Symmetrical and asymmetrical voice systems and processing load: Pupillometric evidence from sentence production in Tagalog and German

Sebastian Sauppe

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

Sebastian Sauppe

Max Planck Institute for Psycholinguistics,
Ruhr University Bochum, and *Heinrich Heine University Düsseldorf*

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 article investigates the psycholinguistic processing consequences of the potential (a)symmetries in the voice systems of Tagalog and German 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 (relative) 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: linking patients to the subject function seems to be associated with greater cognitive effort. This anti-Patient bias in production adds converging evidence to 'subject preferences' reported in the sentence-comprehension literature.*

Keywords: Tagalog, symmetrical voice system, diatheses, sentence production, pupillometry, task-evoked pupillary responses

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 article I 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 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 type of voice system; 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 pro-

duction of sentences are used to investigate whether the grammatical distinction between asymmetrical and symmetrical voice systems may also be reflected in psycholinguistic processes.

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 nondefault voice was chosen. Foley (2008) 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). I briefly discuss each of these systems in turn.

German as an example of asymmetrical voice systems. In nominative-accusative voice systems as found in, for example, 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 the syntactic object and is assigned accusative case, as in the German example in 1.[1] The mapping between semantic roles and syntactic functions is different in passives, where the patient argument is mapped to the subject function.

(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 gefangen.
       the.NOM fish   AUX by   the.DAT man   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, the verb is detransitivized so that the patient can be the syntactic subject. Additional marking is also often required to indicate this nondefault mapping between semantic roles and syntactic functions. In German this is achieved by placing an auxiliary in second 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, for example, ergative-absolutive systems—are thus formally marked for their syntactic (in)transitivity, that is, the number of direct core arguments. It is this formal marking that is at the center of interest in the current article.

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

---

[1] The following abbreviations are used in glosses: ACC: accusative, AUX: auxiliary, AV: agent voice, BV: benefactive voice, CORE: core nonpivot argument, DAT: dative, IPFV: imperfective aspect, IRR: irrealis mood, IV: instrumental voice, LV: locative voice, NOM: nominative, OBL: oblique, PFV: perfective aspect, PTCP: participle, PV: patient voice, PVT: pivot argument, RLS: realis mood, SG: singular.

of its voice system has been debated for over a century (e.g. Blake 1906, Schachter 1976, 1995b, Kroeger 1993a,b), and it has been variously categorized as a nominative-accusative system (Bloomfield 1917), a nominative-accusative-like system (Kroeger 1993b), an ergative-absolutive system (Payne 1982, Aldridge 2012), 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 (see Kroeger 1993b and Schachter 1995b for further constructions that privilege *ang*-marked arguments).

Following Foley (2008), I henceforth refer to the *ang*-marked argument as the PIVOT ARGUMENT in this article (see 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 1995a, respectively). Other arguments that 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; see Kroeger 1993b). Obliques are marked by *sa* (often referred to as dative arguments in the literature).

The sentences in example 2 illustrate how the semantic role of the pivot is reflected on the predicate (pivot arguments and voice markers are in boldface).

(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   **ang=pamingwit**
      IV-IRR~catch   CORE=man CORE=fish 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 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 2b, by contrast, the pivot is the patient of the event so the predicate takes a different voice affix (-*in*). Example sentences 2c–e illustrate locative, instrumental, and benefactive pivots, respectively. The sentences in 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' (Him-

---

[2] It is a matter of debate whether Tagalog exhibits a noun/verb distinction (Himmelmann 2008). To circumvent this discussion, words carrying affixes that cross-reference the semantic role of one of the arguments in the clause are simply 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, these are not dealt with in this article.

melmann 2002:12). In the following, I focus on agent voice (2a) and patient voice (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 2014b). 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 [… nonargument] can appear in initial position in this construction' (Kroeger 1993b:47).[4]

In brief, there are no formally unmarked voice forms in Tagalog because mappings of both agents and patients to the pivot function are marked by voice affixes and detransitivization 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, that is, 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,b). Foley classifies the Tagalog voice system as symmetrical (and the voice systems of other Austronesian languages have also been described as being symmetrical to varying degrees; cf. Cole et al. 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' (Foley 2008:42). Symmetrical voice systems are also characterized by having more than one basic transitive construction and by pivot and core arguments behaving the same 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 Riesberg (2014b) also 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, by contrast, 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 detransitivization 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 detransitivization takes place are symmetrical. Do these differences in the architectures of asymmetrical and symmetrical voice systems lead to different processing signatures of the two systems?

1.2. POTENTIAL CONSEQUENCES FOR SENTENCE PRODUCTION. Different predictions can be made about how sentence production may be influenced by the grammatical ar-

---

[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.

[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.

chitectures of asymmetrical and symmetrical voice systems. In the following, I give a brief overview of psycholinguistic models of sentence production and 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 (Levelt 1989, Bock & Levelt 1994, Bock 1995, Ferreira 2010). 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; that is, 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. Ferreira & Slevc 2007, Ferreira 2010). 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 the 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 second 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 et al. 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, I briefly review the use of pupil-size measurements as an index of cognitive processing load.

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 et al. 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 in order to study cognitive effort or processing load in psychology during the last sixty 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 (Kahneman & Beatty 1966, Ahern & Beatty 1979, Richer & Beatty 1985, Gabay et al. 2011, Laeng et al. 2011, Murphy et al. 2011, Zylberberg et al. 2012, inter alia). 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 et al. 2010). TEPRs have also been shown to be sensitive to other aspects of language processing, such as mismatches between prosody and syntax (Engelhardt et al. 2010), intelligibility of the signal (Zekveld et al. 2010, Zekveld & Kramer 2014), speech rate (Koch & Janse 2016), simultaneous interpretation (Hyönä et al. 1995), frequency effects in lexical decision (Kuchinke et al. 2007), prosody in discourse processing (Zellin et al. 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 (2012) 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).

