mixed effects regression models for fixations to the patient in the three analysis time windows. During the *Verb + Adverb* region, none of the predictors reaches statistical significance, indicating that listeners' fixations to the patient did not differ between sentence types and did not change while hearing the verb and the adverb.

During the *NP1* region, there were more patient fixations in sentences with final agent pivots (5.1a) in which the patient was mentioned during that region. Listeners started to direct their gaze to the patient in this sentence type only toward the end of the time window which might explain that a main effect of sentence type but no interaction with time was found. There were no differences in patient fixations between sentences with sentence-final patient

Table 5.3: Quasi-logistic linear mixed effects regression results predicting empirical logits of fixations to the patient referent in three different sentence types

	by subject					by item				
	$\hat{\beta}$	*SE*	95% Wald CI	*F* statistic		$\hat{\beta}$	*SE*	95% Wald CI	*F* statistic	
Verb + Adverb										
Intercept	-0.75	0.04	-0.84, -0.66			-0.81	0.09	-0.98, -0.64		
Time	-0.02	0.11	-0.24, 0.20	$F(1,10192) = 0.07$		0.01	0.17	-0.32, 0.34	$F(1,3736) = 0.01$	
{PV, RP} vs. AV	-0.07	0.07	-0.20, 0.07	$F(2,14425)$		-0.01	0.09	-0.19, 0.17	$F(2,14124)$	
PV vs. RP	0.04	0.08	-0.11, 0.19	$= 0.65$		0.03	0.06	-0.09, 0.16	$= 0.15$	
Time × {PV, RP} vs. AV	-0.23	0.23	-0.68, 0.22	$F(2,16429)$		>0.01	0.20	-0.40, 0.40	$F(2,27222)$	
Time × PV vs. RP	-0.13	0.24	-0.59, 0.34	$= 0.59$		-0.05	0.24	-0.53, 0.42	$= 0.02$	
NP1										
Intercept	-0.70	0.04	-0.79, -0.62			-0.78	0.07	-0.92, -0.64		
Time	0.10	0.10	-0.11, 0.31	$F(1,5489) = 0.31$		0.04	0.10	-0.16, 0.24	$F(1,3690) = 0.10$	
{PV, RP} vs. AV	-0.34	0.08	-0.49, -0.19	$F(2,2321)$	***	-0.23	0.10	-0.43, -0.02	$F(2,750)$	*
PV vs. RP	-0.15	0.09	-0.33, 0.02	$= 10.66$		-0.12	0.13	-0.37, 0.13	$= 3.89$	
Time × {PV, RP} vs. AV	-0.32	0.28	-0.88, 0.23	$F(2,2395)$		-0.17	0.29	-0.74, 0.40	$F(2,2015)$	
Time × PV vs. RP	-0.12	0.19	-0.50, 0.26	$= 0.83$		-0.09	0.26	-0.60, 0.43	$= 0.20$	
NP2										
Intercept	-0.35	0.04	-0.44, -0.26			-0.39	0.08	-0.56, -0.23		
Time	0.56	0.12	0.33, 0.79	$F(1,2228) = 43.27$	***	0.55	0.14	0.27, 0.83	$F(1,2624) = 31.00$	***
{PV, RP} vs. AV	0.04	0.16	-0.27, 0.35	$F(2,619)$		-0.02	0.14	-0.30, 0.27	$F(2,849)$	
PV vs. RP	-0.02	0.10	-0.22, 0.19	$= 0.28$		-0.04	0.11	-0.26, 0.17	$= 2.97$	
Time × {PV, RP} vs. AV	2.10	0.21	1.70, 2.50	$F(2,10989)$	***	1.97	0.40	1.18, 2.76	$F(2,1458)$	***
Time × PV vs. RP	0.47	0.18	0.11, 0.83	$= 53.99$		0.51	0.26	> -0.01, 1.03	$= 18.13$	

pivots (5.1b) and recent perfective marking (5.2) for which the agent was mentioned during this time window.

Finally, during the *NP2* region, there was a steep increase of patient fixations in sentences in which it was mentioned during this time window, i.e., patient voice and recent perfective sentences. Patient fixations decreased in sentences with agent pivots as they were mentioned sentence-finally. Additionally, in the by-subjects analysis, there was a steeper increase of patient fixations in sentences where it was the pivot (5.1). This effect is, however, barely detectable in the by-items analysis.

To test when listeners began to direct their gaze from the referent of *NP1* to the referent of *NP2*, breakpoint analyses were performed over the corresponding analysis time windows. These analyses test for discontinuities in the linear relations (Baayen, 2008), i.e., changes of direction of the regression lines for agent and patient fixations. Participants' agent fixations began to change before the beginning of *NP2* in all three sentence types (agent pivot sentences: before the first bin of *NP2* by-subjects and by-items; patient pivot sentences: before the last time bin of *NP1* by-subjects and before the first time bin of *NP2* by-items; recent perfective sentences: before the first time bin of *NP2* by-subjects and before the last time bin of *NP1* by-items). Participants' patient fixations began to change with very similar timing (agent pivot sentences: before the first bin of *NP2* by-subjects and by-items; patient pivot sentences: before the first bin of *NP2* by-subjects and by-items; recent perfective sentences: before the last time bin of *NP1* by-subjects and by-items).

In other words, before the onset of the second argument, listeners' fixations to the agent increased in agent voice-marked sentences where it was in sentence-final position and decreased in patient voice and recent perfective-marked sentences where the patient was in sentence-final position. Similarly, before the onset of *NP2*, patient fixations began to increase in the latter sentence types and began to decrease in sentences with agent pivots.

When controlling for agent or patient animacy (humans and animals vs. inanimates) or position within the experiment (first vs. second half), or when only items that occured in all three conditions are included (i.e., excluding scenarios with human patients as sentences with agent pivots are prohibited in these configurations), a similar pattern of results emerges for all three analysis time windows. However, the different slopes for sentences with patient pivots and recent perfective sentences during the *NP2* region that were found in the by-subjects analyses for agent and patient fixations are not consistently found when these control variables were included.

Especially the similar pattern of results that was found when the position of trials in the experiment was controlled (first vs. second half) suggests that participants' behavior was not influenced by an expectation to encounter pronominalized or zero anaphora arguments (cf. Himmelmann, 1999; Kroeger, 1993b). Participants seemed to be primed to encounter sentences with two full NP arguments by the practice trials at the beginning of the experiment; otherwise, some habituation over the course of the experiment modulating the effects of interest would have been expected.

Anticipatory baseline effects (Barr, Gann, & Pierce, 2011) influencing the interaction of time and sentence type are also not detectable when comparing the likelihood of agent or patient fixations during the preview and during the *Verb+Adverb* region (-400–200 ms relative to verb onset vs. 200 ms–*NP1* onset).

5.2.4 *Discussion*

The results of the current visual world experiment on Tagalog suggest that listeners used the lexical semantics of the verb to determine agent and patient referents. They directed their gaze toward the agent after they heard and recognized the verb. Interestingly, listeners focused on the agent in all three sentence types, irrespective of whether it was the pivot or not and therefore also irrespective of whether it could be expected to immediately follow the adverb or not. In contrast, while hearing the verb and the adverb, listeners did not direct their attention toward the patient.

Listeners did not seem to use information provided by the verbal morphology from which the syntactic function and the canonical position of arguments could be inferred for anticipation upon having heard the verb. If there were anticipation processes during the *Verb + Adverb* region based on syntactic information, i.e., if listeners either anticipated the final pivot argument or the linearly first NP, differences between sentences with agent pivots and sentences with patient pivots or recent perfective marking should have been found. Specifically, an increase in patient fixations would have been expected in sentences with agent pivots if anticipation was based on the linear order of NPs because in these sentences the patient canonically precedes the agent. Conversely, if anticipation was based on pivot status, an increase in patient fixation for sentences with patient pivots would have been expected. Yet, only fixations to the agent increased after listeners encountered the initial verb in all three sentence types.

Only after the adverb — during the *NP1* and *NP2* regions — did listeners gaze at agent and patient referents in their linear order. At least for the second argument (*NP2*), listeners seemed to anticipate the respective referent by directing their gaze toward the corresponding element before it was mentioned. Information provided by the verb and the first NP were integrated to predict the referent of the final argument. This interpretation is based on the consideration that programming a saccade typically takes approximately 200 ms (Duchowski, 2007) and there is also a lag between eye movements and the linguistic input of about the same time (Allopenna et al., 1998). Given that the slope of agent and patient fixations changed direction before the onset of the NP2 region in most cases, it may be assumed that listeners programmed their eye movements toward the agent (5.1a) or the patient referent (5.1b and 5.2) already well before having heard and parsed the corresponding noun in the linguistic input.

The results of the current experiment thus indicate that early anticipation of arguments in Tagalog is based on semantic roles and that the agent of the event in particular attracted listeners' attention once enough information about the event had accumulated to allow the identification of agent and patient referents. In Tagalog, the possibilities for prediction upon encountering the verb are not already narrowed down by previous linguistic input, unlike in subject-initial languages where one of the verb's arguments, often the agent, has already been mentioned. Thus, in this verb-initial language, it appears that what is targeted by anticipatory processes is primarily the semantics of the event.

Altmann and Kamide (2007) argue for a linking hypothesis between language processing and eye movements that allows verbs to drive anticipatory eye movements based on the affordances of the linguistic input and the visual display (cf. also Tanenhaus, Magnuson, Dahan, & Chambers, 2000). These affordances are the "properties of the possible interactions [...the depicted referents] could [...] engage in" (Altmann & Kamide, 2007, 513). Accordingly, the presence of a frog and a fly together with the auditory presentation of "eat" conspire to create a representation of the event that makes the frog a potential agent and the fly a potential patient. It is this episodic fit between the semantics of the described event and the depicted referents that drives listeners' eye movements toward the agent upon having heard the verb and before the agent NP was encountered.

5.3 CONCLUSIONS

A visual world experiment on a verb-initial language was presented that was set out to test what kind of information listeners are sensitive to during anticipatory processing in language comprehension. It was found that in Tagalog, listeners focus on the agent of the event upon having heard the sentence-initial verb. The lexical semantics of the verb together with the visual display allowed them to rapidly identify agent and patient referents. It seems that listeners did not use information provided by voice marking to specifically predict the syntactic functions or the linear order of arguments right after having heard the verb.

