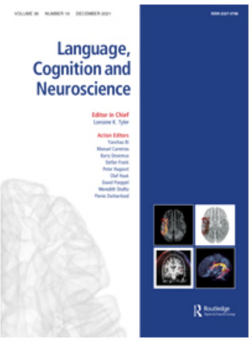

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Sentence formulation is easier when thematic and syntactic prominence align: evidence from psych verbs

Monica L. Do & Elsi Kaiser

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2021.2008458

Sentence formulation is easier when thematic and syntactic prominence align: evidence from psych verbs

Monica L. Do and Elsi Kaiser

Linguistics, University of Southern California, Los Angeles, USA

ABSTRACT
We use the visual world eye-tracking paradigm to investigate how the mapping from thematic event structures to grammatical structures, known as sentence formulation, unfolds during real time sentence production. Experiment 1 contrasted production of SubjExp (*"Leslie$_{EXP}$ loves Ann$_{STIM}$"*) versus ObjExp (*"Leslie$_{STIM}$ scares Ann$_{EXP}$"*) sentences. Experiment 2 investigated passivized SubjExp (*"Leslie$_{STIM}$ was loved by Ann$_{EXP}$"*) and passivized ObjExp sentences (*"Leslie$_{EXP}$ was scared by Ann$_{STIM}$."*). In both studies, we found that speakers were faster to begin speaking and to preferentially fixate the subject when they were able to assign the thematically prominent Experiencer role to the subject of the sentence. We conclude that sentence formulation is easier when speakers can make use of a tight, systematic correspondence between event structures and linguistic structures. We discuss the implications of our work for the relationship between language and thought and for the formal accounts of SubjExp and ObjExp verbs.

ARTICLE HISTORY
Received 3 February 2021
Accepted 10 November 2021

KEYWORDS
Sentence formulation; psych verbs; eye-tracking; language production; passives; thematic roles; subject selection

1. Introduction

Talking is a complicated process. In order to talk about something that happened, speakers first have to create a conceptual representation of the event (Bock & Levelt, 1994; Levelt, 1989; inter alia). These conceptual representations include information about the type of event that is happening, the entities involved, and critically, the roles that each entity plays in the event. They might also include other information like the telicity, temporal aspect, or location of the event. Importantly, though, in order to formulate a sentence about what they just saw, speakers need to encode that conceptual representation into a linguistic representation suitable for the language they speak. This means that in addition to retrieving the right words and the right sounds for the concept they want to talk about, speakers also need to *map* their representation of an event onto a linguistic structure that contains grammatical information about the syntactic roles that each concept plays in the sentence.

1.1 Thematic roles and the mapping process

Prior work on the process of mapping a conceptual representation onto a linguistic one has centred around the question of subject selection: Which entity in a speaker's conceptual representation of an event will be selected as the grammatical subject of the utterance? On this question, prior work has suggested that in most instances, speakers typically map the most conceptually salient entity to the most syntactically prominent role, which in English is typically the subject of the sentence (e.g. Bock et al., 1992; Bock & Warren, 1985). Evidence for this largely comes from production studies of the active/passive alternation. These studies have shown that under most circumstances the entities that are most conceptually salient – entities which are animate (Bock, 1982; Bock et al., 1992; Clark, 1965), human (e.g. Clark & Begun, 1971), or highly imageable/concrete (Bock & Warren, 1985) – tend to be realised as the grammatical subject of a sentence. In describing a "who did what to whom" Agent-Patient (Agt-Pat) event in which Leslie criticises a painting, for instance, it is likely that Leslie, a human, would be selected as the subject and speakers would consequently produce an active voice sentence such as "*Leslie criticized the painting*". Prior work has shown, though, that in some cases, other factors can override this tendency: When perceptual (Gleitman et al., 2007; Rissman et al., 2018) or discourse-pragmatical factors (e.g. Bock, 1982; Bock & Irwin, 1980; Ferreira & Yoshita, 2003; Osgood, 1971; Prat-Salá & Branigan, 2000; Prentice, 1967; inter alia)

CONTACT Monica L. Do monicado@uchicago.edu Department of Linguistics, University of Chicago, 1115 E. 58th Street, Rosenwald Hall Room 229C, Chicago, IL 60637, USA
Supplemental data for this article can be accessed https://doi.org/10.1080/23273798.2021.2008458.

make an otherwise less prominent entity (e.g. the inanimate painting) temporarily more salient, that entity can be promoted to the subject role (for related work on crosslinguistic variation, see e.g. Hwang & Kaiser, 2015). In these cases, speakers might produce a passive such as "*The painting was criticized by Leslie*".

In addition to the individual characteristics of an entity that make it likely to be selected as the subject, work by Ferreira (1994) has shown that the thematic role – the role that an entity plays in the larger event – can also influence subject selection. Specifically, Ferreira (1994) showed that speakers prefer to map the most prominent element in the Thematic Hierarchy (e.g. Fillmore, 1968; Jackendoff, 1972, 1987; Grimshaw, 1990; Stevenson et al., 1994; Figure 1) to the most prominent grammatical role in the sentence.

Ferreira (1994) demonstrated this preference by asking participants to make sentences using two nouns and one verb. Crucially, the verbs came in two varieties: They were either (i) "normal" Agent-Patient verbs (e.g. "*Leslie$_{AGT}$ punished/praised/critized Ann$_{PAT}$.*"; Agt-Pat) and Subject Experiencer verbs (SubjExp; "*Leslie$_{EXP}$ loves/fears/admires Ann$_{STIM}$.*") – where the higher-ranked Agent and Experiencer appear as the subject of an active voice sentence or (ii) they were Object Experiencer verbs (ObjExp; "*Leslie$_{STIM}$ surprises/scares/annoys Ann$_{EXP}$*"), where the thematically *less* prominent Stimulus occurs in the subject position.[1] (The linguistic term "Psych verbs" refers to both SubjExp and ObjExp verbs; we distinguish these two classes in the present work due to the differences in their thematic role alignment.)