1.4. CURRENT EXPERIMENTS. The experiments reported in this article employed TEPRs 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 has been described as grammatically 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 TEPRs. 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 between active and passive because additional material has to be planned and produced in the latter. Therefore, distinct TEPRs are expected to be found. This was tested in experiment 2 on German.

The analyses focus on the temporal unfolding of cognitive load while speakers plan and produce sentences. In particular, experimental trials were not reduced to a single number or two by only analyzing peak dilation amplitudes or latencies. 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 nonlinear changes in curve shapes and is thus well suited to investigate TEPRs and to obtain a detailed picture of pupil-size changes over the course of sentence planning. The reasoning behind employing growth-curve analysis is that it increases the sensitivity of pupillometric analyses by assessing pupil-size changes over time (Kuchinsky et al. 2013:31). 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 in turn are reflected in differential TEPR curve shapes.

2. EXPERIMENTS.

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.

PARTICIPANTS. Fifty-three native speakers of Tagalog (mean age = seventeen years; thirteen male, forty female) 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.

---

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

MATERIALS. Target pictures were forty-four colored line drawings of transitive events (see Figure 1), interspersed among seventy-six filler pictures of intransitive events. 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 of five positions on the upper part of the computer screen against a white background.
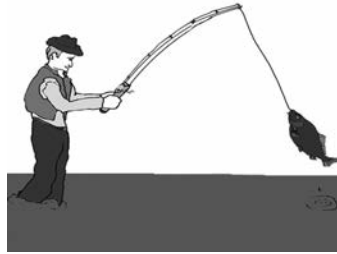

FIGURE 1. Example stimulus picture.

Target pictures were normed by twenty different Tagalog speakers from De La Salle University, who did not participate in the experiment but provided written descriptions of the pictures in a questionnaire. The pictures were then scored with respect to their tendency to be described using agent voice or patient voice or their exhibiting no tendency. Pictures with tendencies toward agent-voice descriptions and toward patient-voice descriptions had an equal share among target items (nineteen pictures with agent-voice tendency, nineteen pictures with patient-voice tendency, five 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-voice sentences and pictures that were likely to elicit patient-voice sentences.

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 of 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 in the depicted event was the pivot argument and two in which the patient 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 ex-

---

[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).

periment started and participants described the target and filler pictures presented to them. Experimental sessions lasted approximately forty minutes.

APPARATUS. Pupil size was recorded with a Tobii T120 remote eye tracker controlled by 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.

DATA PREPROCESSING. 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 a 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 ($\rho = 0.94$); 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 downsampled (Signal Developers 2013) from 120 Hz to 30 Hz, resulting in one sample every ~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).

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 analysis, 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]

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

---

[8] Additionally, participants produced thirty-one locative-voice sentences (2c) and nine benefactive-voice sentences (2e). However, although all nonagent voices (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 nonagent voices could not be assessed with these small numbers of 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.

Linear mixed-effects regression analyses were performed to model pupil-diameter changes by growth-curve analysis (cf. Mirman et al. 2008, Kalénine et al. 2012, Kuchinsky et al. 2013). 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 TEPRs 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 an overall greater increase over the course 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 mean 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, since it mainly describes differences in the tails of the curves. Nevertheless, this time term was included as a 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 and 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 factors other than voice. Speech-onset latency, the mean grayscale 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 (see 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 TEPRs.

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, for example, 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 et al. 2001; cf. also Mathôt et al. 2013). Including the mean grayscale values of the presented pictures in the model accounts for these differences in luminance which could have had an influence on participants' pupil size.

Finally, the codability of the predicate was included as a control variable. Codability reflects how easy or hard the depicted event was to recognize and encode for speakers (cf. van der 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 than or equal to 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.

RESULTS. The upper panel of Figure 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 was 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 a typical shape for pupillary response curves.

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

Table 1 shows the results of the growth-curve analysis for Tagalog. There is a significant interaction between Voice and the linear time term (Time^1), indicating that pupil size increased more over the course of the analysis time window when speakers produced patient-voice sentences than when they produced sentences with agent-voice marking. This effect could be due to a slower decrease of the pupil diameter after having reached its peak for patient-voice trials. The quadratic (Time^2) and cubic (Time^3) time terms did not significantly interact with Voice. This indicates that, overall, the pupillary response curves for agent-voice and patient-voice sentences had similar shapes. Specifically, pupil size increased with similar 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 × Time^1), faster (Speech-onset latency × Time^2), and earlier (Speech-onset latency × Time^3). In addition, the relative luminance of pictures (Mean grayscale value) significantly influenced the shape of pupillary response curves. Speakers' pupils also dilated more in trials in which the predicate codability was low (Predicate codability × Time^1), that is, where it was harder to identify the depicted event and where speakers thus agreed less about which words to use to describe the event.