However, later in the sentence, specifically before the second noun was encountered, listeners did integrate all available information to anticipate the corresponding referent in the sentence-final position. This finding is similar to what has been found in English (Altmann & Kamide, 1999), German (Knoeferle et al., 2005) and Japanese (Kamide, Altmann, & Haywood, 2003). Thus, users of verb-initial languages also exhibit anticipatory gazes based on the linear order of arguments. Prediction of the final NP operates on a temporally more local level and occurs right before it is encountered whereas agent anticipation after the verb is independent of its position in the sentence.

It may be concluded that there are two kinds of anticipatory processes in Tagalog: one is oriented toward the sentence-level which uses verbal semantics to identify and focus on the agent of the event, the other one operates on a local scale and integrates information from the verb and the first argument to anticipate the sentence-final argument. Anticipation of the syntactic object in subject-initial languages could then possibly be seen as an instance of the latter, temporally more local, type.

Altmann and Kamide (2007) argue that anticipatory eye movements in sentence comprehension are driven by overlapping activations between representations of the visually presented objects and conceptual representations induced by the linguistic input. The results from the current experiment suggest that verbs especially facilitate anticipation based on semantic roles. Verbs provide event semantics to which potential referents in the visual display can be associated based on their affordances. Anticipatory eye movements might reflect listeners' knowledge about the dynamics of events in the world and are therefore not only reflecting "unfolding language [...but] an unfolding (mental) world" (Altmann & Kamide, 2007, 515).

One possible interpretation of the findings from Tagalog is thus that language users may engage in simulation-based anticipation when processing verb-initial sentences. Huettig (2016) suggests that there are several anticipatory mechanisms in language comprehension. One of these mechanisms engages perceptual simulation of events in order to predict their outcome and the linguistic structure with which they will be represented. Moulton and Kosslyn (2009) argue that simulation and mental imagery play a vital role for the prediction of future states of the world. Cohn and Paczynski (2013) propose an agent saliency principle that renders agents more prominent than patients in the processing of events in general (cf. also Bornkessel-Schlesewsky & Schlesewsky, 2013b; Kemmerer, 2012). Upon having heard the sentence-initial verb, Tagalog listeners identified the agent referent and might have focused on it because it was the initiator of the described event and was therefore necessary to build an event structural representation and to form expectations about the remainder of the sentence. The results of the current experiment are consistent with the idea that Tagalog listeners mentally simulated the event described by the verb after having encountered it (Pulvermüller, 2005). Agents might attract the most attention during the mental simulation of events because they function as cognitive attractors as they are the instigators of these events (Bornkessel-Schlesewsky & Schlesewsky, 2013a) and because the representation of agents and their actions is probably evolutionary ancient as it is already present in infants (Spelke & Kinzler, 2007).

The current findings are also in accord with approaches to sentence comprehension that assume agent identification to be an early processing step. Bornkessel and Schlesewsky (2006) posit that listeners try to identify the agent as quickly as possible. Many studies also show that sentences in which the agent precedes the patient are easier to process (F. Ferreira, 2003; Schriefers et al., 1995; Traxler et al., 2002 and Wang et al., 2009, inter alia).

Interestingly, the prominence of the agent role in comprehension processes in Tagalog has its reflexes in grammar, too. Schachter (1995a) shows that both pivots and agents are privileged in different syntactic constructions (cf. also Foley & Van Valin, 1984; Schachter, 1976). Riesberg and Primus (2015) argue that even in Tagalog's symmetrical voice system, where verbs are morphologically marked for agent as well as patient pivots, agents have a special grammatical status. For example, agents are always binders of reflexives, independently of their syntactic status (Schachter, 1977). Thus, although there is no grammatical preference for agents as pivots — and patient pivots are in fact more frequent in Tagalog texts —, agents seem to take a prominent role in both processing and grammar. This is surely to be attributed to their centrality for event cognition.

Focusing on a different kind of simulation than the mental simulation of events described above, Pickering and Garrod (2013) proposed that anticipation in language comprehension emerges through prediction by (linguistic) simulation of production processes (cf. also Dell & Chang, 2014; Pickering & Garrod, 2007). Under this view, listeners use the linguistic input that they have encountered at any given point in time to build an impoverished forward production model of what they would say if they were the speaker, just as people construct forward models of motor commands (Wolpert, Doya, & Kawato, 2003). The output of this forward production process is then matched against what was actually heard. Thus, the production system would be routinely employed during comprehension by covertly imitating the speaker's behavior in order to build expectations about the following linguistic material before it is encountered.

Based on eye tracking evidence from sentence production in Tagalog, it seems that the current experiment does not directly support this view. Sauppe et al. (2013) show that in early stages of Tagalog sentence production the pivot argument plays a prominent role — irrespective of its semantic role. In a picture description experiment, speakers preferentially fixated the character that was to become the pivot argument before uttering the sentence-initial verb in order to aid encoding the morphological marking. By contrast, the current experiment found that during sentence comprehension in the presence of visual stimuli, listeners directed their attention toward the agent irrespective of which argument was the pivot of the sentence. Taken together, these results suggest Tagalog speakers and listeners prioritize the processing of distinct kinds of information during the early stages of sentence encoding and decoding.

In other words, during early phases of sentence production Tagalog speakers focus their attention on pivot arguments. During comprehension, on the other hand, Tagalog listeners focus on the agent of the event early after having heard the sentence-initial verb. This suggests that different processes may be at play and that listeners did not immediately build a forward production model of the unfolding sentence to predict upcoming words. If this would have been the case, agent and patient fixations in sentences with agent pivots and patient pivots should have differed based on the differential semantic roles of the pivot arguments. When producing a sentence, Tagalog speakers need to choose a pivot argument and encode the relevant information in form of voice affixes on the verb and case markings on the arguments. When comprehending a sentence, language users do not have to engage in choosing a pivot argument themselves. They can thus rather concentrate on verbal semantics in order quickly build a representation of the described event.

Nevertheless, effects of agent prominence can also be detected in production processes in Tagalog as the planning of sentences with agent pivots exhibits lower cognitive load requirements than the production of sentences with patient pivots (?).

It may be noted that it can not be excluded that local thematic priming between verb and arguments had an influence on listeners' gaze behavior. Kukona et al. (2012) found that anticipatory fixations in a visual world sentence comprehension experiment on English were influenced by semantic priming from verbs when there were strong associations between the verb and its arguments (e.g., *arrest* together with *policeman* and *crook*). Most notably, upon having heard the verb, listeners looked at potential agent referents even if they were not mentioned. It is to be determined in future studies whether these results can also be explained by the relative saliency of agents in the build-up of event structural representations and in how far priming effects influence early agent fixations in Tagalog.

In general it can be concluded that the structure of the input guides the uptake and integration of visual and linguistic information. The current study shows that in addition to selectional restrictions and other structural information (Altmann & Kamide, 1999; Boland, 2005; Kamide, Altmann, & Haywood, 2003), the semantic roles of event participants might also be targeted by anticipation processes. Verb-initial languages might even favor the anticipation of semantic roles because information about the event is presented at the very beginning of an unfolding sentence and neither agent nor patient role are already (lexically) filled upon encountering the verb.

Altmann and Kamide (1999) propose that any information available to the listener is used to anticipate upcoming elements of an unfolding sentence. The results of the current experiment on Tagalog comprehension support this view. As soon as relevant information was available, listeners used selectional restrictions to identify the verb's arguments. Later on, accrued information about the event and the already encountered words was used to anticipate the final noun phrase of a sentence. Interestingly, upon having heard the verb, language users first directed their attention toward the agent, the instigator of the described event, independently of its syntactic status and its position in the sentence.

Going beyond the findings of previous visual world studies on subject-initial languages, the current experiment employed constructions in which the influence of event semantic information and syntactic information could be dissociated. It was shown that it was semantic information that was targeted early by predictive processes although syntactic information was also prominent and became relevant later. During the comprehension of languages with subject-initial word order, predictive processes on the basis of semantic roles might also operate. As mentioned in the introduction, when anticipating sentence-final syntactic objects, listeners could specifically predict the patient referent based on its role in the event described by the verb (cf. Kukona et al., 2012). This, however can not be observed as directly as in verb-initial languages because for the anticipation of sentence-final objects, semantic and syntactic information cannot be disentangled.

Tagalog has a relatively simple verbal morphology in the sense that only the semantic role of one of the arguments is cross-referenced. Future research should address whether the richness of verbal morphology has an influence on anticipatory processes. It could be possible that, e.g., person or number marking of pivot and non-pivot arguments (or subject and object for this purpose) on an initial verb triggers different anticipatory processes because more grammatical information about arguments is provided early.

To date, there are only few studies on online language processing in verb-initial languages (most notably Norcliffe, Konopka, et al., 2015; Sauppe et al., 2013; Wagers, Borja, & Chung, 2015). These languages provide valuable means to put to test processing theories and hypotheses that were developed based on the small set of languages that is usually used in psycholinguistics (such as English, German, Dutch or Japanese; cf. Jaeger & Norcliffe, 2009 on the most studied languages in sentence production research). Making use of the grammatical diversity of the world's languages will help to refine psycholinguistic theories and to uncover processes that cannot be observed by experimentation on the "usual suspect" languages (Levinson, 2012; Norcliffe, Harris, & Jaeger, 2015).

Part IV

CONCLUDING REMARKS

6

CONCLUDING REMARKS

This chapter provides a summary of the individual studies presented in this thesis.

6.1 SUMMARY OF RESULTS

6.1.1 *Chapter 2: The influence of voice marking morphology and word order on incremental sentence planning in Tagalog*

Chapter 2 investigated in two experiments how the time course of sentence production in Tagalog is shaped by its verb-initial word order and how the presence or absence of head-marking morphology influences sentence planning in this language.

At the beginning of this chapter it is argued that the predictions of linear incrementality are not suitable for the planning of sentences in which the verbal predicate is the first word. Linearly incremental production theories that have been proposed in the literature based on evidence from subject-initial languages assume that planning begins with the encoding of the most accessible nominal concept (e.g., Gleitman et al., 2007). By contrast, planning of the initial predicate in languages like Tagalog entails early relational encoding of the event because speakers need to know what action is being carried out in order to select a suitable verbal predicate. A planning account that assumes that formulation proceeds "word by word" must thus also assume that speakers encode event relations early in order to plan the first word. Linear incrementality and hierarchical incrementality thus make quite similar predictions for the planning of predicate-initial sentences. The Chapter 2 subsequently focused on hierarchically incremental accounts of sentence planning.

