Dependencies First: Eye Tracking Evidence from Sentence Production in Tagalog

Sebastian Sauppe (sebastian.sauppe@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, Netherlands;
International Max Planck Research School for Language Sciences, Wundtlaan 1, 6525XD Nijmegen, Netherlands

Elisabeth Norcliffe (elisabeth.norcliffe@mpi.nl)
Agnieszka E. Konopka (agnieszka.konopka@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, Netherlands

Robert D. Van Valin, Jr (vanvalin@phil-fak.uni-duesseldorf.de)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, Netherlands;
Heinrich Heine University Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, Germany;
University at Buffalo, The State University of New York, 609 Baldy Hall, Buffalo, NY 14260, USA

Stephen C. Levinson (stephen.levinson@mpi.nl)
Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, Netherlands;
Radboud University Nijmegen, Erasmusplein 1, 6525HT Nijmegen, Netherlands

Abstract

We investigated the time course of sentence formulation in Tagalog, a verb-initial language in which the verb obligatorily agrees with one of its arguments. Eye-tracked participants described pictures of transitive events. Fixations to the two characters in the events were compared across sentences differing in agreement marking and post-verbal word order. Fixation patterns show evidence for two temporally dissociated phases in Tagalog sentence production. The first, driven by verb agreement, involves early linking of concepts to syntactic functions; the second, driven by word order, involves incremental lexical encoding of these concepts. These results suggest that even the earliest stages of sentence formulation may be guided by a language's grammatical structure.

Keywords: eye tracking; sentence production; incrementality; Austronesian; verb-initial word order.

Introduction

In the process of transforming thoughts into speech, speakers begin with a preverbal message, which must then be encoded linguistically. In English, this process may proceed in a highly lexically incremental manner: for example, when describing events like the one shown in Figure 1, speakers may have encoded as little as the first element (the syntactic subject, e.g., "the boy") of the to-be-uttered sentence prior to speech onset (Gleitman, January, Nappa, & Trueswell, 2007). The encoding of additional event participants (e.g., "the ball") and the relation between them (e.g., "kicking") may be delayed until after speakers finish encoding the first element. This type of incremental planning is compatible with English morphosyntax, arguably in part because full noun phrases do not morphologically mark dependencies with other elements in the sentence. For many sentence types, speakers therefore do not have to commit to a particular syntactic structure upon beginning to encode one of the event participants as the syntactic subject. However, not all languages offer this flexibility: in some languages the first word is overtly marked for a dependency with word(s) occurring only later in the sentence. In such cases, is there an effect of dependency marking on early sentence encoding as speakers begin to map the preverbal message onto linguistic structure?

One such language that exhibits dependency marking on the first word of a sentence is the Austronesian language Tagalog. The predicate is in sentence-initial position and agrees with one of its arguments. Thus, the grammatical properties of Tagalog allow us to test whether and how linguistic structure influences the earliest phases of sentence production; specifically, we test whether the overt dependency marking on the first word in a sentence leads to differences in the time course of sentence formulation in Tagalog compared to languages with no overt dependency marking on the first word (such as English).

In the following, we first sketch the relevant grammatical properties of Tagalog and then report the results of a picture description experiment in which eye-tracked speakers described pictures of simple transitive events.

Tagalog

Tagalog is spoken by approx. 21.5 million speakers in the Philippines; it belongs to the Western Malayo-Polynesian branch of the Austronesian language family. We provide a brief overview of the morphosyntactic properties that are relevant for the reported experiment. For more comprehensive descriptions of Tagalog morphosyntax, see Himmelmann (2005), Kroeger (1993), and Schachter and Otanes (1972).

Basic declarative Tagalog sentences are predicate[1]-initial, i.e., predicates are followed by their arguments. One argu-

[1] To circumvent the discussion on lexical categories (noun/verb distinction) in Tagalog (e.g., Himmelmann, 2008), we will use the term "predicate" throughout this paper to refer to voice-marked words and the term "argument" to refer to heads of case-marked (non-oblique) phrases.

ment phrase in each sentence hosts the case marker *ang*. The semantic relation between the *ang*-marked argument and the predicate is signaled by affixes on the predicate.