The linear mixed-effects regression model in Table 1 ignores word-order differences between sentences and only includes Voice as critical predictor variable. The basic word order in Tagalog is predicate-initial, as the sentences in 2 illustrate. However, the order of arguments following the predicate is relatively free. Example 3 demonstrates that different constituent orders are equally grammatical. Agent-voice and patient-voice
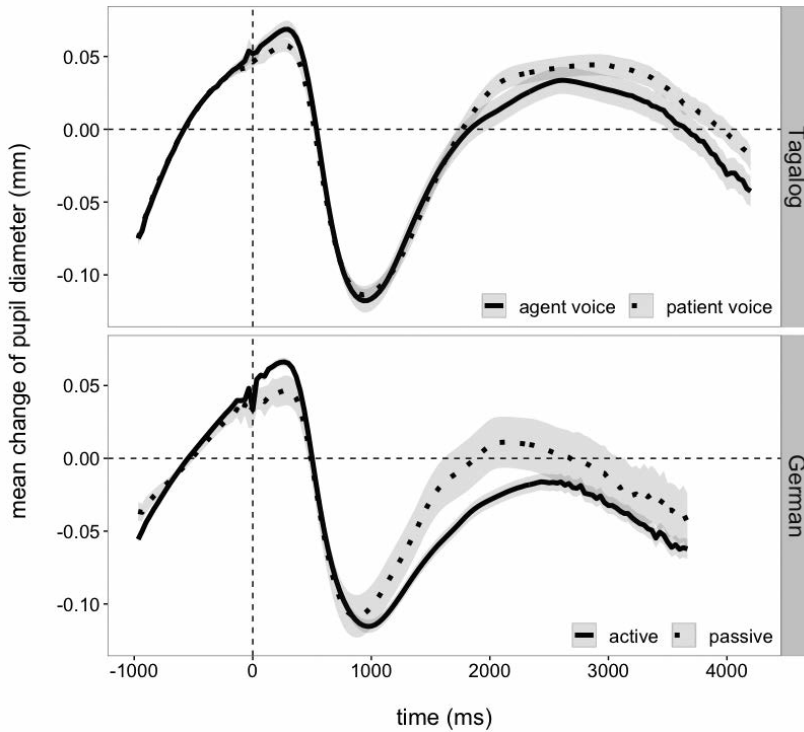
FIGURE 2. Time course of pupil-diameter changes relative to baseline during sentence production in Tagalog (experiment 1, upper panel) and German (experiment 2, lower panel); ribbons indicate one standard error of the mean; dashed vertical line indicates the moment when the to-be-described picture appeared on the screen; dashed horizontal line indicates the baseline pupil diameter.

	$\hat{\beta}$	$|t|$	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 grayscale value	-2.98×10^{-2}	3.54	0.75	
Mean grayscale value × Time^1	-8.53×10^{-2}	2.25	5.06	*
Mean grayscale value × Time^2	8.54×10^{-2}	2.74	7.53	**
Mean grayscale 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	

TABLE 1. Linear mixed-effects regression results (Tagalog).

Note: dependent variable: baseline 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).

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 (Schachter & Otanes 1972, Kroeger 1993b), as in 3a,c.

(3) a. h<**um**>ili ng=isda   **ang=lalaki**
    <AV>catch CORE=fish PVT=man

  b. h<**um**>ili **ang=lalaki** ng=isda
     <AV>catch PVT=man   CORE=fish
      '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
     IRR~catch-PV PVT=fish CORE=man
      '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 TEPRs for sentences in which the *ang*-marked argument was final (as in 3a,c) to sentences in which it was nonfinal (as in 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 ($\hat{\beta}$ = 0.05, |*t*| = 2.24, $\chi^2(1)$ = 5.01, *p* < 0.03). Pivot-final and pivot-medial sentences (as in 3a,b), however, did not differ in their speech-onset latencies ($\hat{\beta}$ = 0.03, |*t*| = 1.02, $\chi^2(1)$ = 1.05, *p* = 0.32). Inclusion of predicate codability as a control variable did not influence the pattern of results for speech-onset latencies.

DISCUSSION. The results of the growth-curve analysis of TEPRs 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 significantly 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 (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 and the linear time term. This means that speakers' pupils on average dilated more over the course of the time window for patient-voice sentences. The TEPR curves in Fig. 2 suggest that speakers' pupils constricted more slowly after having reached their peak dilation during the production of these sentences, which could be the source of this interaction effect. This could mean 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 either of

the quadratic or cubic time terms), but there were more cognitive resources allocated to patient-voice sentences.

The analysis of speech-onset latencies provides additional evidence for the interpretation that the planning of patient-voice sentences demanded 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 suggest that there is no evidence for an asymmetry in the timing and steepness of processing load changes. Rather, the TEPRs resemble what would be expected from a symmetrical voice system. This is in line with Foley's (2008) analysis of the Tagalog voice system (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 increased more when pivot arguments were patients.

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 pupillary responses. 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?

Participants. Thirty-three native speakers of German (mean age = twenty-five years; ten male, twenty-three female) were recruited from Radboud University Nijmegen, from HAN University of Applied Sciences in Nijmegen, and among the Ph.D. 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.

Materials, procedure, and apparatus. The picture stimuli of this experiment consisted of the pictures used in the Tagalog experiment, as well as sixteen additional transitive and seventeen additional filler pictures in order to elicit more picture descriptions from each participant.

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 fifteen example pictures depicting transitive and intransitive events. Pupil size was recorded with a Tobii T120 remote eye tracker running Tobii Studio, sampling both eyes at 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.

DATA PREPROCESSING AND DATA SELECTION. Data preprocessing 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 analysis if they had more than 50% missing data points for the left or right eye before linear interpolation, or 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 nor a passive sentence overtly including the agent (e.g. existentials) or if speakers corrected themselves. This left 1,172 active sentences and 105 passive sentences for analysis.

ANALYSES AND RESULTS. A linear mixed-effects regression model was fit to the TEPRs 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 Fig. 2 above. The overall shape of the pupillary response curves resembles the TEPRs 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 diameters started to increase earlier during the production of passives than during the production of actives. Table 2 shows the regression results for this experiment. The interaction of Voice with the cubic time term was statistically significant. This indicates that pupil diameter increased earlier during the planning of passive sentences.

As in Tagalog, Speech-onset latency and Mean grayscale value 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; that is, pupillary responses were influenced by how much planning time speakers needed before they could initiate articulation, indicating that differences in planning difficulty went hand in hand with variations in mental effort. Predicate codability also significantly influenced TEPRs. 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 × Time^1: $\hat{\beta} = 8.50 \times 10^{-2}$, $|t| = 1.35$, $\chi^2(1) = 1.82$, $p = 0.18$; Voice × Time^2: $\hat{\beta} = -2.68 \times 10^{-3}$, $|t| = 0.06$, $\chi^2(1) < 0.01$, $p = 0.96$; Voice × Time^3: $\hat{\beta} = 5.37 \times 10^{-2}$, $|t| = 1.97$, $\chi^2(1) = 3.90$, $p < 0.05$).