Ferreira (1994) reasoned that if prominence along the Thematic Hierarchy could affect subject selection, then speakers should produce more active voice sentences in the case of Agt-Pat and SubjExp verbs, where the active structure allows the more prominent Agent and Experiencer, respectively, to occur in the subject position. Likewise, they should produce more passive voice sentences in ObjExp conditions, where it is the passive structure ("*Leslie$_{EXP}$ was scared by Ann$_{STIM}$.*") that allows the more thematically prominent Experiencer to become the subject of the sentence. Indeed, when Ferreira (1994) compared the proportion of passives produced in each condition, she found that speakers were significantly more likely to produce passives in the the ObjExp verb conditions than in either the "normal" Agt-Pat verb or SubjExp verb conditions. Ferreira (1994) thus concluded that speakers prefer to map the most prominent element in the Thematic Hierarchy (e.g. Fillmore, 1968; Jackendoff, 1972, 1987; Grimshaw, 1990; Stevenson et al., 1994; Figure 1) to the most prominent grammatical role in the sentence. These results were further corroborated by Gennari and MacDonald (2009), who likewise showed higher rates of passivization among ObjExp verbs. A consistent pattern of results emerged among aphasics in work by Thompson and Lee (2009). In that study, errors rates were higher when describing ObjExp versus SubjExp images using the active voice, but the pattern of results was reversed when describing the same images using passive voice.

Thematic Hierarchy

Agent >> Experiencer >> Goal/Source/Instrument >> Stimulus >> Patient

Figure 1. A version of the Thematic Hierarchy, simplified for present purposes. Accounts differ with respect to the precise number and ordering of some roles in the Thematic Hierarchy. The version shown above, however, is consistent with most Thematic Hierarchy accounts.

Surprisingly, though, in spite of these patterns, the question of whether misalignment of thematic and syntactic prominence has any *real-time* consequences for processes related to sentence formulation remains open. There has been little direct evidence from prior studies, for instance, to suggest that unimpaired speakers have any difficulty producing either SubjExp or ObjExp sentences in the active or passive voice. In the case of Ferreira (1994), for instance, comparison of sentence formulation times did not show any difference between verb types; likewise, the generally low rate of errors in both the active and passive conditions suggest that Thompson and Lee's (2009) unimpaired speakers, unlike aphasics, perform similarly across these conditions. Because little is known about how the mapping between thematic and syntactic prominence can affect the moment-by-moment processes in language production, the related question of what ultimately motivates the tendency to assign the most thematically privileged argument of an event to the subject of the sentence remains open.

1.2. The current study

In the present work, we focus on SubjExp and ObjExp sentences because they allow us to investigate how the mapping between thematic roles and syntactic positions can affect processes in language production when the conceptual roles, number of words, surface structure, and class of event (Levin, 1993) are held constant. We extend prior work by using the visual world eye-tracking paradigm in a real-time language production task. This enables us to tap into measures reflecting incremental sentence formulation processes, even before participants start to speak, to investigate the core question of whether sentence formulation is more difficult when the preference to the map the most prominent thematic role in an event to the most prominent syntactic position in the sentence is violated.

As will become clear below, our results show that, by taking a more fine-grained look at the production process for SubjExp and ObjExp verbs, we are able to detect differences that were not necessarily apparent in earlier studies. At a theoretical level, this helps us to more precisely understand why the link between thematic and syntactic prominence appears so robustly – not only across such domains as varied as language production (Ferreira, 1994; Thompson & Lee, 2009), anaphora resolution (e.g. Schumacher et al., 2017; Stevenson et al., 1994), and language acquisition (Gleitman, 1990; Fisher et al. 2010; Hirsch-Pasek & Golinkoff, 1996), but also across different types of individuals (Thompson & Lee, 2009). Given that the mapping process fundamentally serves as a bridge between speakers' pre-verbal conceptual representations and their linguistic representations, the results of the present work can thus shed light on the nature of the relationship – in particular, the timing and interaction – between event structures and linguistic representations, more generally. Finally, our studies can also inform open question in formal areas of linguistics, where the syntactic representations of SubjExp and ObjExp Verbs is still under debate.

In two language production experiments, we track speakers' speech onset times and eye-movements as they plan active SubjExp sentences (*"Leslie$_{EXP}$ loves/fears/admires Ann$_{STIM}$."*) and active ObjExp (*"Leslie$_{STIM}$ surprises/scares/annoys Ann$_{EXP}$."*) sentences (Experiment 1), and passive SubjExp sentences (*"Leslie$_{STIM}$ was loved/feared/admired by Ann$_{EXP}$."*) and passive ObjExp (*"Leslie$_{EXP}$ was surprised/scared/annoyed by Ann$_{STIM}$."*) sentences (Experiment 2) in real time. (Note that we continue to use active voice labels for SubjExp and ObjExp verbs, even when they are passivized.) In addition to these sentences, active Agent-Patient sentences (Agt-Pat; *"Leslie$_{AGT}$ criticizes/congratulates/blames Ann$_{PAT}$."*) and passive Agent-Patient sentences (*"Leslie$_{AGT}$ was criticized/congratulated/blamed by Ann$_{PAT}$."*) were added as reference conditions in Experiment 1 and Experiment 2, respectively.

If language production is, indeed, more difficult when speakers are unable to map the most prominent role in the event onto the grammatical subject of the sentence, then we expect speakers to be slower to begin uttering "misaligned" sentences where the more prominent Experiencer is the *object* of the sentence. In Experiment 1, on the production of active voice sentences, this misaligned sentence type is exemplified by active voice ObjExp sentences, such as *"Leslie$_{STIM}$ scared Ann$_{EXP}$."*; in Experiment 2, on the production of passive voice sentences, this misalignment occurs during the production of passivized SubjExp sentences, like *"Leslie$_{STIM}$ was loved by Ann$_{EXP}$."*

Data from speakers' eye-movements can also allow us to see whether the inability to map the most prominent thematic argument to the subject position hinders processes in language production, as we can look at the *relative* timing of participants' fixations to the subject character as they are preparing their utterances (i.e. before speech). In particular, prior work on speakers' eye-movements during production has shown that in subject-initial languages, like English, speakers will fixate the character that will become the subject of the sentence almost immediately before naming it (e.g. Griffin & Bock, 2000). If speakers have more difficulty selecting the entity to serve as the subject of the sentence (e.g. because they had to retrieve a less conceptually prominent Stimulus entity or because they had initially retrieved the more prominent Experiencer entity), then it is possible that they will not only be slower to begin speaking, but also be slower to fixate the character that eventually becomes the subject of the sentence. As before, we expect to see this pattern emerge in the ObjExp conditions of Experiment 1, and in the SubjExp conditions of Experiment 2. These predictions are summarised in Table 1.