The *ang*-marked argument will henceforth be referred to as the privileged syntactic argument (PSA). It is morphosyntactically prominent in being the only argument with which the predicate agrees in semantic role (see sentences (1–4)) and also in being the target of many syntactic operations (e.g., Kroeger, 1993)[2]. Arguments marked by *ng* do not exhibit these properties and are therefore referred to as non-privileged syntactic arguments (NPSA).

(1)	*s<um>isipa*	*ng=bola*	*ang=bata*
	<AV>kick[3,4]	NPSA=ball	PSA=child
	predicate	undergoer	actor
	"The child kicks the ball."		
(2)	*s<um>isipa*	*ang=bata*	*ng=bola*
	<AV>kick	PSA=child	NPSA=ball
	predicate	actor	undergoer
	"The child kicks the ball."		
(3)	*s<in>ispa*	*ng=bata*	*ang=bola*
	<UV>kick	NPSA=child	PSA=ball
	predicate	actor	undergoer
	"The child kicks the ball."		
(4)	*s<in>isipa*	*ang=bola*	*ng=bata*
	<UV>kick	PSA=ball	NPSA=child
	predicate	undergoer	actor
	"The child kicks the ball."		

The sentences in (1–4) illustrate three properties of Tagalog grammar that are relevant for this study. First, the predicate always agrees in semantic role with the PSA in basic sentences. In sentences (1) and (2), the PSA denotes the actor[5] of the event so the predicate takes actor voice marking (AV); in sentences (3) and (4) the PSA denotes the undergoer of the event so the predicate takes undergoer voice marking (UV)[6]. NPSAs marked by the case marker *ng* do not agree with the predicate. Second, there are no syntactic constraints on the ordering of arguments for the constructions dealt with in this paper. Sentences (1) and (2) have the same meaning but they differ in their word order: in (1) the PSA is sentence-final, whereas it is sentence-medial in (2); the same holds for (3) and (4), respectively. However, canonically, the PSA is in sentence-final position (as in sentences (1) and (3)). Third, the Tagalog voice system is a so-called "symmetrical voice system" (Foley, 2008): sentences in which the undergoer is selected as PSA and sentences with an actor PSA are equally transitive. This contrasts with languages with asymmetrical voice systems such as English in which valency-changing operations, such as passivization which detranstivizes the verb, are required to allow the patient/undergoer argument to be the PSA (syntactic subject in English). Detransitivization is the key part of the asymmetrical voice system in contrast to Tagalog. Thus, all Tagalog sentences analyzed in this paper are transitive (exhibiting one PSA and one NPSA phrase), regardless of the semantic role of the PSA.

How Do Speakers Plan Sentences in Tagalog?

The sentence-initial position of the predicate in a Tagalog sentence means that speakers must encode enough information about the relationship between the two discourse entities ("boy" and "ball") to select a suitable predicate ("kick") very early in the formulation process. The predicate's agreement in semantic role with the PSA also means that very early in the formulation process one discourse entity from the preverbal message has to be selected to be the PSA and linked to that syntactic function so that appropriate voice marking for the predicate can be selected. Importantly, speakers can produce the PSA immediately after the predicate (as in sentences (2) and (4)) or may delay its production until the end of the sentence (as in sentences (1) and (3)).

To what extent do speakers then have to encode the PSA at the outset of formulation in sentences like (1–4)? We tested whether the processing of the overt dependency between the predicate and the PSA is temporally separate from lexical encoding of the character selected to be the PSA by comparing the time course of formulation for sentences differing in voice and word order.