Finally, it is necessary to directly compare the effects observed in the experiments on Tagalog and German in order to be able to conclude that there are significant differ-

	$\hat{\beta}$	$|t|$	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 $\times$ Time^1	4.53×10^{-2}	0.76	0.57	
Voice $\times$ Time^2	-1.08×10^{-2}	0.25	0.06	
Voice $\times$ Time^3	7.12×10^{-2}	2.68	7.20	**
Speech-onset latency	-5.97×10^{-4}	1.29	4.71	*
Speech-onset latency $\times$ Time^1	1.09×10^{-1}	22.19	492.55	***
Speech-onset latency $\times$ Time^2	1.03×10^{-1}	21.06	443.68	***
Speech-onset latency $\times$ Time^3	2.17×10^{-3}	0.45	0.20	
Mean grayscale value	-3.23×10^{-2}	4.78	1.47	
Mean grayscale value $\times$ Time^1	-1.15×10^{-1}	2.99	8.95	**
Mean grayscale value $\times$ Time^2	6.29×10^{-2}	2.71	7.34	**
Mean grayscale value $\times$ Time^3	-3.42×10^{-2}	2.23	4.99	*
Predicate codability (low)	2.90×10^{-2}	2.15	1.13	
Predicate codability $\times$ Time^1	1.71×10^{-1}	2.25	5.05	*
Predicate codability $\times$ Time^2	-3.59×10^{-2}	0.78	0.61	
Predicate codability $\times$ Time^3	-4.47×10^{-2}	1.47	2.17	

TABLE 2. Linear mixed-effects regression results (German).

Note: dependent variable: baseline 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).

ences in TEPR patterns for different voice types between the two languages (see Gelman & Stern 2006, Nieuwenhuis et al. 2011). Thus, an interaction analysis on the joint data of experiments 1 and 2 was carried out, confirming that the TEPR differences between German and Tagalog also hold when compared directly. For this, a linear mixed-effects regression model was fit to model pupil-size changes in an analysis time window between 0 and 4000 ms (grand mean length of produced utterances in both experiments = 3969 ms). This model included Language as an additional predictor variable, which was allowed to interact with the polynomial time terms and Voice.[9] Crucially, a three-way interaction between Voice, Language, and the cubic time term revealed that TEPR curve shapes for different voice types indeed differed between voice systems (Voice $\times$ Language $\times$ Time^3: $\hat{\beta} = -1.10 \times 10^{-1}$, $|t| = 2.50$, $\chi^2(1) = 6.23$, $p = 0.01$). Additionally, Language and Voice each interacted with the linear time term (Language $\times$ Time^1: $\hat{\beta} = 3.30 \times 10^{-1}$, $|t| = 2.98$, $\chi^2(1) = 8.75$, $p < 0.01$; Voice $\times$ Time^1: $\hat{\beta} = 1.04 \times 10^{-1}$, $|t| = 3.05$, $\chi^2(1) = 10.80$, $p < 0.01$). The three-way interaction between the linear time term, Voice, and Language did not reach statistical significance ($p > 0.69$).

When only taking into account the speech-onset latencies from experiment 2 on German, there was no significant difference between actives (mean = 1699 ms, *SD* = 475 ms) and passives (mean = 1778 ms, *SD* = 498 ms; $p = 0.22$). However, a joint analysis pooling the data from experiment 1 and experiment 2 yields a main effect of Voice (Voice: $\hat{\beta} = 0.04$, $|t| = 2.5$, $\chi^2(1) = 6.07$, $p = 0.01$), but no statistically significant interaction between Voice and Language. Thus, across languages, speech-onset latencies for

[9] In order to avoid redundancy with the language predictor, Voice was recoded in this analysis into two categories: A-Voice, encompassing Tagalog agent-voice sentences and German actives, and P-Voice, encompassing Tagalog patient-voice sentences and German passives. This procedure groups together those voice types that assign the most prominent syntactic function to the agent or the patient, but it is not meant to imply that Tagalog voice types behave grammatically the same as German voice types.

sentences in which the patient is linked to the highest syntactic function appear to be generally greater than speech-onset latencies for sentences with agents in the highest syntactic function.

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 were no statistically significant interactions between Voice and either of the linear and quadratic time terms when considering only the German data. The joint analysis of TEPRs from both experiment 1 and experiment 2, however, showed that in German as well as in Tagalog, the production of sentences linking the patient argument to the highest syntactic function (passives in German and patient-voice sentences in Tagalog) required more cognitive effort. This is indicated by the significant interaction of Voice and the linear time term in the analysis of the joint data from both languages reported at the end of the previous section.

The differential shapes of the pupillary response curves for active and passive sentences (as indexed by the interactions with the cubic time term) 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 different TEPRs 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 on-line 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 shapes of the TEPRs 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, for example, 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 analysis of the joint data from both experiments showed that 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 TEPR 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—the pupil-size changes tell a more nuanced story than what could have been learned from speech-onset latencies alone. The joint analysis of speech-onset latencies

from both experiments suggests that the production of both patient-voice sentences and passives required more cognitive effort than the production of both agent-voice sentences and actives, with similar differences between voice types in both languages. However, 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, the similar TEPRs in Tagalog for sentences linking either the agent or the patient semantic roles to the highest syntactic function suggest that similar planning operations are performed to do this. The two German voice types, by contrast, appear to employ different operations, as indexed by diverging pupillary responses. Analyzing pupillometric data in addition to speech-onset latencies made it possible to 'distinguish mental effort from behavioral performance' (Karatekin et al. 2004:184), showing that patterns of cognitive resource allocation can give different insights from patterns of speech-onset latencies. These two kinds of data are thus best considered synergistically.

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

In Tagalog, both agent-voice and patient-voice sentences exhibited similar pupillary response curve shapes; that is, there was no evidence of asymmetrical changes in cognitive load during the planning and production of different voice types. 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 shapes. 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—relatively similar TEPR curves 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 TEPRs 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. 2012), and reliable effects can occur at least 200–300 ms after a cognitive event (Beatty 1982). However, Wierda and colleagues (2012) propose that pupil-diameter changes result from attentional pulses. This suggests that there is no single neural event that is the source of the

TEPRs 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 (see e.g. Ferreira 2010), and that it is to some degree also a temporal black box for which it is hard to say which process has been carried out at exactly which point in time.