In the first experiment in Chapter 2 we compared agent voice sentences and patient voice sentences in Tagalog that differed in the kind of head-marking carried by the verbal predicate. In the first sentence type, the agent was selected as pivot argument and the predicate signaled its semantic role by carrying agent voice marking. In the second sentence type, the patient functioned as pivot argument. The findings from both structural choices (pivot selection and word order) and eye movement data (time course of fixations) support hierarchical incrementality. During pivot selection, which determined the morphological shape of the predicate, speakers took into account the patient's humanness and always selected human patients to be pivot, following a grammatical constraint (Latrouite, 2011), showing that speakers engaged in relational encoding early in order to determine the semantic roles and humanness of the characters depicted in the stimuli.

The order of agent and patient in voice-marked sentences mainly followed the canonical pivot-final word order. Additionally, human patients were more likely to immediately follow the predicate. Patient humanness plays a prominent role for voice choice because human patients must become pivots; they might therefore have been more accessible during word order selection and were thus positioned earlier.

The patterns of fixations to agent and patient characters showed that Tagalog speakers encode the event relations and select a pivot argument at the outset of formulation. In both agent voice and patient voice sentences the pivot character was fixated with priority in an early time window (0–600 ms after stimulus onset), irrespective of the eventual order of agent and patient after the predicate. After 600 ms, speakers fixated the characters in the stimulus pictures in the order of mention. We interpret this finding to signal relational encoding, pivot selection and the generation of an utterance plan before 600 ms and lexical encoding guided by this utterance plan after 600 ms.

In a second experiment, the production of recent perfective sentences in Tagalog was investigated. These sentences differ from agent voice and patient voice sentences in that there is no head-marking morphology on the predicate and the arguments carry default case marking. When planning recent perfective sentences, speakers thus did not have to select a pivot argument early in the formulation process.

Most notably, speakers were more likely to position conceptually lowly accessible, non-human agents immediately after the predicate when producing recent perfective sentences. This findings strongly supports hierarchically incremental production and is incompatible with linearly incremental accounts because it suggests that speakers' planning scope was wide and that lexical retrieval was guided by an utterance representation. We discussed the possibility that this is an effect of hearer-oriented planning processes (Kurumada & Jaeger, 2015). Tagalog speakers may have been more likely to place non-human agents in their canonical position in order to avoid potential semantic role ambiguities because unlike in voice-marked sentences, voice affixes and case marking were not available as cues to the arguments' semantic roles. However, it is left to future research to explore the possibility of hearer-oriented effects on the planning of recent perfective sentences in Tagalog as the experiments in Chapter 2 were not designed to test this.

Finally, the fixation patterns observed during the production of recent perfective sentences with differing orders of agent and patient indicate that speakers also began planning with relational encoding of the events. To investigate the influence of head-marking morphology on the time course of formulation, we compared agent voice sentences and recent perfective sentences with [agent patient] order. Compared to the production of voiceless recent perfective sentences, speakers preferentially fixated pivot characters when planning voice-marked sentences in an early time window. We take this to confirm that the pivot fixations observed in Experiment 1 of this chapter indeed reflect pivot selection and encoding of the voice affix because the two sentence types only differed with respect to the presence or absence of voice marking. Thus, we were able to show a reflex of head-marking morphology on the online planning of sentences.

In sum, the research in Chapter 2 for the first time — to the best of our knowledge — reported online evidence from sentence production in an Austronesian language and showed the effects of planning head-marking morphology on the time course of sentence formulation and on the planning strategies that speakers employ.

6.1.2 *Chapter 3: Voice, word order and verb planning in German sentence production*

Chapter 3 investigated relational encoding and the selection of verbs during the production of German sentences. Using the same paradigm and picture stimuli as in Experiment 2 from Chapter 2, fixations to the agent and patient characters in active and passive sentences with verb-medial and verb-final word orders were compared in order to assess whether the position of the verb influences the time course of sentence planning.

In German subject-initial sentences, the lexical main verb may either occur immediately after the subject (V-medial) or at the end of the sentence if the second position is filled by another inflected verbal element, such as an auxiliary (V-final). Linear and hierarchical incrementality make different predictions for the planning of these sentence types, which were tested by analyzing how speakers distributed their visual attention to subject characters (agent in actives, patient in passives) and object/oblique characters before and after speech onset. Additionally, the factors driving speakers' choice between active and passive structures were investigated.

Hierarchical incrementality assumes that speakers always encode the event first and that lexical retrieval is then guided by a structural-relational representation of the utterance. Two different versions of hierarchical incrementality were introduced in Chapter 3. These approaches differ with respect to whether they assume that the generation of the utterance plan only includes conceptual encoding of the action in the described event (weakly hierarchical incrementality; e.g., Griffin & Bock, 2000) or whether speakers already select a verb early in the planning process (strongly hierarchical incrementality; e.g., Kempen & Huijbers, 1983). Consequently, weakly and strongly hierarchical incrementality make different assumptions about whether speakers can assign syntactic functions and make structural choices on the basis of a conceptual and relational representation of the event or whether verb lemmas control syntactic function assignment so that articulation of the subject could not be initiated before a verb is selected.

Strongly hierarchical incrementality predicted that the patterns of fixations to agent and patient characters in a picture description task should not differ before speech onset for different German sentence types (V-medial actives, V-final actives, and passives) because speakers should engage in early relational encoding and verb selection followed by linguistic encoding irrespective of the sentence's word order. Weakly hierarchical incrementality, by contrast, predicted different fixation patterns for V-medial and V-final sentences. This view assumes that linguistic encoding proceeds in the order of mention and while this means that verbs are not necessarily always planned at the outset of formulation, it implies that speakers should begin to encode the verb in V-medial sentences already before speech onset in order to avoid disfluencies in speaking (Griffin, 2003). It was predicted that this would lead to more evenly distributed agent and patient fixations in these sentences as compared to V-final sentences where speakers would not need to begin to prepare the verbs already before speech onset.

With respect to voice choice, both weakly and strongly hierarchical incrementality allow speakers to take into account the humanness configuration between agent and patient to select a subject. For this, it would be necessary to assign semantic roles to the depicted characters first, based on a representation of the to-be-described action.

Linear incrementality, on the other hand, assumes that speakers begin planning by encoding the sentence-initial subject. Under this account, first-fixated or human characters should be generally more likely to function as subject. Linear incrementality also predicted that relational encoding might be postponed in V-final sentences as compared to V-medial sentences because the verb would only need to be produced at the end of the sentence in the former but immediately after the subject in the former. Thus, in V-medial sentences, verb would need to be prepared earlier to ensure that speakers can maintain fluency (Griffin, 2003). The prediction for the experiment in Chapter 3 was therefore that speakers would distribute their fixations more between agent and patient in the stimulus pictures when planning V-medial actives, where the verb was positioned early, as compared to V-final actives and passives.

Analyses of eye movements revealed that German speakers indeed distributed their attention more between agent (subject) and patient (object) before speech onset when they planned V-medial actives as compared to V-final actives, suggesting that they engaged in early relational encoding and verb planning to a higher degree when the verb was positioned early — in line with the predictions from both linear and weakly hierarchical incrementality. The fixation patterns during the production of both V-medial and V-final actives also differed from passives. Speakers distributed their fixations even more between agent and patient characters when planning passives as during the planning of the two kinds of active sentences.

Crucially, the relative humanness of agent and patient influenced speakers' choice between producing active or passive sentences, while first fixations had no effect on what voice speakers produced. Human patients were more likely to become subjects, even more so when the agent was non-human. This finding is consistent with hierarchically incremental production accounts because it presupposes that speakers first engaged in relational encoding in order to know the semantic roles of event participants before assigning syntactic functions. Moreover, this is incompatible with linear incrementality which does assumed that the accessibility of individual referents, but not conditional on their roles in the event, could have influenced structural choices.

The results of the experiment reported in Chapter 3 therefore support weakly hierarchical incrementality and suggest that the timing of verb planning in German sentences is influenced by the word order of the sentence currently under production. Speakers always engaged in relational encoding early in the sentence planning process, encoding the depicted action and the semantic roles and humanness of the picture characters and using this information to generate a structural-relational representation of the utterance. However, speakers only prepared the verb before speech onset when planning V-medial sentences, where they had to utter the verb immediately after the subject.

Furthermore, the fixation differences between actives and passives showed that speakers engaged in more relational planning for the latter. This could be due to the fact that passives are often used to describe unusual humanness relations where a human patient is being acted upon, the relative infrequency of passives and the potentially different planning mechanisms for actives and passives in German (cf. the results for the temporal development of cognitive processing load for actives vs. passives in Chapter 4).

In sum, the study in this Chapter 3 showed that German speakers begin the planning of simple declarative sentences with the formation of an event representation that allows them to assign semantic roles and to assess the humanness of agent and patient in the depicted event. The eye movement patterns before speech onset suggest that the timing of fine-grained relational encoding for the purpose of verb selection in actives is contingent on the position of the verb in the sentence. This pattern of results — the early generation of an utterance plan in all sentence types, but different timing of verb selection — support a weakly hierarchically incremental view of sentence production.

6.1.3 *Chapter 4: Voice systems and processing load during sentence planning*

After Chapter 2 and Chapter 3 explored when and how Tagalog speakers and German speakers encode event relations, decide on an utterance plan and retrieve the individual words of their sentences, Chapter 4 took a broader perspective by exploring the processing consequences of the grammatical architectures of the voice systems of two languages. Specifically, differences in the mental effort required to plan and produce actives as compared to passives in German and to plan and produce agent voice sentences as compared to patient voice sentences in Tagalog were investigated.

The voice systems of Tagalog and German exemplify two different kinds of voice systems. In the symmetrical voice system of Tagalog there are no unmarked voice forms, all voices are equally transitive and overtly marked by voice affixes on the predicate. In the asymmetrical voice system of German, by contrast, the active is the unmarked voice form for the description of transitive events. The passive is the marked voice form in German because it goes in hand with overt morphological marking and de-transitivisation. Thus, while the different voices behave morphosyntactically similar in Tagalog, there are sharp differences between the behaviour and properties of actives and passives in German.

For the first time, this study tested for processing consequences of the distinction between symmetrical and asymmetrical voice systems. For this, pupillometric data that were recorded together with eye movement data in the experiments in Chapter 2 (Experiment 1, Tagalog) and Chapter 3 (German) were analysed. The mental effort associated to the planning and production of different voices in the two languages was investigated by comparing the time course of pupil size changes while Tagalog speakers produced agent voice and patient voice sentences and German speakers produced actives and passives.

Pupil size changes are a suitable psychophysiological measure of mental effort and processing load because they are indicative of activity in the locus coeruleus, which plays a major role in the modulation of attention and arousal in the brain (Sara, 2009). This measure can thus be used to observe processing load detached from behavioral performance in psycholinguistic experiments.