2. Experiment 1a: active sentences

2.1 Methods

2.1.1 Participants

Thirty-eight adult native speakers of American English were recruited from the University of Southern California and given course/extra credit for their participation. Of these, four were excluded due to a failure to understand the task (i.e. repeatedly failed to use the verb prompt, incorrectly named characters). Data from the remaining 34 participants were submitted for analysis.

Table 1. Summary of predictions for Experiment 1a and Experiment 2.

	Verb Type		Predictions
Exp 1a	SubjExp *Leslie$_{EXP}$ loves Ann$_{STIM}$.*	Easier	• Faster speech onset • Earlier fixations to subject character
	ObjExp *Leslie$_{STIM}$ scares Ann$_{EXP}$.*	Harder	• Slower speech onset • Later fixations to subject character
Exp 2	SubjExp *Leslie$_{STIM}$ was loved by Ann$_{EXP}$.*	Harder	• Slower speech onset • Later fixations to subject character
	ObjExp *Leslie$_{EXP}$ was scared by Ann$_{STIM}$.*	Easier	• Faster speech onset • Earlier fixations to subject character

Table 2. Examples of verbs in each of the three conditions, as well as the kinds of sentences participants are expected to produce. Sample images are shown in Figure 1.

Verb Type	Sample Prompt Verbs [presented on the screen in third person, present tense]	Sample Target Sentence
Experiencer-Stimulus	Loves, fears, admires	Leslie loves Ann.
Stimulus-Experiencer	Surprises, scares, annoys	Leslie scares Ann.
Agent-Patient	Criticises, congratulates, blames	Leslie criticises Ann.

2.1.2 Materials and design

Experiment 1a was a 1-way, within-subjects design involving Verb Type. Following Thompson and Lee (2009), participants were presented with a SubjExp, ObjExp, or an Agt-Pat prompt verb prior to seeing an image on the screen. The task was to describe the image on the screen. Although Agt-Pat verbs are not central to the main research question at hand, we included them as a reference condition because their real-time production is better understood. In particular, given the types of stimuli needed for SubjExp and ObjExp verbs (e.g. images of psychological states), we believed the Agt-Pat condition would be useful in "diagnosing" any potential problems with our design. Participants, therefore, saw eight different verbs in each of the three conditions, yielding a total of 24 target items. Conditions with sample verbs are listed in Table 2, with the full list of verbs in Appendix A.

Verbs were selected using the classification outlined by Levin (1993). ObjExp and SubjExp verbs were drawn exclusively from the "Psych Verb" class (Levin, 1993; Class 31). Although many traditionally-investigated Agt-Pat vebs (e.g. "*hit*", "*shoot*", "*spray*") typically refer to physical actions between two people, the Agt-Pat verbs in our study were drawn primarily from Levin's class of "Judgement Verbs" (Class 33). This is because Judgement Verbs do not refer to physical actions between two people, which meant that we were able to use the same type of non-interactive images that we used with Psych verbs.[2]

Prompt verbs were presented to participants in writing on the computer screen with simple present tense marking (e.g. *scares*) and participants were instructed to use the words in the exact form shown. Given findings by Ferreira (1994), we opted to do this in order to prevent participants from creating passive utterances, especially in the ObjExp conditions (e.g. "*Ann was scared by Leslie*"). Verbs were shown in black, bolded Times New Roman, size 44 font and displayed at the top centre of the screen (about 1.25 in. from the top).

Each critical image screen involved drawings of two characters expressing the psychological states corresponding to the verb shown on the preceding screen. As shown in Figure 2, the characters have a range of facial expressions. The characters were rotated throughout the study such that each character appeared twice as the subject and twice as the object in each condition and was paired an equal number of times with each of the other characters across conditions. In addition, characters were left-right balanced such that they appeared on both the left and right sides of the screen as the subject of the sentence at least once in each condition. Finally, to minimise any effect of familiarity, no character occurred more than once with the same facial expression.

Thirty-six filler items, which also required participants to describe a critical image using a prompt word, were included. In addition to the four human characters, some fillers also depicted one or more of eight distinct animal characters. For the human characters, which were used in both targets and fillers, none of the facial expressions for a particular human in the target items occurred in the fillers.

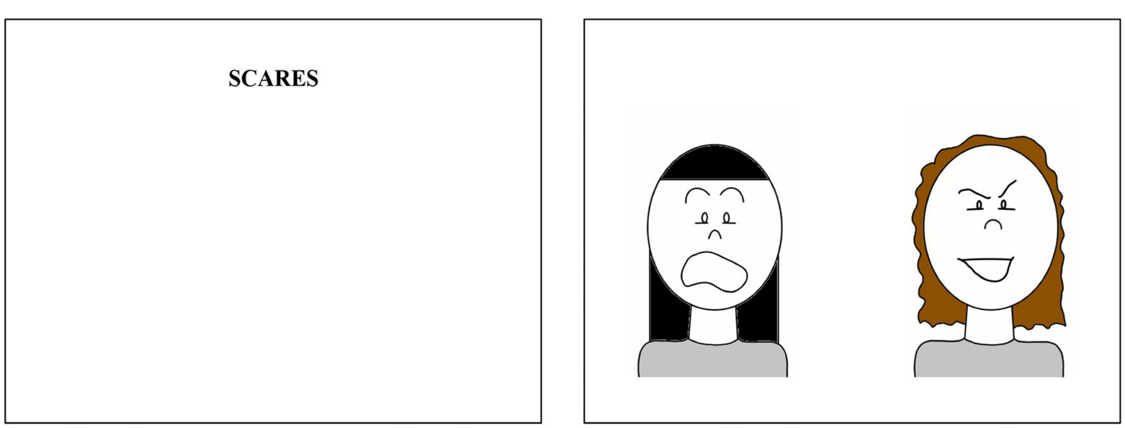

Figure 2. (a) Example ObjExp prompt verb and (b) Example to-be-described critical image corresponding to "*Leslie scares Ann*".