Native speakers of Tagalog performed a picture description experiment similar to Griffin and Bock (2000) while their eye movements and speech were recorded. The pictures showed events with one actor and one undergoer (Figure 1). We compared the distribution of fixations to the actor and the undergoer in these pictures for different sentence

[2] In English the syntactic subject is the PSA: it triggers agreement with the verb and it is the target of many syntactic operations (to the exclusion of the syntactic object). In Tagalog, however, we refrain from using the term "subject" for the *ang*-marked argument phrase in order to underscore the fact that the Tagalog PSA is different from the syntactic subject in an accusatively aligned language (such as English).

[3] In this paper, we adhere to the Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php); the following abbreviations are used: AV = actor voice, NPSA = non-privileged syntactic argument, PSA = privileged syntactic argument, UV = undergoer voice. The first line of a glossed example shows the sentence in Tagalog with the relevant morphemes separated, the second line provides a word-by-word translation of the words and morphemes, the third line shows the word order of the sentence again in terms of semantic roles, the last line gives an English translation.

[4] For the sake of brevity, we waive glossing aspect-mood morphology because it is irrelevant for the morphosyntactic issues discussed in this paper.

[5] We use Foley and Van Valin's (1984) notions of "actor" and "undergoer" to refer to semantic relations between predicates and arguments.

[6] More precisely, the predicate in (3) and (4) takes patient voice marking because the PSA denotes the patient of the action. Predicates may also take series of other voices. Following Himmelmann (2005), we subsume patient voice and these other voices under the label "undergoer voice" because they share a couple of semantic and formal characteristics in contrast to actor voice.

Figure 1: example target stimulus picture (eliciting the example sentences (1–4))

types. First, we compared sentences differing in voice marking and word order, i.e., the actor voice sentence type in (1) and the undergoer voice sentence type in (3), to investigate whether the semantic role of the PSA has an early influence on sentence planning. Second, comparisons between sentences with different voice marking but the same word order (such as in (2) vs. (3)) and sentences with the same voice marking but different word order (such as in (1) vs. (2)) were carried out to investigate whether a possible PSA effect on planning is solely due to the planning of the dependency between the predicate and the PSA (i.e., voice marking) or whether it is also influenced by word order.

Experiment

Method

Participants 53 native speakers of Tagalog (13 male; mean age = 17.5 years; range = 15–28 years) were recruited at De La Salle University in Manila and participated for payment. All participants reported that they spoke Tagalog a total of at least five hours a day and to at least one of their parents.

Materials and Design Target pictures were 44 cartoon drawings of transitive events (see Figure 1). They were 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. Pictures were then arranged in four lists created by randomizing the order of the target and filler pictures and counterbalancing the two mirror-reversed versions of each target picture.

Equipment The experiment was run with a Tobii T120 eye tracker (120 Hz sampling frequency) on a Panasonic CF-F9 computer. Participants' responses were recorded with a microphone.

Procedure Each experimental session lasted approx. 40 minutes. Participants first read instructions for the experiment in Tagalog and completed a questionnaire about their linguistic background. The experimenter (a native speaker of Tagalog) then explained the procedure and repeated the instructions: participants were asked to describe the events shown in the pictures with one sentence that named all event participants as accurately and as quickly as possible.

Stimuli were presented in two blocks, each block lasting approx. 10–15 minutes. Calibration was performed before each block. The experiment began with a practice phase in which participants saw 11 pictures presented one at a time and heard a recorded description of each depicted event; these example sentences had predicate-initial word order and were mostly PSA-final. After presentation of the example descriptions, participants saw the same pictures again and were asked to describe them themselves. The experimenter provided feedback after each training picture if participants produced non-predicate-initial structures (e.g., existential constructions) or started speaking very late after picture onset.

In the experimental phase, each picture trial was preceded by a display showing a fixation dot at the top of the screen. Participants were asked to look at the fixation dot and the experimenter initiated the trial with a mouse click. Participants completed the experiment without further instructions from the experimenter; however, the experimenter monitored the entire experimental session and repeated the instructions if participants started consistently using non-predicate-initial structures or dropping arguments.