The analyses revealed that the time course of pupil size changes was similar for agent voice and patient voice sentences in Tagalog, indicating that processing load over time developed in similar ways. This supported the hypothesis that speakers have to perform the same planning operations during the formulation of both voice types, namely selecting one argument to function as pivot and encoding the corresponding voice affix (cf. Chapter 2). This is in line with what was predicted for a symmetrical voice system.

On the other hand, processing load developed differently over time for the formulation of actives and passives in German. When planning passives, speakers' pupils dilated earlier, suggesting that different planning operations might have been carried than during the preparation of active sentences (cf. Chapter 3). This might reflect the fact that passives are morphologically marked in comparison to actives.

An additional finding of the study in Chapter 4 was the identification of an anti-Patient bias in sentence production in Tagalog. While both agent voice and patient voice sentences exhibited the same temporal dynamics of processing load changes, pupil size stayed more dilated towards the end of the analysis time window for patient voice sentences than for agent voice sentences, indicating sustained mental effort. In the sentence comprehension literature it is well-established that sentences which link patients to the highest syntactic function are more difficult to process. The sustained cognitive processing load for patient voice sentences in the current study is thus interpreted to show a similar effect in production, being a reflex of the mentally effortful operation of linking the patient to pivot function during sentence planning.

In sum, the study in Chapter 4 demonstrates that the distinction between asymmetrical and symmetrical voice systems from the typological literature has a measurable effect on online sentence planning. Processing load changes follow different time courses during the production of sentences in different kinds if voice systems. Thus, the grammatical analysis of Tagalog as being a symmetrical voice language is supported by these results.

6.1.4 *Chapter 5: Voice and word order in Tagalog sentence comprehension*

Finally, Chapter 5 turned to the role of verb-initial word order and voice marking in the comprehension of Tagalog sentences, focusing on prediction and anticipatory processing.

Anticipation is an important aspect of sentence comprehension. Chapter 5 reviews the evidence from the literature that shows that listeners generally predict upcoming linguistic input. However, heretofore only studies on subject-initial languages were conducted. In these languages, semantic roles, syntactic functions and linear order of arguments are often conflated. In this chapter it is argued that it is therefore dissociate what kind of linguistic information listeners direct their attention to during anticipation when only subject-initial languages are investigated.

Tagalog's verb-initial word order and its voice system allowed to disentangle whether semantic roles, syntactic functions or linear order were the target of anticipation. In a visual world experiment, Tagalog speakers heard transitive sentences in agent voice, patient voice, or recent perfective aspect and at the same time saw a display that contained a potential agent, a potential patient and an unrelated distractor. Analyses of participants' eye movements while listening to these sentences revealed that agent fixations increased immediately after the verb was presented, irrespective of its morphological marking. This suggests that participants used the lexical semantics of the sentence-initial verbs together with contextual information from the visual display to rapidly identify the agents and patients of the described events and to consequently orient their visual attention to the agent referents.

Apparently, listeners did not make use of information about whether the agent or the patient was the pivot (based on voice marking in agent voice and patient voice sentences) or whether the agent or the patient would be mentioned first (based on the canonical pivot-final word order of voice-marked sentences and the canonical [agent patient] order in recent perfective sentences), although this information was also provided by or could have been inferred from the sentence-initial verb.

However, later in the sentence, listeners did integrate all available information and directed their gaze towards the second referent (the agent in agent voice sentences and the patient in patient voice and recent perfective sentences), anticipating its mention.

It is concluded that there are two kinds of anticipation processes in Tagalog sentence comprehension. The first process is oriented towards prediction on the sentence-level and focuses on the agent of the event that is described; it is based on the semantics if the verb and is independent of voice marking. Verbs may especially facilitate anticipation based on semantic roles because they provide immediate access to semantics of the described event. Thus, early during Tagalog comprehension, listeners may identify referents from the visual displays based on the affordances and selectional restrictions of these verbs.

The second process operates on a smaller scale and occurs later in the sentence where listeners integrate the information from the verb and the first argument phrase to anticipate the sentence-final argument, similarly to what has been observed in subject-initial languages.

Thus, this visual world study suggests that anticipation in verb-initial languages is primarily based on semantic roles because this can be extracted from the information provided at the beginning of a sentence. Chapter 5 thus contributes to the literature on sentence comprehension by shifting the focus from investigating the anticipation of sentence-final elements towards including constructions that allow for a wider set of possible anticipation targets, helping to identify the kinds of information that are used by comprehenders during predictive processing.

6.2 CONCLUSIONS

The studies in reported in Chapters 2–5 explored how voice and word order influence the incremental production and comprehension of sentences in Tagalog and German. Each of these chapters contributes to the psycholinguistic literature by exploring novel aspects of sentence production and comprehension.

Chapter 2 reports the first eye-tracked production study on an Austronesian language and shows for the first time how head-marking morphology influences planning processes. Chapter 3 demonstrates the timing of verb selection during German sentence production in a temporally more fine-grained manner than previous studies (e.g., Schriefers et al., 1998).

Taken together, these picture description experiments on German and Tagalog showed that speakers always engaged in relational encoding early during the formulation process in order to be able to assess the semantic roles and the humanness of the depicted characters. The production of Tagalog recent perfective sentences even suggested hearer-oriented planning processes, which I argue is based on an utterance plan generated early on. The evidence for early relational encoding found in all of these experiments broadly supports hierarchically incrementality. Interestingly enough, German and Tagalog speakers planned their sentences hierarchically incrementally even when linearly incremental planning would have been easily available, namely in V-final active (and passive) sentences in German and in recent perfective sentences in Tagalog. However, the differences in eye movement patterns for the planning of V-medial and V-final actives and passives in German suggest that the early relational encoding and the generation of an utterance plan might leave some flexibility for carrying out more fine-grained encoding and planning operations later in the sentence, contingent on word order and thus only at a time at which this information is finally needed.

Chapter 4 reports the first study to show the processing consequences of the distinction between asymmetrical and symmetrical voice systems as proposed by, e.g., Foley (2008) and Riesberg (2014b), substantiating these typological proposals (which are based solely on grammatical observations) with experimental psychophysiological evidence. Additionally, this study is the first to investigate the temporal dynamics of cognitive processing load during the planning and formulation of sentences. This demonstrates that time course analyses of pupil size changes can be used to investigate more complex production phenomena than single word planning (e.g., Papesh & Goldinger, 2012).

Finally, Chapter 5 reports the first visual world sentence comprehension experiment on a verb-initial language (cf. Wagers et al., 2015, for another recent study on verb-initial comprehension). By exploiting Tagalog's word order and voice system, it was possible to disentangle different kinds of targets during anticipatory processing. The results suggest that verb-initial word may enforce the anticipation of upcoming arguments on the basis of the affordances of verbs and their arguments' semantic roles.

Additionally, this study suggests that different kinds of linguistic information take precedence during sentence production and comprehension in Tagalog because comprehenders focused on semantic roles during listening (disregarding morphosyntactic voice marking cues), whereas Chapter 2 showed that during sentence formulation, syntactic functions play an important role because they have to be assigned early in order to plan the morphological form of the sentence-initial verb. This suggests that sentence production and sentence comprehension afford different operations in Tagalog, which is interesting in the light of recent proposals spelling out the relationship between production and comprehension. Pickering and Garrod (2013) present a model in which comprehension is a central part of production because speakers use their comprehension system to monitor their own production, as originally proposed by Levelt (Levelt, 1989; Wheeldon & Levelt, 1995). In Pickering and Garrod's model, sentence planning is also a central part of sentence comprehension because listeners use their production system to predict the upcoming words in the utterance they are currently parsing. How the comprehension and production of sentences interact in Tagalog, or verb-initial languages in general, and how a language's word order influences whether and how these two modes of language processing are intertwined is left for future research.

In sum, the research presented in this thesis shows that the grammatical properties of languages influence how speakers plan sentences and how comprehenders parse the linguistic input. In order to build better psycholinguistic models of language processing in the future, it is necessary to incorporate more evidence from heretofore under-studied languages because this will help to make the theories more universally applicable and to eliminate biases in theory building that result from a narrow empirical basis that misrepresents the grammatical diversity of the world's languages (Evans & Levinson, 2009; Jaeger & Norcliffe, 2009; Levinson, 2012; Levinson & Evans, 2010).

REFERENCES

Adams, K. L., & Manaster-Ramer, A. (1988). Some questions of topic/focus choice in Tagalog. *Oceanic Linguistics*, *27*(1–2), 79–101.

Adelaar, A., & Himmelmann, N. P. (Eds.). (2005). *The Austronesian languages of Asia and Madagascar*. Oxon: Routledge.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: John Wiley and Sons.

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119–126. doi: 10.1080/00031305.1998.10480550

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science*, *205*(4412), 1289–1292. doi: 10.1126/science.472746

Aissen, J. (1999). Markedness and subject choice in optimality theory. *Natural Language & Linguistic Theory*, *17*(4), 673–711. doi: 10.1023/A:1006335629372

Alday, P. M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). Towards a computational model of actor-based language comprehension. *Neuroinformatics*, *12*(1), 143–179. doi: 10.1007/s12021-013-9198-x

Aldridge, E. (2012). Antipassive and ergativity in Tagalog. *Lingua*, *122*(3), 192–203. doi: 10.1016/j.lingua.2011.10.012

Allopenna, P. D., Magnu, & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. doi: 10.1006/jmla.1997.2558

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 791–810. doi: 10.1037/0278-7393.33.4.791

Alnæs, D., Handal Sneve, M., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14*(4). doi: 10.1167/14.4.1

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi: 10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. doi: 10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. doi: 10.1016/0010-0277(88)90020-0

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language*, *76*(1), 28–55.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. doi: 10.1016/j.jml.2007.09.002

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*(328). doi: 10.3389/fpsyg.2013.00328

Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, *137*(2), 201–207. doi: 10.1016/j.actpsy.2010.09.011

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. *Frontiers in Psychology*, *7*(1858). doi: 10.3389/fpsyg.2016.01858

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015, June). Parsimonious Mixed Models. *ArXiv e-prints*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. Retrieved from `http://CRAN.R-project.org/package=lme4` doi: 10.18637/jss.v067.i01

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. doi: 10.1037/0033-2909.91.2.276

Bergamin, O., & Kardon, R. H. (2003). Latency of the pupil light reflex: Sample rate, stimulus intensity, and variation in normal subjects. *Investigative Ophthalmology & Visual Science*, *44*(4), 1546–1554. doi: 10.1167/iovs.02-0468

Bickel, B., Witzlack-Makarevich, A., Choudhary, K. K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE*, *10*(8), e0132819. doi: 10.1371/journal.pone.0132819

Blake, F. R. (1906). Expression of case by the verb in Tagalog. *Journal of the American Oriental Society*, *27*, 183–189.