On filler trials, participants were asked to produce sentences using three different types of prompt words: (i) locative adjectives (e.g. *above*), which prompted participants to describe the location of one or more of the characters on the critical image screen; (ii) scalar adjectives (e.g. *faster*), which prompted participants to compare two or more of the characters on the critical image screen; or (iii) unergative verbs (e.g. *sneezes*).

Image Interpretability Questionnaire: A potential complication associated with using Psych verbs (e.g. SubjExp: "*fears*", ObjExp: "*scares*") is that the to-be-described events tend to be inherently more difficult to visually represent than the Agent-Patient events used in prior work (e.g. "*chase*", "*spray*"), because Psych verbs represent *internal* states. Thus, we anticipated that a possible concern was that participants might have difficulty interpreting the depicted event. To assess the extent to which participants' eye-movements might be linked to the interpretability of to-be-described images, participants were asked to complete an Image Interpretability questionnaire immediately after finishing the eye-tracking portion of the experiment. In this questionnaire, participants were shown all the target images that they had encountered during the study, and were instructed to indicate all images for which they were unsure "who did what to whom".

Emotiveness Questionnaire: Work outside the domain of language (e.g. Pritsch et al., 2017; Humphrey et al., 2012) has found that speakers "look more attentively and longer at emotionally salient facial expressions (…) compared to neutral faces" (Pritsch et al., 2017, p. 1). Given this, we were concerned that eye-movements in our data may be driven by the features of some facial expressions relative to others, rather than by processes unique to language production. Participants were therefore given a questionnaire in which they were asked to rate (on a 5- point scale, 1=Neutral and 5=Strongly Emotive) the emotiveness or expressiveness of each expression used in the study. This questionnaire was administered after the main eye-tracking experiment and Image Interpretability questionnaire.

2.1.3 Procedure

The study was conducted in a quiet room. Upon entering, participants were given instructions and then seated a comfortable distance away from the display monitor. Participants completed three practice trials. Then, they were trained and tested on the names of the four distinct characters (two male, two female). We opted for a relatively small number of distinct characters to minimise memory load. During this process, participants were first given the name and image of each character twice. They were then given character-name pairs and asked to decide (Y/N) whether the name matched the character shown on the screen. Finally, participants were trained on the characters' names one more time, and asked to name each character out loud. Participants were only able to continue to the main experiment once they had demonstrated that were able to name all the characters fluently and accurately.

Each trial consisted of two different screens (Figure 2): Participants first saw a prompt screen, which provided participants with the word they needed to use in their description (the critical verb, on target trials). Once they were ready to describe the image, participants pushed a button on a game controller to advance to the critical image screen, which depicted the two-character event. Thus, information about the event (i.e. the verb) and information about the entities involved in the event (i.e. the characters) were presented on separate screens. This was done to ensure that speakers could not begin formulating their sentences until after critical image onset, even if they knew (based on the verb) what type of event to expect.

Visual stimuli were presented on a 22-inch CRT monitor at a resolution of 1024 × 768 pixels (72ppi). Eye-gaze was recorded with an Eyelink II eye-tracker at a sampling rate of 500 Hz (SR Research) and eye-movements were calibrated using a 9-point procedure.

After participants had completed the eye-tracking portion of the experiment, they were given the post-experimental pen-and-paper questionnaires designed to help address concerns linked to the interpretability of the images and the visual salience of each individual characters' facial expressions.[3] In total, the entire session lasted 45–60 min.

2.2 Predictions

If producing ObjExp sentences *("Leslie$_{STIM}$ scares Ann$_{EXP}$.")*, where the *less* thematically prominent entity is the subject, is more difficult than producing SubjExp sentences *("Leslie$_{EXP}$ loves Ann$_{STIM}$.")*, where the more thematically prominent entity is the subject, then speakers should be comparatively slower to begin speaking in the ObjExp than in the SubjExp conditions. In addition, selecting and retrieving an entity to serve as the sentential subject may be more difficult when that entity is the less prominent Stimulus. If so, then we also expect fixations to the the entity that ultimately becomes the subject of the sentence to occur later in ObjExp conditions than in SubjExp conditions. Finally, Agt-Pat verbs were included in our study as a reference condition. Indeed, speech onset and eye-movement results from this condition were as expected given prior work (e.g. Griffin & Bock, 2000), so we do not discuss them further.

2.3 Results

Utterances were transcribed and coded by a native speaker of American English (the first author). We excluded trials which contained errors (explained below) and, to maximise comparability, also excluded cases in which participants produced fully grammatical utterances that did not conform to the target "X verbs Y" form. In total, this exclusion criteria affected roughly 28% of the data. Given the relatively high proportion of exclusions, it is useful to further discuss the types of errors produced in our study. First, in roughly 9% of utterances, speakers misnamed characters or used a non-target verb form (e.g. "scaring" instead of "scares"). This is comparable to error rates reported in other production work (Brown-Schmidt & Konopka, 2008; Do & Kaiser, 2019; Ganushchak et al., 2014, 2017). The remainder of the error-based exclusions (19%) involved "semantic role reversals" errors, in which participants saw a "Leslie scares Ann" image but produced a "Ann scares Leslie" sentence. Numerically, these types of errors occurred more frequently in the SubjExp conditions (7.9%) for verbs such as "*loves*", "*fears*", or "*admires*", than in the ObjExp (6.0%) and Agt-Pat (5.1%) conditions. However, linear mixed effect regression models revealed no statistically significant difference in the distribution of errors across these three conditions (p's >.1). Given that images for verbs of psychological state are, by nature, both difficult to depict and difficult to interpret, the relatively high proportion of "semantic role reversal" errors was not surprising. Indeed, this was a concern that we foresaw and aimed to address using the Image Interpretability questionnaire (discussed more below). Disfluencies (e.g. repairs, stutters, repetitions, pauses) were also excluded from speech onset and eye-movement analyses, affecting about 6% of the total data. This is comparable to rates of disfluencies seen in other production studies with ambiguous stimuli (e.g., Lee et al., 2013). Finally, we also excluded any outliers in speech onset times using the Mad-Median Rule (Wilcox, 2012), resulting in the exclusion of 8.37% of the remaining data.[4] Rates of exclusion varied between 1.5 - 3.4% by condition. (See Appendix B for additional details about the distribution of errors, disfluencies, and outliers by condition). In total, 482 utterances were submitted for analysis.