Results

Picture descriptions Speakers produced 384 sentences with actor voice marking and predicate$_{AV}$-undergoer$_{NPSA}$-actor$_{PSA}$ word order (as in sentence (1)), 67 sentences with actor voice marking and predicate$_{AV}$-actor$_{PSA}$-undergoer$_{NPSA}$ word order (as in sentence (2)), 787 sentences with undergoer voice marking and predicate$_{UV}$-actor$_{NPSA}$-undergoer$_{PSA}$ word order (as in sentence (3)), and 26 sentences with undergoer voice marking and predicate$_{UV}$-undergoer$_{PSA}$-actor$_{NPSA}$ word order (as in sentence (4)). Analyses were limited to the first three sentence types.

First Fixations The majority of first fixations[7] (58.3%) across all trials fell on the actor in the event. Contrary to earlier work on English (e.g., Gleitman et al., 2007), first fixations did not predict choice of voice or word order (both z's<1.4, *n.s.*).

Time Course of Fixations Consecutive fixations to each character were aggregated into "runs" of fixations directed to those characters. The distributions of fixations directed to the actor and to the undergoer were then compared across the three most frequent sentence types in this dataset (i.e., the sentence types in (1–3)) with quasi-logistic regressions (Barr, 2008, for details about random effects). We selected three time windows for analysis (0–600 ms, 600–1600 ms, and 1600–2600 ms after picture onset). Selection of time windows was based on three theoretically important distinc-

[7] The Tobii Fixation Filter as implemented in Tobii Studio 2.3 was used to determine fixations.

a)
b)
c)

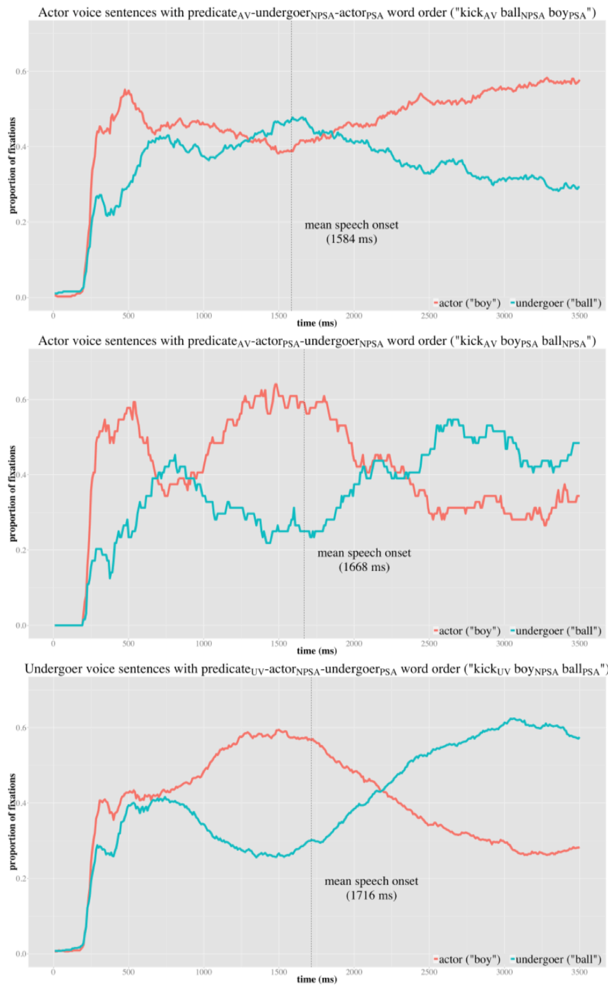

Figure 2: graphs showing fixation proportions to actor and undergoer characters over time for three sentence types

tions and to facilitate comparisons across the three sentence types, as explained below.