Another finding of the current experiments is that, in Tagalog, 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 sizes might have decreased more slowly after having reached their peak dilation and speakers also needed more time to start speaking 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 productions 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 (McFarland 1984, Latrouite 2011).

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, 1995b) argues that syntactic privileges for different constructions are divided between the pivot and the agent argument; agents are, for example, binders of reflexives irrespective of whether they are the pivot (Schachter 1977).

That processing load increased more 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 showing that speakers started speaking later for patient voice.

By the same token, however, this effect is compatible with the notion of (relative) symmetry in the Tagalog voice system. Cognitive load 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 similar 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—cognitive effort increased earlier for the production of passives. The joint analysis of both experiments also showed that processing load increased more over the course of the analysis time window and that speech-onset latencies were longer when German speakers produced passives. It is, however, not possible to disentangle the mental effort of mapping the patient argument to the subject function from 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 (Bornkessel-Schlesewsky & Schlesewsky 2009, 2013b, Alday et al. 2014; see also Sauppe 2016). 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 et al. 1995), and English (Traxler et al. 2002), among other languages. A similar effect has also been demonstrated in Chinese (Wang et al. 2009), Hindi (Choudhary et al. 2009, Bickel et al. 2015), and Turkish (Demiral et al. 2008).

This bias may help listeners keep the structures 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): that is, they can launch actions without patients (as in *Mary was working all day*); this does not hold true for patients that are affected by (causal) actions that must be instigated by an agent.

Together with the evidence from the current experiments, 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. Kempen et al. 2012, MacDonald 2013, Pickering & Garrod 2013, Dell & Chang 2014). 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 nonagents 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 that 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 (Decety & Sommerville 2003, David et al. 2006, Frith & Frith 2010). Additionally, parts of Broca's area, a brain region that is also involved in language processing, are involved in the representation of actions and goal-directed human movements (Clerget et al. 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', that is, 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 involve 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 not marked: 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 et al. 2011), among 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 article support 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 article contributes to the literature on understudied languages (Jaeger & Norcliffe 2009, Norcliffe et al. 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.

REFERENCES

Ahern, Sylvia, and Jackson Beatty. 1979. Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science* 205.1289–92. DOI: 10.1126/science.472746.

Alday, Philip M.; Matthias Schlesewsky; and Ina Bornkessel-Schlesewsky. 2014. Towards a computational model of actor-based language comprehension. *Neuroinformatics* 12.143–79. DOI: 10.1007/s12021-013-9198-x.

Aldridge, Edith. 2012. Antipassive and ergativity in Tagalog. *Lingua* 122.192–203. DOI: 10.1016/j.lingua.2011.10.012.

Alnæs, Dag; Markus Handal Sneve; Thomas Espeseth; Tor Endestad; Steven Harry Pieter van de Pavert; and Bruno Laeng. 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):1. DOI: 10.1167/14.4.1.

Barr, Dale 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, Dale J.; Roger Levy; Christoph Scheepers; and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–78. DOI: 10.1016/j.jml.2012.11.001.

Bates, Douglas; Martin Mächler; Ben Bolker; and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48. DOI: 10.18637/jss.v067.i01.

Beatty, Jackson. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91.276–92. DOI: 10.1037/0033-2909.91.2.276.

Bergamin, Oliver, and Randy H. Kardon. 2003. Latency of the pupil light reflex: Sample rate, stimulus intensity, and variation in normal subjects. *Investigative Ophthalmology & Visual Science* 44.1546–54. DOI: 10.1167/iovs.02-0468.

Bickel, Balthasar; Alena Witzlack-Makarevich; Kamal K. Choudhary; Matthias Schlesewsky; and Ina Bornkessel-Schlesewsky. 2015. The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE* 10:e0132819. DOI: 10.1371/journal.pone.0132819.

Blake, Frank R. 1906. Expression of case by the verb in Tagalog. *Journal of the American Oriental Society* 27.183–89. DOI: 10.2307/592858.

Bloomfield, Leonard. 1917. *Tagalog texts with grammatical analysis*. Urbana: University of Illinois.

Bock, Kathryn. 1995. Sentence production: From mind to mouth. *Speech, language, and communication: Handbook of perception and cognition*, ed. by Joanne L. Miller and Peter D. Eimas, 181–216. San Diego: Academic Press.

Bock, Kathryn, and W. J. M. Levelt. 1994. Language production: Grammatical encoding. *Handbook of psycholinguistics*, 1st edn., ed. by Morton Ann Gernsbacher, 945–84. San Diego: Academic Press.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2009. The role of prominence information in the real-time comprehension of transitive constructions: A crosslinguistic approach. *Language and Linguistics Compass* 3.19–58. DOI: 10.1111/j.1749-818X.2008.00099.x.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2013a. Neurotypology: Modeling crosslinguistic similarities and differences in the neurocognition of language comprehension. *Language down the garden path: The cognitive and biological basis for linguistic structures*, ed. by Montserrat Sanz, Itziar Laka, and Michael K. Tanenhaus, 241–52. Oxford: Oxford University Press.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2013b. Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain and Language* 125.60–76. DOI: 10.1016/j.bandl.2013.01.010.

Choudhary, Kamal Kumar; Matthias Schlesewsky; Dietmar Roehm; and Ina Bornkessel-Schlesewsky. 2009. The N400 as a correlate of interpretively relevant linguistic rules: Evidence from Hindi. *Neuropsychologia* 47.3012–22. DOI: 10.1016/j.neuropsychologia.2009.05.009.

Clerget, Emeline; Aline Winderickx; Luciano Fadiga; and Etienne Olivier. 2009. Role of Broca's area in encoding sequential human actions: A virtual lesion study. *NeuroReport* 20.1496–99. DOI: 10.1097/WNR.0b013e3283329be8.