We first discuss results from speech onset times and eye-movements before moving to discussion of the post-experiment questionnaires which probed for potential differences in the interpretability of our images and the visual salience of characters' facial expressions.[5] The full set of data, code, and analyses for this and subsequent experiments are available through the first author's OSF (https://osf.io/bpa9r/).

2.3.1 Speech onset times

Speech onset latencies were computed as the time between the onset of the to-be-described image and the onset of the subject phrase. Onset times were measured in Praat (Boersma & Weenink, 2009) and analyzed using linear mixed effect regression models (Bates et al., 2015) in R. Given our predictions for speech onset times, contrasts were coded such that utterance onset times for the Agt-Pat verbs were first compared with the onset times for SubjExp and ObjExp verbs, combined. A second contrast set up our primary comparison of interest; this compared the onset times for the SubjExp and ObjExp verbs directly. Verb Type was included both as a fixed effect and as a random effect by-subjects and by-items.

We started out with the maximal model and reduced via model comparison so that the final model included only the random effects justified by the data (e.g., Bates et al., 2015; Matuschak et al., 2017). Model selection was done using a backwards selection schema wherein only those random effects which significantly improved the fit of the model ($p < .05$) were included.

Of interest for the research question at hand are the onset times for the ObjExp versus SubjExp verbs. Here, we predicted that speakers should have more difficulty planning sentences – i.e. we should see longer speech onset times – with ObjExp verbs than SubjExp. This is because the *more* prominent thematic element of ObjExp verbs, the Experiencer, is mapped onto the syntactically *less* prominent object position. In line with this prediction, we find that speech onset times in the ObjExp condition were significantly slower than in the SubjExp condition ($\beta = 300.40$, SE = 82.37, $|t| = 3.65$, $p < .01$) while speech onset times for Agt-Pat verbs did not differ from the other two verb types statistically ($\beta = -44.17$, SE = 70.61, $|t| = .63$, $p = .54$). In line with our predictions, then, evidence from speech onset times showed that speakers were significantly slower to begin speaking in the ObjExp condition than they were in the SubjExp condition.

One possibility, though, given some evidence that ObjExp verbs are less frequently produced in the active voice than SubjExp verbs (e.g. Ferreira, 1994; Wilson & Dillon, 2020; see also Engelberg, 2018 for similar results in German), is that speakers may have been slower to begin speaking in the ObjExp condition for reasons related to the comparatively lower frequency of active-voice ObjExp constructions.

To investigate the extent to which usage frequency for different verb types could account for the data in our studies, we used the COCA (the Corpus of Contemporary American English; Davies, 2008) to extract the active voice frequency (occurrences per million words) for each of the verbs used in our experiment. We then fitted a model with just active voice frequency as a fixed effect predictor for Exp 1a's speech onset times (random effect structure was determined as before) and compared this Frequency-Only model to the Verb-Type model reported above. Contra what is predicted from a frequency-based account, the results of the Frequency-Only model showed no significant effect of frequency ($\beta = -6.96$, SE = 5.61, $|t| = 1.24$, $p > .22$), indicating that active voice frequency was not a significant predictor of speech onset times. Moreover, comparison with the VerbType model using ANOVA yielded lower AIC, lower BIC, and larger logLikelihood scores for the VerbType model, indicating that the VerbType model provided a significantly better fit ($p < .01$) than the Frequency-Only model. Finally, we also fitted a separate model with both Frequency and VerbType as predictors. Here, too, we found only a significant effect of VerbType ($\beta = 317.536$, SE = 113.674, $|t| = 2.793$, $p < .05$) with no significant main or interaction effects related to Frequency (p's >.2). Taken together, these results make a purely frequency-driven account highly unlikely; they, instead, suggest that differences between the SubjExp and ObjExp conditions are better accounted for via the mapping between event roles and linguistic structures Table 3.

2.3.2 Eye-movements

Because we were interested in processes that occur as speakers are planning their utterances, we analyzed speaker's eye-movements during the 0–2000ms window after image onset. This window reflects the time from critical image onset to the time when speakers began uttering their sentences and is consistent with what has been used in other eye-tracking studies of language production to index processes associated sentence formulation (e.g. Do & Kaiser, 2019; Ganushchak et al., 2014, 2017; Griffin & Bock, 2000; Konopka, 2019; Konopka & Kuchinsky, 2015; Konopka & Meyer, 2014; Norcliffe et al., 2015; Sauppe et al., 2013).

We chose to analyze eye-movements during this window using Generalized Additive Mixed Models (GAMMs; Porretta et al., 2018; van Rij, 2015; Winter & Wieling, 2016) because they allowed us to (i) compare non-linear effects over time and (ii) account for the autocorrelation of residuals that occurs from sampling eye-movements at consecutive time points. This was done using the *mgcv* (Wood, 2017) and *itsadug* packages (van Rij et al., 2020) in R (version 3.6.2; R Core Team, 2016).