In all three sentence types included in the analysis (Figure 2), speakers directed more fixations to the character selected to be PSA than to the NPSA character in an early time window (0–600 ms). The 0–400 ms time window is argued to correspond to a period of event apprehension (Griffin & Bock, 2000), and here we extended this window to 600 ms based on the distribution of fixations in all three sentence types in Figure 2. Fixations to the two characters in this time window were aggregated into 50 ms time bins. Griffin and Bock (2000) propose that after the initial period of event apprehension, speakers begin fixating the two characters in the order of mention in order to retrieve their names; indeed, the distribution of fixations after 600 ms in this dataset largely shows that speakers fixated characters in the order of mention. Thus between 600 ms and speech onset (approx. 1600 ms after picture onset), speakers preferentially fixated the character that was mentioned immediately after the predicate (independently of syntactic function – i.e., whether it was the PSA or the NPSA – or semantic role). After speech onset, speakers then began shifting their gaze to the second character, and we compared the distribution of fixations to the two characters up to 1 second after speech onset (i.e., between 1600 and 2600 ms). Fixations were aggregated into 200 ms time bins for the analysis of the 600–1600 ms and 1600–2600 ms time windows.

Three analyses were performed to compare the distribution of actor-directed fixations in sentences differing in voice and word order in more detail. All analyses included time bin and sentence type as predictors. All models included random slopes for the two predictors. In the text, we report only the interactions between these factors from the by-participant analyses using the full random structure.[8] Effects were considered to be reliable at $p<0.05$ (most effects were also reliable according to the more conservative pMCMC estimates calculated for models without random slopes; in cases of discrepancy between p values calculated for models with random slopes and pMCMC values calculated for models without random slopes, we used the more conservative criterion to indicate significance and provide the corresponding pMCMC value).[9]

[8] By-item analyses showed analogous patterns; the full set of by-participant and by-item results is available from the first author on request.

[9] Linear mixed models were run in R using the lmer() function of the lme4 package. pMCMC values were calculated with the pvals.fnc() function of the languageR package. Figures were created using the ggplot2 package in R.

First, to test whether differences in voice marking and word order predict differences in early encoding of the PSA and NPSA characters, we compared actor voice sentences with predicate$_{AV}$-undergoer$_{NPSA}$-actor$_{PSA}$ word order (example (1), Figure 2a) and undergoer voice sentences with predicate$_{UV}$-actor$_{NPSA}$-undergoer$_{PSA}$ word order (example (3), Figure 2c). Between 0 and 600 ms, speakers fixated the actor character more often and more quickly when it was selected to be the PSA (i.e., in the actor voice sentences with predicate$_{AV}$-undergoer$_{NPSA}$-actor$_{PSA}$ word order, Figure 2a) than when it was not (Figure 2c; sentence type × time bin: β=-0.92, t=-7.16). After 600 ms, speakers fixated the characters in the order of mention. Between 600 and 1600 ms they fixated the actor character more often when the actor argument phrase immediately followed the predicate, i.e., in the undergoer voice sentences with predicate$_{UV}$-actor$_{NPSA}$-undergoer$_{PSA}$ word order (Figure 2c), then when it was sentence-final (Figure 2a; sentence type × time bin: β=1.14, t=49.48). They shifted their gaze to the second character between 1600 and 2600 ms, and thus fixated the actor character less often when the actor argument phrase was not sentence-final (Figure 2a; sentence type × time bin: β=-1.70, t=-73.93). These results suggest early encoding of the PSA character only for the purposes of selecting the appropriate agreement marking on the predicate; lexical encoding of the PSA character occurred either before or after encoding of the NPSA character, according to word order.

Second, we tested whether differences in voice marking alone can influence the time course of formulation by comparing sentences with the same word order, i.e., actor voice sentences with predicate$_{AV}$-actor$_{PSA}$-undergoer$_{NPSA}$ word order (example (2), Figure 2b) and undergoer voice sentences with predicate$_{UV}$-actor$_{NPSA}$-undergoer$_{PSA}$ word order (example (3), Figure 2c). Speakers fixated the actor character more often and more quickly between 0 and 600 ms when it was selected to be the PSA (Figure 2b) than when it was not selected to be the PSA, i.e., in the undergoer voice sentences (Figure 2c; sentence type × time bin: β=-1.76, t=-7.46). However, there were no differences between the actor voice and the undergoer voice sentences in the overall likelihood of speakers to fixate the actor character or direct fixations to it over time between 600 and 1600 ms (sentence type: β=0.13, t=4.40, pMCMC=0.65; sentence type × time bin: β=-0.03, t=-0.56, pMCMC=0.93) and between 1600 and 2600ms after picture onset (sentence type: β=-0.06, t=-1.64, pMCMC=0.84; sentence type × time bin: β=-0.18, t=-3.09, pMCMC=0.70) because speakers produced the actor character first and the undergoer character second in both sentence types. This confirms that differences in the time course of early sentence formulation reflect encoding of features of the PSA character relevant only for agreement marking.