Cohen, Jacob; Patricia Cohen; Stephen G. West; and Leona S. Aiken. 2003. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd edn. Mahwah, NJ: Lawrence Erlbaum.

Cole, Peter; Gabriella Hermon; and Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118.1500–1553. DOI: 10.1016/j.lingua.2007.08.008.

David, Nicole; Bettina H. Bewernick; Michael X. Cohen; Albert Newen; Silke Lux; Gereon R. Fink; N. Jon Shah; and Kai Vogeley. 2006. Neural representations of self versus other: Visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience* 18.898–910. DOI: 10.1162/jocn.2006.18.6.898.

Decety, Jean, and Jessica A. Sommerville. 2003. Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences* 7.527–33. DOI: 10.1016/j.tics.2003.10.004.

Dell, Gary S., and Franklin Chang. 2014. The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369:20120394. DOI: 10.1098/rstb.2012.0394.

Demiral, Şükrü Barış; Matthias Schlesewsky; and Ina Bornkessel-Schlesewsky. 2008. On the universality of language comprehension strategies: Evidence from Turkish. *Cognition* 106.484–500. DOI: 10.1016/j.cognition.2007.01.008.

Donohue, Mark. 2008. Malay as a mirror of Austronesian: Voice development and voice variation. *Lingua* 118.1470–99. DOI: 10.1016/j.lingua.2007.08.007.

Dunn, Michael; Simon J. Greenhill; Stephen C. Levinson; and Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473.79–82. DOI: 10.1038/nature09923.

Engelhardt, Paul E.; Fernanda Ferreira; and Elenea G. Patsenko. 2010. Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology* 63.639–45. DOI: 10.1080/17470210903469864.

Evans, Nicholas, and Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32.429–48. DOI: 10.1017/S0140525X0999094X.

Fazio, Patrik; Anna Cantagallo; Laila Craighero; Alessandro D'Ausilio; Alice C. Roy; Thierry Pozzo; Ferdinando Calzolari; Enrico Granieri; and Luciano Fadiga. 2009. Encoding of human action in Broca's area. *Brain* 132.1980–88. DOI: 10.1093/brain/awp118.

Ferreira, Victor S. 2010. Language production. *Wiley Interdisciplinary Reviews: Cognitive Science* 1.834–44. DOI: 10.1002/wcs.70.

Ferreira, Victor S., and L. Robert Slevc. 2007. Grammatical encoding. *The Oxford handbook of psycholinguistics*, ed. by M. Gareth Gaskell, 453–69. Oxford: Oxford University Press.

Foley, William A. 2008. The place of Philippine languages in a typology of voice systems. *Voice and grammatical relations in Austronesian languages*, ed. by Peter K. Austin and Simon Musgrave, 22–44. Stanford, CA: CSLI Publications.

Foley, William A., and Robert D. Van Valin, Jr. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Fox, John, and Sanford Weisberg. 2011. *An R companion to applied regression*. 2nd edn. R package version 2.1-1. Thousand Oaks, CA: Sage.

Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory* 5.519–59. DOI: 10.1007/BF00138988.

FRITH, UTA, and CHRIS FRITH. 2010. The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.165–75. DOI: 10.1098/rstb.2009.0160.

GABAY, SHAI; YONI PERTZOV; and AVISHAI HENIK. 2011. Orienting of attention, pupil size, and the norepinephrine system. *Attention, Perception, & Psychophysics* 73.123–29. DOI: 10.3758/s13414-010-0015-4.

GELMAN, ANDREW, and HAL STERN. 2006. The difference between 'significant' and 'not significant' is not itself statistically significant. *The American Statistician* 60(4).328–31. DOI: 10.1198/000313006X152649.

GOLDINGER, STEPHEN D., and MEGAN H. PAPESH. 2012. Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science* 21.90–95. DOI: 10.1177/0963721412436811.

GRIFFIN, ZENZI M., and KATHRYN BOCK. 2000. What the eyes say about speaking. *Psychological Science* 11.274–79. DOI: 10.1111/1467-9280.00255.

HAGGARD, PATRICK. 2008. Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience* 9.934–46. DOI: 10.1038/nrn2497.

HESS, ECKHARD H., and JAMES M. POLT. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143.1190–92. DOI: 10.1126/science.143.3611.1190.

HIMMELMANN, NIKOLAUS P. 2002. Voice in western Austronesian: An update. *The history and typology of western Austronesian voice systems* (Pacific linguistics 518), ed. by Fay Wouk and Malcolm Ross, 7–16. Canberra: Australian National University.

HIMMELMANN, NIKOLAUS P. 2004. Tagalog (Austronesian). *Morphology: An international handbook on inflection and word-formation*, ed. by Geert Booij, Christian Lehmann, Joachim Mugdan, Stavros Skopeteas, and Wolfgang Kesselheim, 1473–90. Berlin: Mouton de Gruyter.

HIMMELMANN, NIKOLAUS P. 2005a. The Austronesian languages of Asia and Madagascar: Typological characteristics. *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 110–81. Oxon: Routledge.

HIMMELMANN, NIKOLAUS P. 2005b. Tagalog. *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 350–76. Oxon: Routledge.

HIMMELMANN, NIKOLAUS P. 2008. Lexical categories and voice in Tagalog. *Voice and grammatical relations in Austronesian languages*, ed. by Peter K. Austin and Simon Musgrave, 247–93. Stanford, CA: CSLI Publications.

HYÖNÄ, JUKKA; JORMA TOMMOLA; and ANNA-MARI ALAJA. 1995. Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* 48.598–612.

JAEGER, T. FLORIAN, and ELISABETH J. NORCLIFFE. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3.866–87. DOI: 10.1111/j.1749-818x.2009.00147.x.

JUST, MARCEL A., and PATRICIA A. CARPENTER. 1993. The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology* 47.310–39. DOI: 10.1037/h0078820.