The proportion of looks to the syntactic subject during the 0–2000ms time window was aggregated into 20 ms bins. The response variable was the (binned) proportion of looks to the subject. To model the different trends over time for each Verb Type condition, Verb Type was treatment coded with the ObjExp condition set as the reference level and included as an ordered factor predictor. This allowed us to test for intercept differences (i.e. whether there were any significant differences in the proportion of looks to the subject at 0ms across conditions) and differences in the shapes of the curves over time (i.e. whether the smooths for each condition difference significantly from each other over time). In addition, we also included separate smooths for by-subject and by-item random effects; each trial was also given its own random intercept. After fitting this model, an appropriate autocorrelation parameter (AR1; ρ = .93), which captures the amount of autocorrelation introduced into the model residuals as a consequence of sampling eye-movements at successive time intervals, was selected using `acf_resid()`. Finally, model comparison was done using `compareML()` to test whether the AR1 parameter and each of our predictors significantly improved model fit. Predictors that did not improve model fit were omitted from the final model.

The model summary is reported in Table 4. The parametric coefficients show only a significant effect of intercept, meaning that at time 0ms (i.e. the onset of the critical image), the proportion of looks to the subject in the ObjExp reference condition differ significantly from zero. However, the proportion looks to the

Table 3. Mean Speech Onset Times, standard deviations, standard errors and 95% confidence interval for each verb condition in Experiment 1a.

Verb type	Mean onset times (ms)	SD (ms)	SE (ms)	95% CI (ms)
Agt-Pat	1818	431	31	(1756, 1880)
SubjExp	1739	457	35	(1671, 1807)
ObjExp	2029	457	37	(1957, 2102)

Table 4. Final model summary for Experiment 1a. The results of this model are presented visually in Figure 3.

	Parametric coefficients			
	Estimate	Standard Error	$|t|$-value	p-value
Intercept	0.52	0.03	15.43	<.001
OFVerbType2	−0.02	0.04	0.41	0.68
OFVerbType3	0.04	0.04	0.89	0.37
	Smooth Terms			
	Edf	Ref.df	F-value	p-value
s(Time)	7.39	8.22	18.65	<.001
s(Time): OFVerbTypeSubjExp	7.54	8.42	5.75	<.001
s(Time):OFVerbTypeAgtPat	6.65	7.78	1.74	0.06
s(Time, Subject)	43.74	338.00	0.35	<.001
s(Time, itemid)	35.24	234.00	0.42	<.001

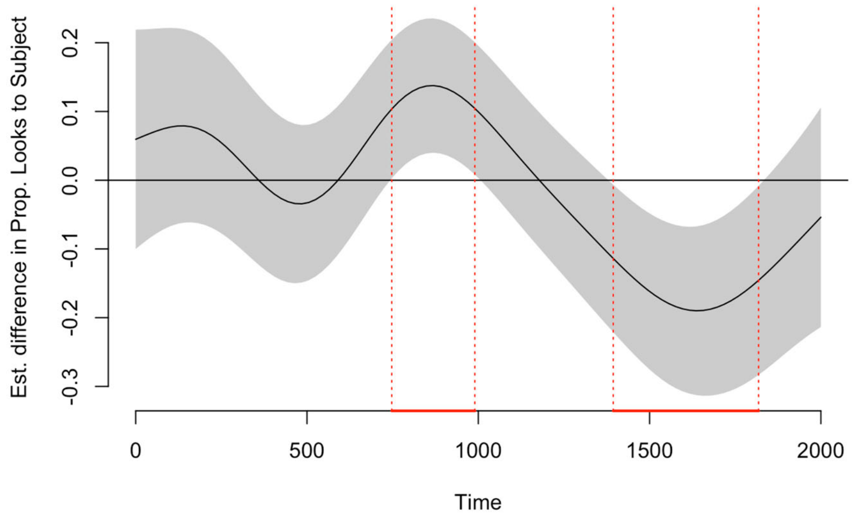

Figure 3. Difference plot with the mean proportion of looks to the subject in the SubjExp condition minus the mean proportion of looks in the ObjExp condition in Experiment 1a. Shading represents 95% CI. Positive regions indicate more looks to the subject in the SubjExp condition than in the ObjExp condition while negative regions indicate more looks to the subject in the ObjExp than SubjExp conditions. Red dotted lines indicate time window where eye-movement differences in the SubjExp versus ObjExp conditions significantly differed from zero.

subject in the SubjExp or Agt-Pat condition do not appear to differ significantly from those in the ObjExp condition at 0ms after image onset.

Turning our attention to the smooth terms, we see that the smooth for our ObjExp reference level differs significantly from zero, meaning that the smooth for this condition does, indeed, show a non-linear trend over time. Critically, we also see that the smooth indexing the *difference* between the smooths for SubjExp versus ObjExp conditions is also significantly different from zero, meaning that the SubjExp versus ObjExp conditions show different trends over time. Finally, a separate smooth of the difference between the Agt-Pat versus ObjExp conditions shows a marginally different trend over time for these two conditions.

While the model summary does allow us to conclude that the SubjExp versus ObjExp conditions differ over time, the summary cannot tell us when those differences occur or how the curves differ from each other. For this, we need to visually inspect the difference plot given in Figure 3 (Porretta et al., 2018; Winter & Wieling, 2016; Wood, 2017).

Specifically, Figure 3 plots the *difference* in the proportion of looks to the subject in the SubjExp versus ObjExp conditions, with 0ms corresponding to the onset of the critical image. A positive difference score indicates more looks to the subject in the SubjExp condition than in the ObjExp condition; a negative score indicates more looks to the subject in ObjExp than in SubjExp. Periods where the shading does not encompass the x-axis represent significant differences between the proportion of looks to the subject for these two conditions and are indicated by red dashed lines. As can be seen in Figure 3, there were significantly more looks to the subject in the SubjExp condition than in the ObjExp condition during the 750–1000ms time window (highlighted in red). By contrast, it is not until the 1400–1800ms window after image onset that, as as indicated by negative difference scores, we begin to see more looks to the subject in the ObjExp condition than in the SubjExp condition. These results are in line with our prediction that subjects should preferentially fixate the subject later in the "misaligned" ObjExp condition than in the SubjExp condition where syntactic and thematic prominence are aligned. Consistent with our predictions, then, the results of participants' eye-movements during real-time production of ObjExp and SubjExp sentences show an earlier preference for the subject character – i.e. the entity that would eventually become the subject – in the SubjExp condition, where syntactic and thematic prominence are aligned, than in the ObjExp condition, where the thematically less prominent Stimulus needs to be mapped on the syntactically prominent subject role.