Finally, we compared actor voice sentences with predicate$_{AV}$-undergoer$_{NPSA}$-actor$_{PSA}$ word order (example (1), Figure 2a) and actor voice sentences with predicate$_{AV}$-actor$_{PSA}$-undergoer$_{NPSA}$ word order (example (2), Figure 2b) to test whether word order influences fixations to picture characters when voice marking is kept constant. Comparing fixations to the actor character between 0 and 600 ms showed no reliable differences between the actor voice sentences with predicate$_{AV}$-undergoer$_{NPSA}$-actor$_{PSA}$ word order (Figure 2a) and the actor voice sentences with predicate$_{AV}$-actor$_{PSA}$-undergoer$_{NPSA}$ word order (Figure 2b) in this time window (sentence type: β=-0.36, t=-3.15, pMCMC=0.23; sentence type × time bin: β=0.88, t=3.00, pMCMC=0.16). The distribution of fixations to the actor and undergoer characters did, however, differ after 600 ms because the linear order of these characters in the two sentence types was different. Thus in the 600 and 1600 ms time window, speakers fixated the actor character more often if the actor argument phrase immediately followed the predicate (predicate$_{AV}$-actor$_{PSA}$-undergoer$_{NPSA}$ word order) then if it was sentence-final (sentence type × time bin: β=1.13, t=15.79), whereas speakers fixated the actor character less often between 1600 and 2600 ms if the actor argument phrase directly followed the predicate than if it was sentence-final (sentence type × time bin: β=-1.75, t=-18.39).

Speech onsets Speech onsets are shown in Figure 2 for each sentence type. Onsets were somewhat shorter in actor voice sentences with predicate$_{AV}$-undergoer$_{NPSA}$-actor$_{PSA}$ word order than in actor voice sentences with predicate$_{AV}$-actor$_{PSA}$-undergoer$_{NPSA}$ word order and undergoer voice sentences with predicate$_{UV}$-actor$_{NPSA}$-undergoer$_{PSA}$ word order (word order × voice: β=-0.19, t=-1.99, pMCMC=0.052). More importantly, we note a difference from results obtained with English speakers (Griffin & Bock, 2000): here, speech onsets occurred while speakers were still fixating the character that was mentioned first, suggesting that they had only fully encoded the predicate before initiating production, whereas English speakers begin their sentences only after encoding the first character. The sentence-initial position of the predicate may have allowed speakers to begin their sentences before completing the encoding of the first-mentioned character.

Discussion and Conclusion

We interpret the results of this experiment as evidence for linguistic guidance in the earliest stages of sentence production in Tagalog and for a temporal dissociation of the mapping of message-level concepts to syntactic functions and the lexical encoding of these concepts.

Linguistic guidance in early sentence production is suggested by differences in fixation patterns in the 0–600 ms time window across the three sentence types we analyze here: a depicted character was fixated more often if it was to become the sentence's PSA than when it was not. Specifically, speakers fixated the actor character more often than the undergoer character before 600 ms if the actor character was selected as the PSA, regardless of the position of the actor argument in the sentence (i.e., this pattern held for both, actor$_{PSA}$-medial or actor$_{PSA}$-final word orders). In contrast, there was no difference in early fixations directed to the actor character in actor voice sentences with different word orders (i.e., in actor$_{PSA}$-medial or actor$_{PSA}$-final sentences).