Bloomfield, L. (1917). *Tagalog texts with grammatical analysis*. Urbana: University of Illinois.

Bock, K. (1986). Exploring levels of processing in sentence production. In G. Kempen (Ed.), *Natural language generation: New* *results in artificial intelligence, psychology, and linguistics* (pp. 351–363). Dordrecht / Boston / Lancaster: Nijhoff.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication: Handbook of perception and cognition* (pp. 181–216). San Diego, CA: Academic Press.

Bock, K., & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 21–46). Oxford: Oxford University Press.

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting First Things First. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 249–278). New York / Hove: Psychology Press.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*(4), 653–685. doi: 10.1016/S0749-596X(03)00007-X

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (1st ed., pp. 945–984). San Diego, CA: Academic Press.

Bock, K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47–67. doi: 10.1016/0010-0277(85)90023-X

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9–10), 341–345.

Boland, J. E. (2005). Visual arguments. *Cognition*, *95*(3), 237–274. doi: j.cognition.2004.01.008

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1995). Verb Argument Structure in Parsing and Interpretation: Evidence from wh-Questions. *Journal of Memory and Language*, *34*(6), 774–806. doi: 10.1006/jmla.1995.1034

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821. doi: 10.1037/0033-295X.113.4.787

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass*, *3*(1), 19–58.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013a). Neurotypology: Modeling crosslinguistic similarities and differences in the neurocognition of language comprehension. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language down the garden path: The cognitive and biological basis for linguistic structures* (pp. 241–252). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199677139.003.0012

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013b). Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76. doi: 10.1016/j.bandl.2013.01.010

Branigan, H. P., & Feliki, E. (1999). Conceptual accessibility and serial order in Greek speech production. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st annual conference of the cognitive science society* (pp. 96–101). Mahwah, NJ: Lawrence Erlbaum Associates.

Branigan, H. P., Pickering, M. J., & Tanaka, M. N. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*(2), 172–189. doi: 10.1016/j.lingua.2007.02.003

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, *109*(2), 274–280. doi: 10.1016/j.cognition.2008.07.011

Brown-Schmidt, S., & Konopka, A. E. (2015). Processes of incremental message planning during conversation. *Psychonomic Bulletin & Review*, *22*(3), 833–843. doi: 10.3758/s13423-014-0714-2

Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, *4*(25). doi: 10.3389/fnhum.2010.00025

Bunger, A., Papafragou, A., & Trueswell, J. C. (2013). Event structure influences language production: Evidence from structural priming in motion event description. *Journal of Memory and Language*, *69*(3), 299–323. doi: 10.1016/j.jml.2013.04.002

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. doi: 10.1037/0033-295X.113.2.234

Choudhary, K. K., Schlesewsky, M., Roehm, D., & Bornkessel-Schlesewsky, I. (2009). The N400 as a correlate of interpretively relevant linguistic rules: Evidence from Hindi. *Neuropsychologia*, *47*(13), 3012–3022. doi: 10.1016/j.neuropsychologia.2009.05.009

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–558. doi: 10.1017/S0140525X08004998

Christianson, K. T., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, *98*(2), 105–135. doi: 10.1016/j.cognition.2004.10.006

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. doi: 10.1017/S0140525X12000477

Clerget, E., Winderickx, A., Fadiga, L., & Olivier, E. (2009). Role of Broca's area in encoding sequential human actions: a virtual lesion study. *NeuroReport*, *20*(16), 1496–1499. doi: 10.1097/WNR.0b013e3283329be8

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied* *multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology*, *67*(3), 73–97. doi: 10.1016/j.cogpsych.2013.07.002

Cole, P., Hermon, G., & Yanti. (2008). Voice in Malay/Indonesian. *Lingua*, *118*(10), 1500–1553. doi: 10.1016/j.lingua.2007.08.008

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., ... Vogeley, K. (2006). Neural representations of Self versus Other: Visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience*, *18*(6), 898–910. doi: 10.1162/jocn.2006.18.6.898

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in Cognitive Sciences*, *7*(12), 527–533. doi: 10.1016/j.tics.2003.10.004

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. doi: 10.1037/0033-295X.93.3.283

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of The Royal Society B*, *369*(1634), 1–9. doi: 10.1098/rstb.2012.0394

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–11121. doi: 10.1038/nn1504

Demiral, Ş. B., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: Evidence from Turkish. *Cognition*, *106*(1), 484–500.

De Smedt, K. J. M. J. (1994). Parallelism in Incremental Sentence

Generation. In G. Adriaens & U. Hahn (Eds.), *Parallel Natural Language Processing* (pp. 421–447). Norwood, NJ: Ablex.

Dobel, C., Glanemann, R., Kreysa, H., Zwitserlood, P., & Eisenbeiß, S. (2011). Visual encoding of coherent and non-coherent scenes. In J. Bohnemeyer & E. Pederson (Eds.), *Event representation in language and cognition* (Vol. 11, pp. 189–215). Cambridge: Cambridge University Press.

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, *125*(2), 129–143. doi: 10.1016/j.actpsy.2006.07.004

Donnelly, S., & Verkuilen, J. (2017). Empirical logit analysis is not logistic regression. *Journal of Memory and Language*, *94*, 28–42. doi: 10.1016/j.jml.2016.10.005

Donohue, M. (2008). Malay as a mirror of Austronesian: Voice development and voice variation. *Lingua*, *118*(10), 1470–1499. doi: 10.1016/j.lingua.2007.08.007

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (chap. 81). Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `http://wals.info/`

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (2nd ed.). London: Springer.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79–82. doi: 10.1038/nature09923

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. doi: 10.1016/S0022-5371(81)90220-6

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension.

The Quarterly Journal of Experimental Psychology, *63*(4), 639–645. doi: 10.1080/17470210903469864

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–492. doi: 10.1017/S0140525X0999094

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., ... Fadiga, L. (2009). Encoding of human action in Broca's area. *Brain*, *132*(7), 1980–1988. doi: 10.1093/brain/awp118

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, *33*(6), 715–736. doi: 10.1006/jmla.1994.1034

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Philadelphia: Psychology Press.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*(2), 164–203. doi: 10.1016/S0010-0285(03)00005-7

Ferreira, F., & Engelhardt, P. E. (2006). Syntax and production. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 61–91). London: Academic Press.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. doi: 10.1006/jmla.2001.2797

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. In B. H. Ross (Ed.), *Advances in research and theory* (Vol. 49, pp. 209–246). London: Academic Press. doi: 10.1016/S0079-7421(08)00006-6

Ferreira, V. S. (2010). Language production. *Wiley Interdisciplinary*

Reviews: Cognitive Science, *1*(6), 834–844. doi: 10.1002/wcs.70

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. doi: 10.1006/cogp.1999.0730

Ferreira, V. S., & Slevc, L. R. (2007). Grammatical encoding. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 453–469). Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198568971.001.0001

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, *32*(6), 669–692. doi: 10.1023/A:1026146332132

Foley, W. A. (2008). The place of Philippine languages in a typology of voice systems. In P. K. Austin & S. Musgrave (Eds.), *Voice and grammatical relations in Austronesian languages* (pp. 22–44). Stanford, CA: CSLI Publications.

Foley, W. A., & Van Valin, R. D. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (Second ed.). Thousand Oaks, CA: Sage. Retrieved from `http://socserv.socsci.mcmaster.ca/jfox/Books/Companion` (R package version 2.1-1)

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory*, *5*(4), 519–559.

Frith, U., & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of The Royal Society B*, *365*(1537), 165–175. doi: 10.1098/rstb.2009.0160

Gabay, S., Pertzov, Y., & Henik, A. (2011). Orienting of attention, pupil size, and the norepinephrine system. *Attention, Perception, & Psychophysics*, *73*(1), 123–129. doi: 10.3758/s13414-010-0015-4

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007).

On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–596. doi: 10.1016/j.jml.2007.01.007

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, *21*(2), 90–95. doi: 10.1177/0963721412436811

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), B1–B14. doi: 10.1016/S0010-0277(01)00138-X

Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, *10*(3), 603–609. doi: 10.3758/BF03196521

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. doi: 10.1111/1467-9280.00255

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, *142*(3), 880–905. doi: 10.1037/a0030045

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934–946. doi: 10.1038/nrn2497

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models — the R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–32.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192. doi: 10.1126/science.143.3611.1190

Himmelmann, N. P. (1999). The lack of zero anaphora and incipient person marking in Tagalog. *Oceanic Linguistics*, *38*(2), 231–269.

Himmelmann, N. P. (2002). Voice in western Austronesian: an update. In F. Wouk & M. Ross (Eds.), *The history and typology of western Austronesian voice systems* (pp. 7–16). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Himmelmann, N. P. (2004). Tagalog (Austronesian). In G. Booj, C. Lehmann, J. Mugdan, S. Skopeteas, & W. Kesselheim (Eds.), *Morphology: An International Handbook on Inflection and Word-Formation* (Vol. 17, pp. 1473–1490). Berlin / New York: Mouton de Gruyter.

Himmelmann, N. P. (2005a). The Austronesian languages of Asia and Madagascar: Typological characteristics. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 110–181). Oxon: Routledge.

Himmelmann, N. P. (2005b). Tagalog. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 350–376). Oxon: Routledge.

Himmelmann, N. P. (2008). Lexical categories and voice in Tagalog. In P. K. Austin & S. Musgrave (Eds.), *Voice and grammatical relations in Austronesian languages* (pp. 247–293). Stanford, CA: CSLI Publications.

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *The American Statistician*, *52*(2), 181–184. doi: 10.1080/00031305.1998.10480559

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Huettig, F. (2016). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135. doi: 10.1016/j.brainres.2015.02.014

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. doi: 10.1016/j.actpsy.2010.11.003

Huettig, F., Singh, N., & Mishra, R. K. (2011). Language-mediated visual orienting behavior in low and high literates. *Frontiers in Psychology*, *2*(285). doi: 10.3389/fpsyg.2011.00285

Hwang, H., & Kaiser, E. (2014). The role of the verb in grammatical function assignment in English and Korean. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1363–1376. doi: 10.1037/a0036797

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *48*(3), 598–612.