This same pattern of results is reflected in Figure 4, which plots the proportion of looks to the subject in the SubjExp (dotted orange), ObjExp (solid blue), and Agt-Pat (dashed green) conditions. Here, we see that roughly 600ms after image onset, looks to the eventual subject in the SubjExp condition begin to climb, eventually reaching a peak at approximately 750ms after image onset. In contrast, looks to the eventual subject

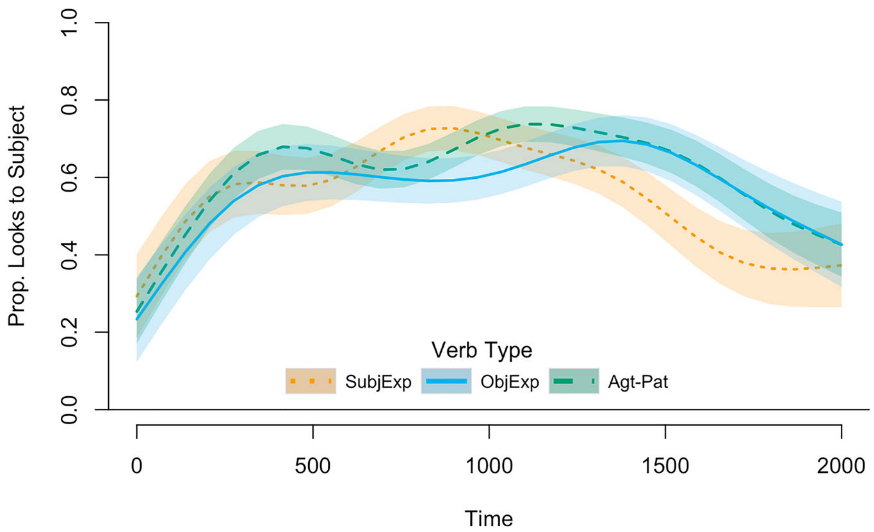

Figure 4. Proportion of looks to the subject from critical image onset (0 ms) to 2000ms after image onset in each Verb Type condition of Experiment 1a. Shading indicates ± 1 standard error.

in the ObjExp condition plateau around chance levels (0.5) until roughly 1000ms after image onset and do not peak until later – about 1400ms after image onset.

2.3.3 Image Interpretability questionnaire

As previously noted, we anticipated that the interpretability of some images used in our study might complicate the interpretation of our results. To help address this concern, participants were asked to complete an Image Interpretability questionnaire immediately after the eye-tracking portion of the experiment. Responses from this questionnaire were analyzed using logistic mixed effects regression models in R. Verb Type was included as a fixed effect and as a random effect in both the by-subject and by-items terms. We started with the maximal model, but only effects that contributed significantly to the model were included in the final model ($p < .05$).

Our results showed that the proportion of "unclear" images was significantly higher in the SubjExp condition than in the ObjExp condition ($\beta = 1.0598$, $SE = 0.1740$, $|z| = 6.090$; $p < .001$). In other words, there were more unclear images in the condition where syntactic and thematic prominence are *aligned* (SubjExp) than in the condition where syntactic and thematic prominence are *misaligned* (ObjExp). We return to this in the discussion.

2.3.4 Emotiveness questionnaire

Recall that we also wanted to address possible concerns that the eye-movement results might be linked to the visual properties of one character in our images over another. One concern, for instance, may be that speakers were slower to fixate the subject in the ObjExp condition because of visual properties associated with the object, the Experiencer (e.g. the Experiencer was more emotive and thus, more visually interesting). If so, results from our ObjExp condition would not necessarily speak to processes associated with sentence formulation. To assess this possibility, participants were also asked to complete a second questionnaire where they rated the emotiveness of each facial expression in the study. Responses from this 5-point rating questionnaire were first converted to z-scores; those z-scores were separately analyzed for each condition using linear mixed effect regression models in R. For interpretability, we report the uncoverted scores below.

In the critical ObjExp condition, there was no statistical difference in emotiveness ratings between the Stimulus ($\bar{x}$=4.05; SE=.06) versus the Experiencer ($\bar{x}$=3.66; SE=.06; $\beta = -0.35$, $SE = 0.32$, $|t| = 1.09$, $p > .3$). Likewise, in the SubjExp condition, we found no evidence of difference between ratings for the Experiencer ($\bar{x}$=3.35; SE=.06) versus the Stimulus ($\bar{x}$=3.66; SE=.06; $\beta = 0.29$, $SE = 0.34$, $|t| = 0.86$, $p > .4$). Thus, in the critical conditions of comparison, we did not find evidence from the Emotiveness Questionnaire to suggest that there were systematic imbalances associated with the visual properties of our images that could have contributed to the eye-movement results observed in our language production task. Although the Agt-Pat condition was not central to our predictions, we also tested for differences in the emotiveness of of the Agent ($\bar{x}$=3.88; SE=.06) versus Patients ($\bar{x}$=3.16; SE=.06) characters. Here, Agents were found to be reliably more emotive than Patients ($\beta = -0.71$, $SE = 0.07$, $|t| = 10.48$, $p < .001$).

2.4 Discussion

In Experiment 1a, we were interested in speech onset times and eye-movements as speakers produced sentences with two different classes of verbs – SubjExp verbs, where the more thematically prominent Experiencer can be assigned to the subject role, and ObjExp verbs, where the more thematically prominent Experiencer must be assigned to the object role. We found that speakers are slower to begin speaking (Section 2.3.1) and also slower to preferentially fixate the subject character (Section 2.3.2) in the ObjExp conditions than in SubjExp conditions. These results thus lend support to the possibility that speakers have greater difficulty producing active voice ObjExp sentences (where syntactic and thematic prominence are misaligned) than they do producing active voice SubjExp sentences (where syntactic and thematic prominence are aligned).