In other words, the results suggest that differences in voice marking (signaling differing semantic roles of the PSAs) but not differences in word order have an effect on fixation patterns in the earliest stage of sentence planning.

Early fixations of the PSA character suggest that the PSA effect is a reflex of linking message-level concepts of discourse entities to prominent syntactic functions. Speakers select a participant of the depicted event to be the PSA and encode its semantic role in order to produce an appropriate voice affix at the predicate. We propose that this process happens very early during formulation as speakers begin encoding information about the relationship between the two characters in the event.

Comparisons of the fixation patterns in the two later time windows (600–1600 ms and 1600–2600 ms) suggest that the PSA effect, i.e., the linking of a discourse entity concept to a prominent syntactic function, and the lexical encoding of the PSA are temporally dissociated. Whereas speakers are more likely to fixate the character selected to be the PSA before 600 ms, fixations to the two characters after 600 ms are contingent on word order. In the 600–1600 ms time window, the character that is to be mentioned immediately after the predicate is fixated more often by speakers than the character that is to be mentioned sentence-finally. Specifically, in actor voice sentences with predicate_{AV}-undergoer_{NPSA}-actor_{PSA} word order, speakers shift their gaze from the actor character (the PSA) to the undergoer character (the NPSA) after 600 ms, and similarly, in undergoer voice sentences with predicate_{UV}-actor_{NPSA}-undergoer_{PSA} word order, the speakers' gaze shifts from the undergoer character (the PSA) to the actor character (the NPSA) after 600 ms. Finally, in actor voice sentences with predicate_{AV}-actor_{PSA}-undergoer_{NPSA} word order speakers continue looking at the actor character (the PSA) because it is to be mentioned directly after the predicate. In the 1600–2600 ms window, speakers then fixate the character to be mentioned sentence-finally more often than the other character (i.e., the actor character in the first mentioned sentence type and the undergoer character in the two latter types). We interpret this as incremental encoding of the two character names in the order of mention that is distinct from the early phase of linking concepts to syntactic functions (0–600 ms).

The results suggest that there are two observable phases in the sentence production process in Tagalog: an early phase of sentence planning that includes the planning of the dependency relation between the predicate and the PSA (i.e., the voice marking), which is neither influenced by the actual semantic role of the PSA nor the word order of the to-be-uttered sentence, and a later phase that involves the incremental lexical encoding of the two arguments of the predicate.

Importantly, these analyses of the time course of sentence formulation in Tagalog provide insight into a process that is not easily observable in a language like English, namely the linking of conceptual discourse entities to prominent syntactic functions. The rigid subject-initial word order of English prevents dissociating the linking of concepts to syntactic functions from planning and encoding of the subject argument; thus our results on Tagalog highlight the need for controlled studies on typologically diverse languages that allow dissociations between different processes at the interface of thinking and speaking.

Ultimately, more fine-grained models of early message and sentence formulation are needed to address the relationship between formulation of a preverbal message and the mapping of this message onto language, and it is important for the development of such models to consider languages with grammatical properties that support investigations of these phenomena (Jaeger & Norcliffe, 2009).

Acknowledgments

This research was supported by a PhD fellowship from the International Max Planck Research School for Language Sciences to the first author and by the Language & Cognition Department of the MPI for Psycholinguistics, Nijmegen. We thank Danilo Dayag for administrative help, Philip Rentillo and Erwin Roxas for help with conducting the experiment and transcriptions, and Ronald Fischer for technical assistance.

References

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.

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

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

Gleitman, L. R., January, D., Nappa, R., and Trueswell, J. C. (2007). On the *give* and *take* between event apprehension and utterance formulation. *Journal of Memory and Language*, 57, 544–596.

Griffin, Z. M. and Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279.

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

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

Jaeger, T. F. and Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, 3, 866–887.

Kroeger, P. (1993*). Phrase structure and grammatical relations in Tagalog.* Palo Alto: CSLI Publications.

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