Iwasaki, N. (2011). Incremental sentence production: Observations from elicited speech errors in Japanese. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and producing head-final structures* (Vol. 38, pp. 131–151). Dordrecht: Springer. doi: 10.1007/978-90-481-9213-7_7

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. doi: 10.1016/j.jml.2007.11.007

Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1439–1449. doi: 10.1037/a0027862

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, *3*(4), 866–887. doi: 10.1111/j.1749-818x.2009.00147.x

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing.

Canadian Journal of Experimental Psychology, *47*(2), 310–339. doi: 10.1037/h0078820

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. doi: 10.1126/science.154.3756.1583

Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1274–1295. doi: 10.1037/a0027626

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647–670. doi: 10.1111/j.1749-818X.2008.00072.x

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. doi: 10.1016/S0749-596X(03)00023-8

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55. doi: 10.1023/A:1021933015362

Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*(2), 175–185. doi: 10.1111/j.1469-8986.2004.00147.x

Kemmerer, D. (2012). The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area. *Language and Linguistics Compass*, *6*(1), 50–66. doi: 10.1002/lnc3.322

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*(2), 201–258. doi: 10.1016/S0364-0213(87)80006-X

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: indirect election of words. *Cognition*, *14*(2), 185–209. doi: 10.1016/0010-0277(83)90029-X

Kempen, G., Olsthoorn, N., & Sprenger, S. (2012). Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes*, *27*(3), 345–380. doi: 10.1080/01690965.2010.544583

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. doi: 10.2307/2533558

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, *7*(12), 1299–1301. doi: 10.1038/nn1355

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. doi: 10.1016/j.cognition.2004.03.002

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *Journal of the Acoustical Society of America*, *139*(4), 1618–1636. doi: 10.1121/1.4944032

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162. doi: 10.1016/j.jml.2011.08.003

Konopka, A. E., & Kuchinsky, S. E. (2015). How message similarity shapes the timecourse of sentence formulation. *Journal of*

Memory and Language, *84*, 1–23. doi: 10.1016/j.jml.2015.04.003

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40. doi: 10.1016/j.cogpsych.2014.04.001

Kratzer, A. (1996). Severing the external argument from its verb. In J. Rooryck & L. Zaring (Eds.), *Phrase structure and the lexicon* (Vol. 33, pp. 109–137). Dordrecht: Springer. doi: 10.1007/978-94-015-8617-7

Kroeger, P. (1993a). Another look at subjecthood in Tagalog. *Philippine Journal of Linguistics*, *24*(2), 1–16.

Kroeger, P. (1993b). *Phrase structure and grammatical relations in Tagalog*. Stanford, CA: CSLI Publications.

Kuchinke, L., Võ, M. L.-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*(2), 132–140. doi: /10.1016/j.ijpsycho.2007.04.004

Kuchinsky, S. E. (2009). *From seeing to saying: Perceiving, planning, producing* (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.

Kuchinsky, S. E., Ahlstrom, J. B., Vaden Jr, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*(1), 23–34. doi: 10.1111/j.1469-8986.2012.01477.x

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 748–756. doi: 10.1037/a0022637

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2012). The time course of anticipatory constraint integration. *Cognition*, *119*(1), 23–42. doi: 10.1016/j.cognition.2010.12.002

Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, *83*, 152–178. doi: 10.1016/j.jml.2015.03.003

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. doi: 10.1126/science.7350657

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cognitive Processing*, *12*(1), 13–21. doi: 10.1007/s10339-010-0370-z

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. doi: 10.1177/1745691611427305

Latrouite, A. (2011). *Voice and case in Tagalog: the coding of prominence and orientation* (Unpublished doctoral dissertation). Heinrich-Heine-Universität Düsseldorf.

Latrouite, A. (2015). Shifting perspectives: Case marking restrictions and the syntax-semantics-pragmatics interface. In J. Fleischhauer, A. Latrouite, & R. Osswald (Eds.), *Exploring the syntax-semantics-pragmatics interface* (pp. 285–314). Düsseldorf: Düsseldorf University Press.

Lee, E.-K., Brown-Schmidt, S., & Watson, D. G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition*, *129*(3), 544–562. doi: 10.1016/j.cognition.2013.08.007

Lemon, J. (2006). Plotrix: a package in the red light district of R. *R-News*, *6*(4), 8-12.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: a blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 83–122). Oxford: Oxford University Press.

Levelt, W. J. M. (2013). *A history of psycholinguistics: The pre-Chomskyan* *era*. Oxford: Oxford University Press.

Levinson, S. C. (2012). The original sin of cognitive science. *Topics in Cognitive Science*, *4*(3), 396–403. doi: 10.1111/j.1756-8765.2012.01195.x

Levinson, S. C., & Evans, N. (2010). Time for a sea-change in linguistics: Response to comments on 'The Myth of Language Universals'. *Lingua*, *120*(12), 2733–2758. doi: 10.1016/j.lingua.2010.08.001

Lindsley, J. R. (1975). Producing simple utterances: How far ahead do we plan? *Cognitive Psychology*, *7*(1), 1–19. doi: 10.1016/0010-0285(75)90002-X

Lockwood, H. T., & Macaulay, M. (2012). Prominence hierarchies. *Language and Linguistics Compass*, *6*(7), 431–446. doi: 10.1002/lnc3.345

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*(226). doi: 10.3389/fpsyg.2013.00226

Maclachlan, A. (1996). *Aspects of ergativity in Tagalog* (Unpublished doctoral dissertation). McGill University, Montreal.

Magyari, L., Bastiaansen, M., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530–2539. doi: 10.1162/jocn_a_00673

Magyari, L., & de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, *3*(376). doi: 10.3389/fpsyg.2012.00376

Marslen-Wilson, W. D. (1975). Sentence Perception as an Interactive Parallel Process. *Science*, *189*(4198), 226–228. doi: 10.1126/science.189.4198.226

Mathôt, S., van der Linden, L., Grainger, J., & Vitu, F. (2013). The pupillary light response reveals the focus of covert visual attention. *PLoS ONE*, *8*(10), e78168. doi: 10.1371/journal.pone

.0078168

Maurits, L., & Griffiths, T. L. (2014). Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(37), 13576–13581. doi: 10.1073/pnas.1319042111

McDonald, J. L., Bock, K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, *25*(2), 188–230. doi: 10.1006/cogp.1993.1005

McFarland, C. D. (1984). Tagalog causative verbs. *Philippine Journal of Linguistics*, *16*(2), 19–47.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton: Chapman & Hall/CRC.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. doi: 10.1016/j.jml.2007.11.006

Mishra, R. K., Singh, N., Pandey, A., & Huettig, F. (2012). Spoken language-mediated anticipatory eye-movements are modulated by reading ability – Evidence from Indian low and high literates. *Journal of Eye Movement Research*, *5*(1), 1–10.

Momma, S., Slevc, L. R., & Phillips, C. (2016). The timing of verb selection in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 813–824. doi: 10.1037/xlm0000195

Moulton, S. T., & Kosslyn, S. M. (2009). Imagining predictions: mental imagery as mental emulation. *Philosophical Transactions of The Royal Society B*, *364*(1521), 1273–1280. doi: 10.1098/rstb.2008.0314

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology*, *48*(11), 1532–1543. doi: 10.1111/j.1469-8986.2011.01226.x

Myachykov, A., Garrod, S., & Scheepers, C. (2009). Attention and syntax in sentence production: A critical review. *Discours: Revue de linguistique, psycholinguistique et informatique*, *4*. doi: 10.4000/discours.7594

Myachykov, A., Scheepers, C., Garrod, S., Thompson, D., & Fedorova, O. (2013). Syntactic flexibility and competition in sentence production: The case of English and Russian. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1601–1619. doi: 10.1080/17470218.2012.754910

Myachykov, A., Thompson, D., Garrod, S., & Scheepers, C. (2012). Referential and visual cues to structural choice in visually situated sentence production. *Frontiers in Psychology*, *2*(396). doi: 10.3389/fpsyg.2011.00396

Myachykov, A., Tomlin, R. S., & Posner, M. I. (2005). Attention and empirical studies of grammar. *The Linguistic Review*, *22*(2–4), 347–364. doi: 10.1515/tlir.2005.22.2-4.347

Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, *62*(1), 56–119.

Norcliffe, E., Harris, A. C., & Jaeger, T. F. (2015). Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Language, Cognition and Neuroscience*, *30*(9), 1009–1032. doi: 10.1080/23273798.2015.1080373

Norcliffe, E., & Konopka, A. E. (2015). Vision and language in cross-linguistic research on sentence production. In R. K. Mishra, N. Srinivasan, & F. Huettig (Eds.), *Attention and vision in language processing* (pp. 77–96). New Delhi: Springer. doi: 10.1007/978-81-322-2443-3_5

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*(9), 1187–1208.

doi: 10.1080/23273798.2015.1006238
Olsen, A. (2012). *Tobii I-VT Fixation Filter: Algorithm Description* (Tech. Rep.). Danderyd: Tobii Technology. Retrieved from http://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/tobii-pro-i-vt-fixation-filter.pdf/?v=2012
Papesh, M. H., & Goldinger, S. D. (2012). Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics*, *74*(4), 754–765. doi: 10.3758/s13414-011-0263-y
Paul, I., & Travis, L. (2006). Ergativity in Austronesian languages: What it can do, what it can't, but not why. In A. Johns, D. Massam, & J. Ndayiragije (Eds.), *Ergativity: Emerging issues* (Vol. 65, pp. 315–335). Dordrecht: Springer.
Payne, T. E. (1982). Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language*, *6*(1), 75–106.
Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, *11*(3), 105–110. doi: 10.1016/j.tics.2006.12.002
Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. doi: 10.1017/S0140525X12001495
Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effect models in S and S-PLUS*. New York: Springer.
Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–569. doi: 10.1111/j.1469-8986.2009.00947.x
Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*,

2(5), 509–522. doi: 10.1037/0278-7393.2.5.509

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*(2), 168–182. doi: 10.1006/jmla.1999.2668

Primus, B. (1999). *Cases and thematic roles: Ergative, accusative and active*. Tübingen: Max Niemeyer Verlag.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), 576–582. doi: 10.1038/nrn1706

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna. Retrieved from `http://www.R-project.org`

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. doi: 10.3758/BF03214555

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, *22*(2), 204–207. doi: 10.1111/j.1469-8986.1985.tb01587.x

Riesberg, S. (2014a). Passive actors are not adjuncts – consequences for the distinction between symmetrical and asymmetrical voice alternations. In I. W. Arka & N. L. K. M. Indrawati (Eds.), *Argument realisations and related constructions in austronesian languages: papers from 12-ICAL* (Vol. 2, pp. 281–302). Canberra: Asia-Pacific Linguistics.