KAHNEMAN, DANIEL, and JACKSON BEATTY. 1966. Pupil diameter and load on memory. *Science* 154.1583–85. DOI: 10.1126/science.154.3756.1583.

KALÉNINE, SOLÈNE; DANIEL MIRMAN; ERICA L. MIDDLETON; and LAUREL J. BUXBAUM. 2012. Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38.1274–95. DOI: 10.1037/a0027626.

KARATEKIN, CANAN; JANE W. COUPERUS; and DAVID J. MARCUS. 2004. Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology* 41.175–85. DOI: 10.1111/j.1469-8986.2004.00147.x.

KEMMERER, DAVID. 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* 6.50–66. DOI: 10.1002/lnc3.322.

KEMPEN, GERARD; NOMI OLSTHOORN; and SIMONE SPRENGER. 2012. Grammatical workspace sharing during language production and language comprehension: Evidence from grammatical multitasking. *Language and Cognitive Processes* 27.345–80. DOI: 10.1080/01690965.2010.544583.

Koch, Xaver, and Esther Janse. 2016. Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America* 139.1618–36. DOI: 10.1121/1.4944032.

Kroeger, Paul. 1993a. Another look at subjecthood in Tagalog. *Philippine Journal of Linguistics* 24.1–16.

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

Kuchinke, Lars; Melissae L.-H. Võ; Markus Hofmann; and Arthur M. Jacobs. 2007. Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology* 65.132–40. DOI: 10.1016/j.ijpsycho.2007.04.004.

Kuchinsky, Stefanie E.; Jayne B. Ahlstrom; Kenneth I. Vaden, Jr.; Stephanie L. Cute; Larry E. Humes; Judy R. Dubno; and Mark A. Eckert. 2013. Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology* 50.23–34. DOI: 10.1111/j.1469-8986.2012.01477.x.

Laeng, Bruno; Marte Ørbo; Terje Holmlund; and Michele Miozzo. 2011. Pupillary Stroop effects. *Cognitive Processing* 12.13–21. DOI: 10.1007/s10339-010-0370-z.

Laeng, Bruno; Sylvain Sirois; and Gustav Gredebäck. 2012. Pupillometry: A window to the preconscious? *Perspectives on Psychological Science* 7.18–27. DOI: 10.1177/1745691611427305.

Latrouite, Anja. 2011. *Voice and case in Tagalog: The coding of prominence and orientation*. Düsseldorf: Heinrich-Heine-Universität dissertation.

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. *The neurocognition of language*, ed. by Colin M. Brown and Peter Hagoort, 83–122. Oxford: Oxford University Press.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4:226. DOI: 10.3389/fpsyg.2013.00226.

Maclachlan, Anna. 1996. *Aspects of ergativity in Tagalog*. Montreal: McGill University dissertation.

Mathôt, Sebastiaan; Lotje van der Linden; Jonathan Grainger; and Françoise Vitu. 2013. The pupillary light response reveals the focus of covert visual attention. *PLoS ONE* 8:e78168. DOI: 10.1371/journal.pone.0078168.

McFarland, Curtis D. 1984. Tagalog causative verbs. *Philippine Journal of Linguistics* 16.19–47.

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

Mirman, Daniel; James A. Dixon; and James S. Magnuson. 2008. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language* 59.475–94. DOI: 10.1016/j.jml.2007.11.006.

Murphy, Peter R.; Ian H. Robertson; Joshua H. Balsters; and Redmond G. O'Connell. 2011. Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology* 48.1532–43. DOI: 10.1111/j.1469-8986.2011.01226.x.

Nieuwenhuis, Sander; Birte U. Forstmann; and Eric-Jan Wagenmakers. 2011. Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience* 14(9).1105–7. DOI: 10.1038/nn.2886.

Norcliffe, Elisabeth; Alice C. Harris; and T. Florian Jaeger. 2015. Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Language, Cognition and Neuroscience* 30.1009–32. DOI: 10.1080/23273798.2015.1080373.

Papesh, Megan H., and Stephen D. Goldinger. 2012. Pupil-BLAH-metry: Cognitive effort in speech planning reflected by pupil dilation. *Attention, Perception, & Psychophysics* 74.754–65. DOI: 10.3758/s13414-011-0263-y.

Paul, Ileana, and Lisa Travis. 2006. Ergativity in Austronesian languages: What it can do, what it can't, but not why. *Ergativity: Emerging issues*, ed. by Alana Johns, Diane Massam, and Juvenal Ndayiragije, 315–35. Dordrecht: Springer.

Payne, Thomas E. 1982. Role and reference related subject properties and ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6.75–106. DOI: 10.1075/sl.6.1.05pay.

PICKERING, MARTIN J., and SIMON GARROD. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36.329–47. DOI: 10.1017/S0140525X12001495.

PIQUADO, TEPRING; DEREK ISAACOWITZ; and ARTHUR WINGFIELD. 2010. Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* 47.560–69. DOI: 10.1111/j.1469-8986.2009.00947.x.

PRIMUS, BEATRICE. 1999. *Cases and thematic roles: Ergative, accusative and active*. Tübingen: Max Niemeyer.

R CORE TEAM. 2015. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: https://www.r-project.org/.

RICHER, FRANCOIS, and JACKSON BEATTY. 1985. Pupillary dilations in movement preparation and execution. *Psychophysiology* 22.204–7. DOI: 10.1111/j.1469-8986.1985.tb01587.x.

RIESBERG, SONJA. 2014a. Passive actors are not adjuncts—Consequences for the distinction between symmetrical and asymmetrical voice alternations. *Argument realisations and related constructions in Austronesian languages: Papers from 12-ICAL*, vol. 2, ed. by I Wayan Arka and N. L. K. Mas Indrawati, 281–302. Canberra: Australian National University.

RIESBERG, SONJA. 2014b. *Symmetrical voice and linking in western Austronesian languages*. Berlin: De Gruyter Mouton.