Riesberg, S. (2014b). *Symmetrical voice and linking in Western Austronesian languages*. Boston / Berlin: De Gruyter Mouton.

Riesberg, S., & Primus, B. (2015). Agent prominence in symmetrical voice languages. *STUF – Language Typology and Universals*, *68*(4), 551–564. Manuscript. doi: 10.1515/stuf-2015-0023

Robertson, S. S., & Suci, G. J. (1980). Event perception by children in the early stages of language production. *Child Development*, *51*(1), 89–96.

Rohde, H., & Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition*, *133*(3), 667–691. doi: 10.1016/j.cognition.2014.08.012

Ross, M. (2002). The history and transitivity of western Austronesian voice and voice-marking. In F. Wouk & M. Ross (Eds.), *The history and typology of western Austronesian voice systems* (pp. 17–62). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function. Part I: Principles of functional organisation. *Current Neuropharmacology*, *6*(3), 235–253.

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*, 211–223. doi: 10.1038/nrn2573

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, *162*, 42–45. doi: 10.1016/j.bandl.2016.08.001

Sauppe, S. (2016). Verbal semantics drives early anticipatory eye movements during the comprehension of verb-initial sentences. *Frontiers in Psychology*, *7*(95). doi: 10.3389/fpsyg.2016.00095

Sauppe, S. (2017). *Word order and voice influence the timing of verb planning in German sentence production.* Manuscript under review.

Sauppe, S., Norcliffe, E. J., Konopka, A. E., Van Valin, R. D., & Levinson, S. C. (2013). Dependencies first: Eye tracking evidence from sentence production in Tagalog. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 1265–1270). Austin, TX.

Schachter, P. (1976). The subject in Philippine languages: Topic, actor,

actor-topic, or none of the above. In C. N. Li (Ed.), *Subject and topic* (pp. 491–518). New York: Academic Press.

Schachter, P. (1977). Reference-related and role-related properties of subjects. In P. Cole & J. M. Sadock (Eds.), *Grammatical relations* (pp. 279–306). New York: Academic Press.

Schachter, P. (1995a). *The subject in Tagalog: Still none of the above* (Vol. 15). Los Angeles: Department of Linguistics, University of California.

Schachter, P. (1995b). Tagalog. In J. Jacobs, A. von Stechow, W. Sternefeld, & T. Vennemann (Eds.), *Syntax: An International Handbook of Contemporary Research* (Vol. 9, pp. 1418–1430). Berlin / New York: Walter de Gruyter. doi: 10.1515/9783110142631.2

Schachter, P., & Otanes, F. T. (1972). *Tagalog reference grammar*. Berkeley: University of California Press.

Schluroff, M., Zimmermann, T. E., Freeman Jr, R. B., Hofmeister, K., Lorscheid, T., & Weber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and Language*, *27*(2), 322-344. doi: 10.1016/0093-934X(86)90023-4

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology*, *5*(137). doi: 10.3389/fpsyg.2014.00137

Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, *34*(4), 499–520.

Schriefers, H., Teruel, E., & Meinshausen, R. (1998). Producing simple sentences: Results from picture–word interference experiments. *Journal of Memory and Language*, *39*(4), 609–632. doi: 10.1006/jmla.1998.2578

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. doi: 10.1111/j.1756-8765.2009.01024.x

Shannon, C. E. (1948). A mathematical theory of communication.

The Bell System Technical Journal, *27*(3), 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Signal Developers. (2013). signal: Signal processing [Computer software manual]. Retrieved from `http://r-forge.r-project.org/projects/signal/` (R package version 0.7-4)

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. doi: 10.1002/wcs.1323

Sison, C. P., & Glaz, J. (1995). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, *90*(429), 366–369. doi: 10.2307/2291162

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. doi: 10.1111/j.1467-7687.2007.00569.x

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*(8), 311–327. doi: 10.1111/lnc3.12151

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: evidence from eye-movements. *Language and Cognition*, *6*(1), 12–44. doi: 10.1017/langcog.2013.5

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*, 318–330. doi: 10.1016/j.jml.2011.04.009

Tanangkingsing, M., & Huang, S. (2007). Cebuano passives revisited. *Oceanic Linguistics*, *46*(2), 554–584.

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of The Royal Society B*, *363*(1493), 1105–1122. doi: 10.1098/rstb.2007.2162

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557–580. doi: 10.1023/A:1026464108329

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. doi: 10.1126/science.7777863

Tomlin, R. S. (1995). Focal attention, voice, and word order: An experimental, cross-linguistic study. In P. A. Downing & M. Noonan (Eds.), *Word order in discourse* (Vol. 30, pp. 517–554). Amsterdam / Philadelphia: John Benjamins.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90. doi: 10.1006/jmla.2001.2836

van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124–144. doi: 10.1016/j.jml.2013.11.001

van Nice, K. Y., & Dietrich, R. (2003). Task sensitivity of animacy effects: evidence from German picture descriptions. *Linguistics*, *41*(5), 825–849. doi: 10.1515/ling.2003.027

Verney, S. P., Granholm, E., & Dionisio, D. P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology*, *38*(1), 76–83. doi: 10.1111/1469-8986.3810076

Villacorta, P. J. (2012). Multinomialci: Simultaneous confidence intervals for multinomial proportions according to the method by Sison and Glaz [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=MultinomialCI` (R package version 1.0)

Wagers, M., Borja, M. F., & Chung, S. (2015). The real-time comprehension of WH-dependencies in a WH-agreement language. *Language*, *91*(1), 109–144. doi: 10.1353/lan.2015.0001

Wang, L., Schlesewsky, M., Bickel, B., & Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the 'subject'-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes*, *24*(7), 1180–1226. doi: 10.1080/01690960802159937

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, *34*(3), 311–334. doi: 10.1006/jmla.1995.1014

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. Dordrecht / Heidelberg / London / New York: Springer. doi: 10.1007/978-0-387-98141-3

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8456–8460. doi: 10.1073/pnas.120185810910.1073/pnas.120185810910.1073/pnas.1201858109

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of The Royal Society B*, *358*(1431), 593–602. doi: 10.1098/rstb.2002.1238

Wouk, F., & Ross, M. (Eds.). (2002). *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Wundt, W. (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythus und Sitte. Erster Band: Die Sprache*. Leipzig: Wilhelm Engelmann.

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does *not* do). *Journal of Memory and Language*, *72*, 37–48. doi: 10.1016/j.jml.2013.12.003

Zeileis, A., & Grothendieck, G. (2005). zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, *14*(6), 1–27.

Zekveld, A. A., & Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology*, *51*(3), 277–284. doi: 10.1111/psyp.12151

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear & Hearing*, *31*(4), 480–490. doi: 10.1097/AUD.0b013e3181d4f251

Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, *81*(3), 133–141.

Zylberberg, A., Oliva, M., & Sigman, M. (2012). Pupil dilation: a fingerprint of temporal selection during the "Attentional Blink". *Frontiers in Psychology*, *3*(316). doi: 10.3389/fpsyg.2012.00316

Part V

SYNOPSES

SYNOPSIS IN ENGLISH

Both the production and the comprehension of sentences are fundamentally incremental, meaning they are processed in steps and piece-by-piece rather than as a whole. Previous research has shown that we do not first prepare all the details of a sentence down to the sound structure of each word when we speak but rather may start speaking as soon as we do have enough information planned. The comprehension of sentences has also been shown to be an incremental process. When hearing a sentence, we immediately start to form an interpretation and try to merge the words that we already heard into a meaningful whole instead of waiting until the sentence we hear is finished. Moreover, we often even go further and actively anticipate the upcoming words.

The studies in this thesis are concerned with incremental sentence comprehension and sentence production and how the grammars of languages affect these incremental processes. With respect to sentence production, they explore what is "enough information" that needs to planned before speaking can start and whether this depends on the word order of a language or the way it marks dependencies between verbs and their arguments. With respect to sentence comprehension, they explore what kind of information is predicted when anticipating how a sentence may continue.

The two languages that were investigated are German and Tagalog (an Austronesian language of the Philippines). Tagalog is especially interesting for psycholinguistics because its grammatical properties allow us to test hypotheses about language production and comprehension that cannot be tested with the commonly investigated

European languages. These grammatical properties are its verb-initial word order and the language's "Philippine-type" voice system.

Chapter 2 investigates whether the verb-initial word order of Tagalog has consequences for how speakers plan sentences. It is argued that the approach of linear incrementality, where formulation proceeds strictly word-by-word, is not applicable to this language because sentence formulation is influenced by the need to express the relationship between event participants (the "who does what to whom") early in the sentence with a verb as the first word. In two experiments, eye movements and grammatical choices during picture descriptions were analyzed. Tagalog speakers started the planning process by encoding the event and by generating a grammatical structure of their utterance, as it was predicted by the account of hierarchically incremental sentence production.

Chapter 3 turns to German and explores differences in the planning of sentences with different word order. In German, the verb may either occur sentence-medially after the subject or in sentence-final position. In a picture description experiment, it was found that verbs are only prepared shortly before they are uttered, i.e. early in the formulation process when the verb is in medial position and late when it is in final position. In addition, it was found that German speakers base their grammatical choices on the humanness of agents (the "doers" of events) and patients (the "done-tos"). These findings suggest that speakers always start planning with analyzing the to-be-described event to form a representation of the utterance; however, individual words are only planned later.

Chapter 4 deals with the development of cognitive processing load over the time course of sentence production in Tagalog and German. The two languages differ in how the verb and its arguments are grammatically related to each other, i.e. in their voice marking systems. In unmarked active voice sentences in German, the verb takes the agent as the subject. Speakers must use passives with overt voice marking morphology on the verb if the patient is to be the subject. In Tagalog, by contrast, the verb is morphologically marked in either case. This kind of symmetrical voice system is only found in Austronesian languages and is therefore also called "Philippine-type" voice system. Tagalog and German speakers' processing load was investigated using pupillometry, measuring changes in the size of their pupil as they described pictures with sentences with different voice markings. It was found that actives and passives in German and agent voice and patient voice sentences in Tagalog exhibited different dynamics of pupil size changes over time. This suggests that the different grammatical architectures of the Tagalog and German voice systems actually have real-time consequences for language processing.