RIESBERG, SONJA, and BEATRICE PRIMUS. 2015. Agent prominence in symmetrical voice languages. *STUF—Language Typology and Universals* 68.551–64. DOI: 10.1515/stuf-2015-0023.

ROSS, MALCOLM. 2002. The history and transitivity of western Austronesian voice and voice-marking. *The history and typology of western Austronesian voice systems* (Pacific linguistics 518), ed. by Fay Wouk and Malcolm Ross, 17–62. Canberra: Australian National University.

SAMUELS, EBONY R., and ELEMER SZABADI. 2008. Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function. Part 1: Principles of functional organisation. *Current Neuropharmacology* 6.235–53. DOI: 10.2174/157015908785777229.

SARA, SUSAN J. 2009. The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience* 10.211–23. DOI: 10.1038/nrn2573.

SAUPPE, SEBASTIAN. 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, SEBASTIAN. 2017. Word order and voice influence the timing of verb planning in German sentence production. Nijmegen: Max Planck Institute for Psycholinguistics, MS.

SAUPPE, SEBASTIAN; ELISABETH NORCLIFFE; AGNIESZKA E. KONOPKA; ROBERT D. VAN VALIN, JR.; and STEPHEN C. LEVINSON. 2013. Dependencies first: Eye tracking evidence from sentence production in Tagalog. *Proceedings of the 35th annual conference of the Cognitive Science Society (CogSci 2013)*, 1265–70. Online: https://mindmodeling.org/cogsci2013/papers/0242/.

SCHACHTER, PAUL. 1976. The subject in Philippine languages: Topic, actor, actor-topic, or none of the above. *Subject and topic*, ed. by Charles N. Li, 491–518. New York: Academic Press.

SCHACHTER, PAUL. 1977. Reference-related and role-related properties of subjects. *Grammatical relations*, ed. by Peter Cole and Jerrold M. Sadock, 279–306. New York: Academic Press.

SCHACHTER, PAUL. 1995a. Tagalog. *Syntax: An international handbook of contemporary research*, ed. by Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, 1418–30. Berlin: Walter de Gruyter.

SCHACHTER, PAUL. 1995b. *The subject in Tagalog: Still none of the above*. (UCLA occasional papers in linguistics 15.) Los Angeles: Department of Linguistics, University of California, Los Angeles. Online: http://www.linguistics.ucla.edu/publications/OPL_15.pdf.

SCHACHTER, PAUL, and FE T. OTANES. 1972. *Tagalog reference grammar*. Berkeley: University of California Press.

SCHLUROFF, MICHAEL; THOMAS E. ZIMMERMANN; R. B. FREEMAN, JR.; KLAUS HOFMEISTER; THOMAS LORSCHEID; and ARNO WEBER. 1986. Pupillary responses to syntactic ambiguity of sentences. *Brain and Language* 27.322–44. DOI: 10.1016/0093-934X (86)90023-4.

SCHMIDTKE, JENS. 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, HERBERT; ANGELA D. FRIEDERICI; and KATJA KÜHN. 1995. The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language* 34.499–520. DOI: 10.1006/jmla.1995.1023.

SHANNON, CLAUDE E. 1948. A mathematical theory of communication. *Bell Labs Technical Journal* 27.379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

SIGNAL DEVELOPERS. 2013. signal: Signal processing. R package version 0.7-4. Online: https://cran.r-project.org/web/packages/signal/.

SIROIS, SYLVAIN, and JULIE BRISSON. 2014. Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science* 5(6).679–92. DOI: 10.1002/wcs.1323.

TANANGKINGSING, MICHAEL, and SHUANFAN HUANG. 2007. Cebuano passives revisited. *Oceanic Linguistics* 46.554–84. DOI: 10.1353/ol.2008.0006.

TRAXLER, MATTHEW J.; ROBIN K. MORRIS; and RACHEL E. SEELY. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47.69–90. DOI: 10.1006/jmla.2001.2836.

van de VELDE, MAARTJE; ANTJE S. MEYER; and AGNIESZKA E. KONOPKA. 2014. Message formulation and structural assembly: Describing 'easy' and 'hard' events with preferred and dispreferred syntactic structures. *Journal of Memory and Language* 71.124–44. DOI: 10.1016/j.jml.2013.11.001.

VERNEY, STEVEN P.; ERIC GRANHOLM; and DAPHNE P. DIONISIO. 2001. Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology* 38.76–83. DOI: 10.1111/1469-8986.3810076.

WANG, LUMING; MATTHIAS SCHLESEWSKY; BALTHASAR BICKEL; and INA BORNKESSEL-SCHLESEWSKY. 2009. Exploring the nature of the 'subject'-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes* 24.1180–1226. DOI: 10.1080/01690960802159937.

WIERDA, STEFAN M.; HEDDERIK van RIJN; NIELS A. TAATGEN; and SANDER MARTENS. 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* 109.8456–60. DOI: 10.1073/pnas.1201858109.

ZEILEIS, ACHIM, and GABOR GROTHENDIECK. 2005. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software* 14.1–27. DOI: 10.18637/jss.v014 .i06.

ZEKVELD, ADRIANA A., and SOPHIA E. KRAMER. 2014. Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology* 51.277–84. DOI: 10.1111/psyp.12151.

ZEKVELD, ADRIANA A.; SOPHIA E. KRAMER; and JOOST M. FESTEN. 2010. Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear & Hearing* 31.480–90. DOI: 10.1097/AUD.0b013e3181d4f251.

ZELLIN, MARTINA; ANN PANNEKAMP; ULRIKE TOEPEL; and ELKE van der MEER. 2011. In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology* 81.133–41. DOI: 10.1016/j.ijpsycho.2011.05.009.

ZYLBERBERG, ARIEL; MANUEL OLIVA; and MARIANO SIGMAN. 2012. Pupil dilation: A fingerprint of temporal selection during the 'attentional blink'. *Frontiers in Psychology* 3:316. DOI: 10.3389/fpsyg.2012.00316.

[sauppe.s@gmail.com]