Finally, chapter 5 presents a study on Tagalog investigating what kind of information is anticipated during sentence comprehension. The verb is the first word in a Tagalog sentence and provides the listener with information about what is happening in the sentence (for example, whether the event that is described is a "kicking" event or a "sleeping" event) as well as about what order of agent and patient will follow the verb. In a visual-world eye tracking experiment, participants saw different objects on a computer screen while hearing a sentence about these objects in order to test what kind of information from the verb they would exploit for eye movements that anticipate the linguistic input. Analyses of eye movements after the sentence-initial verb was heard reveal that listeners always looked at the agent first, i.e. they exploited the meaning of the verb in order to predict which object on the display is carrying out the described action. Only later, they also predicted word order by looking towards the second-mentioned object while the first object's name was still being uttered.

In sum, the research presented in this thesis shows that the grammatical properties of languages influence how speakers plan sentences and how comprehenders parse the linguistic input. Different word orders and different voice marking systems lead to observable differences in how sentences are processed in Tagalog and German. Taken together, the results from chapters 2–5 also highlight the cross-linguistic importance of event representations for both formulating and comprehending sentences.

NEDERLANDSE SYNOPSIS

Zowel de productie als het begrijpen van zinnen zijn fundamenteel incrementeel: dit betekent dat de zinnen in stappen en stuk voor stuk worden verwerkt in plaats van als een geheel. Eerder onderzoek heeft aangetoond dat we, wanneer we spreken, niet als eerste alle details van een zin tot op de geluidsstructuur van elk woord voorbereiden, maar beginnen te spreken zodra we genoeg informatie gepland hebben. Het begrijpen van zinnen is tevens aangetoond als een incrementeel proces. Wanneer we een zin horen, beginnen we onmiddellijk een interpretatie te vormen en proberen de woorden, die we al gehoord hebben, te combineren tot een zinvol geheel, in plaats van te wachten tot de zin die we horen is afgerond. Bovendien, we gaan vaak nog verder en anticiperen de woorden die nog gaan volgen.

De studies in dit proefschrift hebben betrekking op incrementeel zinsbegrip en zinsproductie en hoe de grammatica van een taal invloed heeft op deze incrementele processen. Het zinsproductie-onderzoek gaat na wat "genoeg informatie" is dat gepland moet worden voordat men kan gaan beginnen met spreken en of dit afhankelijk is van de woordvolgorde van een taal, of de manier waarop een taal duidelijk maakt hoe werkwoorden en hun argumenten afhankelijk zijn van elkaar. Het zinsbegrip-onderzoek bestudeert wat voor informatie voorspeld wordt wanneer een luisteraar anticipeert hoe de zin misschien wordt vervolgd.

De twee talen die hier worden onderzocht zijn Duits en Tagalog (een Austronesische taal uit de Filipijnen). Tagalog is vooral interessant voor de psycholinguïstiek omdat de grammaticale eigenschappen van deze taal ons in staat stellen hypotheses over taalproductie en

-begrip te testen die niet getoetst kunnen worden op de algemeen onderzochte Europese talen. Deze grammaticale eigenschappen zijn de werkwoord-eerst woordvolgorde en het "Filippijns-typische" *voice*-systeem van de taal.

Hoofdstuk 2 onderzoekt of de werkwoord-eerst woordvolgorde in het Tagalog gevolgen heeft voor de manier waarop sprekers hun zinnen plannen. Men beargumenteert dat de aanpak van lineaire incrementaliteit, waarbij formulering strikt woord voor woord gebeurt, niet toepasbaar is op deze taal omdat de formulering van zinnen beïnvloed wordt door de noodzaak om de relatie tussen de deelnemers van de gebeurtenissen (de "wie doet wat aan wie") vroeg in de zin met een werkwoord als eerste woord uit te drukken. Door middel van twee experimenten zijn de oogbewegingen en grammaticale keuzes geanalyseerd tijdens de omschrijving van afbeeldingen. Tagalog sprekers begonnen het planningsproces door de gebeurtenis te coderen en door een grammaticale structuur van hun uitspraak te genereren, zoals voorspeld is door hiërarchisch incrementele zinsproductie.

Hoofdstuk 3 betreft het Duits en onderzoekt de verschillen in het plannen van zinnen met verschillende woordvolgordes. In het Duits kan het werkwoord of mediaal in de zin na het onderwerp staan of de zin afsluiten. In het experiment waarbij afbeeldingen omschreven werden, werd gevonden dat werkwoorden alleen worden voorbereid kort voordat deze gesproken worden, dat wil zeggen vroeg in het formuleringsproces wanneer het werkwoord zich in een mediale positie bevindt en laat wanneer het in eindpositie staat. Daarnaast bleek dat Duitse sprekers hun grammaticale keuzes baseren op de menselijkheid van de agenten (de "doeners" van gebeurtenissen) en patiënten (degenen die het "ondergaan"). Deze bevindingen geven aan dat sprekers altijd het plannen beginnen met het analyseren van de te omschrijven gebeurtenis om een weergave van de uitspraak te vormen; echter, individuele woorden worden later gepland.

Hoofdstuk 4 behandelt de ontwikkeling van de cognitieve belasting tijdens de productie van een zin in Tagalog of Duits. De twee talen verschillen in hoe het werkwoord en zijn argumenten grammaticaal met elkaar verbonden zijn, dat wil zeggen in hun *voice*-markeringssystemen. In ongemarkeerde actieve zinnen in het Duits neemt het werkwoord de agent als onderwerp. Sprekers moeten passieve werkwoorden met een openlijke morfologische markering gebruiken als de patiënt het onderwerp is. In Tagalog, daarentegen, is het werkwoord morfologisch in beide gevallen gemarkeerd. Dit soort symmetrisch *voice*-systeem is alleen te vinden in sommige Austronesische talen en wordt daarom tevens het "Filippijns-typische" *voice*-systeem genoemd. De verwerkingslast van Tagalog en Duitse sprekers was onderzocht door middel van pupillometrie, het meten van veranderingen in de grootte van hun pupil terwijl ze afbeeldingen in zinnen omschreven met verschillende *voice*-markeringen. Er werd gevonden dat actieve en passieve zinnen in het Duits en *agent-voice* en *patiënt-voice* zinnen in Tagalog in de loop van de tijd verschillende veranderingen in pupilgroottes vertoonden. Dit wijst erop dat de verschillende *voice*-systemen in Tagalog en Duits daadwerkelijk gevolgen hebben voor taalverwerking.

Ten slotte, hoofdstuk 5 presenteert een studie van het Tagalog waarin onderzocht wordt wat voor soort informatie geanticipeerd wordt tijdens het begrijpen van de zin. Het werkwoord is het eerste woord in een Tagalog zin en geeft de luisteraar informatie over zowel wat er gebeurt in de zin (bijvoorbeeld, of de omschreven gebeurtenis een schopbeweging of een slaaptoestand betreft) als wat de volgorde van agent of patiënt is na het werkwoord. In een visueel-wereld-eye-tracking-experiment, zagen participanten verschillende objecten op een computerscherm terwijl zij een zin over deze objecten hoorden om te testen wat voor soort informatie van het werkwoord ze zouden gebruiken voor de oogbewegingen, die de volgende woorden anticiperen. Analyses van oogbewegingen onthulden, nadat de zin met het werkwoord op de eerste positie was gehoord, dat luisteraars altijd eerst naar de agent keken; dat wil zeggen dat zij de betekenis van het werkwoord gebruikten om te voorspellen welk object op het scherm de omschreven actie uitvoerde. Pas later voorspelden zij de woordvolgorde door te kijken naar het tweede genoemde object, terwijl de naam van het eerste object nog uitsproken werd.

Kortom, het onderzoek dat gepresenteerd wordt in dit proefschrift toont aan dat de grammaticale eigenschappen van talen invloed hebben op hoe sprekers een zin plannen en hoe luisteraars de woorden, die zij horen, begrijpen. Verschillende woordvolgordes en verschillende *voice*-markeringssystemen leiden tot waarneembare verschillen in hoe zinnen in Tagalog en Duits worden verwerkt. Dit in acht genomen, benadrukken de resultaten van hoofdstukken 2–5 tevens het belang van gebeurtenisweergaves in verschillende talen voor zowel de formulering als het begrijpen van zinnen.

CURRICULUM VITAE

Sebastian Sauppe graduated as Magister Artium in General Linguistics and Sociology from the University of Leipzig, Germany, in 2010. In 2011, he began working on his PhD project at Language and Cognition department of the Max Planck Institute for Psycholinguistics in Nijmegen, Netherlands. The results of his PhD research are detailed in this thesis. Currently, he is a postdoctoral researcher at the Department of Comparative Linguistics of the University of Zurich, Switzerland.

MPI SERIES IN PSYCHOLINGUISTICS

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*

17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. Petra M. van Alphen
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*
27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*
35. Retrieving semantic and syntactic word properties. Oliver Müller
36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*
37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*
41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*
42. The acquisition of auditory categories. *Martijn Goudbeek*
43. Affix reduction in spoken Dutch. *Mark Pluymaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*
45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*
63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*
64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*
66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*
67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*
69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*
70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*
71. Gaze behavior in face-to-face interaction. *Federico Rossano*
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*
74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*
75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*

76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*
77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*
78. Contributions of executive control to individual differences in word production. *Zeshu Shao*
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*
80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*
81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*
82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*
83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*
84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*
85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*
86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*
87. The acquisition of morphophonological alternations across languages. *Helen Buckler*
88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*
89. Rediscovering a forgotten language. *Jiyoun Choi*
90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*
91. Infants'understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*
92. Information structure in Avatime. *Saskia van Putten*
93. Switch reference in Whitesands. *Jeremy Hammond*
94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*

95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sümer*
96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*
97. Incrementality and Flexibility in Sentence Production. *Maartje van de Velde*
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*
99. The request system in Italian interaction. *Giovanni Rossi*
100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*
101. Assessing birth language memory in young adoptees. *Wencui Zhou*
102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*
103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*
104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition. *Rósa Signý Gisladottir*
105. Modelling Multimodal Language Processing. *Alastair Smith*
106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*
107. Situational variation in non-native communication. *Huib Kouwenhoven*
108. Sustained attention in language production. *Suzanne Jongman*
109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntatic, and orthographic effects. *Malte Viebahn*
110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*
111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*
112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*
113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*
114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*
115. Music and language comprehension in the brain. *Richard Kunert*

116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*
117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*
118. Language processing in a conversation context. *Lotte Schoot*
119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*
120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*
121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*
122. Sensorimotor experience in speech perception. *Will Schuerman*
123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*
124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*
125. Lapse organization in interaction. *Elliott Hoey*
126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*
127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*
128. The role of voice and word order in incremental sentence processing. Studies on sentence production and comprehension in Tagalog and German. *Sebastian Sauppe*