In addition, the results of our post-experimental Image Interpretability and Emotiveness Questionnaires do not provide any clear evidence for an alternative account that attributes our finding to visual artefacts in the stimuli. In particular, if our speech onset and eye-movement results were driven by difficulty interpreting the images in the ObjExp condition, then we might expect a higher proportion of "unclear" images for the ObjExp conditions in the Image Interpretability questionnaire. As reported in Section 2.2.3, this turned out not to be the case; we found the opposite pattern.[6] Likewise, if speakers in our study were quicker to select the subject of SubjExp sentences simply because they disproportionately attended to the facial expression of the Experiencer character, then we would expect to find higher emotiveness ratings for Experiencers in that condition. However, the results of the Emotiveness Questionnaire (Section 2.2.4) did not provide evidence to that effect. Nevertheless, given that judgements from the Emotiveness questionnaires were collected after participants had finished producing their sentences, it is possible that they do not accurately reflect the sub-conscious eye-movements that occurred during the experiment itself. To better assess the extent to which the patterns observed in Experiment 1a were driven by the mismatch between syntactic and thematic prominence, rather than an artefact of visual properties of our images, we conducted Experiment 1b.

3. Experiment 1b: testing the role of emotiveness

Experiment 1b was a designed to see whether the pattern of results reported in Experiment 1a was primarily driven by difficulty associated with processes specific to sentence planning or by non-linguistic factors related to the emotional expressions depicted in our stimuli. On this alternative account of the data, delays in speech onset times and the timing of preferential fixations to the subject character in the ObjExp conditions in Experiment 1a may have been due to visual properties associated with the emotiveness of the Stimulus characters in the ObjExp condition: they may have been less visually salient (e.g. Pritsch, 2017) or the emotions depicted were less interpretable or more difficult to detect (e.g. Becker et al., 2011; Fox et al. 2000; Savage et al., 2013). If this is the case – if the results of Experiment 1a were fundamentally driven by properties associated with the way that our Stimulus versus Experiencer characters were visually depicted, rather than by processes associated with linguistic encoding – then it might be reasonable to expect that the same pattern of eye-movements found in Experiment 1a should surface (perhaps even more so) when participants are not engaged in any linguistic task. In particular, we should likewise find a distinct pattern of eye-movements in the ObjExp versus SubjExp condition. To that end, we analyzed the eye-movements of a new group of participants as they completed the non-linguistic Picture Inspection Task (e.g. Griffin & Bock, 2000) used in Experiment 1b.

3.1 Methods

3.1.1 Participants

18 native speakers of American English who did not participate in Experiment 1a were recruited from the University of Southern California. They were given course credit for their participation.

3.1.2 Materials and design

The exact same images, presentation order and list structure of Experiment 1a was used for Experiment 1b. [7] However, because there was no spoken component in Experiment 1b, the word prompts used in Experiment 1a were replaced with a simple fixation crosses located at the top centre of the screen, in the same location as the verb in Experiment 1a.

3.1.3 Procedure

Participants in Experiment 1b were instructed to carefully examine each image to get a sense of the "content and characteristics" of each image. To ensure that participants would attend to the images, they were told that while they did not need to memorise every image, they would be asked questions about the images at random points throughout the study. Because we did not want

Table 5. Final model summary for Experiment 1b. The results of this model are presented visually in Figure 5.

	Parametric Coefficients			
	Estimate	Standard Error	$\|t\|$-value	*p*-value
Intercept	0.83	0.41	2.01	<.05
OFVerbType2	−0.36	0.35	1.05	0.30
OFVerbType3	−0.04	0.04	1.01	0.49
	Smooth Terms			
	Edf	Ref.df	*F*-value	*p*-value
s(Time)	5.71	6.97	4.41	<.001
s(Time):OFVerbTypeSubjExp	1.01	1.01	0.003	0.96
s(Time):OFVerbTypeAgtPat	1.00	1.00	0.15	0.70
S(Event)	110.16	393.00	0.41	<.001
s(Time, Subject)	130.00	178.00	4.67	<.001
s(Time, Item)	175.37	234.00	4.83	<.001

to explicitly focus participants on any particular aspect of the stimuli, these questions asked participants to rate different aesthetic characteristics (e.g. intricateness, naturalness, ugliness) of the images on a scale of 1 = "Not at all" to 5 = "Very". Following the eye-tracking portion of the study, participants were asked to complete the Emotiveness Rating questionnaire. The experimental session lasted roughly 30 min.

3.2 Results

3.2.1 Eye-movements

To maximise comparability with Experiment 1a, our statistical analyses were also conducted over the 0–2000 ms window after image onset. Analyses were again conducted using GAMMs with fixed effects and random effects specified as in Experiment 1a. Although no verbs were actually provided to participants (prompt words were replaced by fixation crosses), we treated the items and conditions in Experiment 1b the same way as in Experiment 1a. That is, images that were analyzed in the SubjExp conditions of Experiment 1a were also included in SubjExp conditions of Experiment 1b; the same was true for the images in the ObjExp and Agt-Pat conditions.

Thus, the ObjExp Verb Type was again specified in the treatment contrasts to be our reference level. An AR1 autocorrelation value of $\rho = .92$ was selected using `acf_resid()` and model specification and selection was done as before. The results of the model summary are given in Table 5.

Here, the parametric coefficients indicate that the intercept for the ObjExp Verb Type – i.e. the proportion of looks to the character that would have been the subject in Experiment 1a at the onset of the critical image – was significantly different from 0, but did not differ significantly from either of the other two Verb Type conditions. Again turning our attention to the critical smooth terms, we see that the presence of a statistically significant smooth over time shows that the smooth for the ObjExp reference level in Experiment 1b, like Experiment 1a, shows significant departures from linearity. However, contrary to what is predicted by the emotiveness account and unlike what was seen in Experiment 1a, we find no significant differences between the ObjExp versus SubjExp or the ObjExp versus Agt-Pat conditions. In other words, there was no evidence to suggest that eye-movements in the ObjExp, SubjExp, or Agt-Pat conditions differed over time.

This is most evident in Figure 5, which again shows a *difference plot* of the proportion of looks to (i) the character that would have been the subject in the ObjExp condition of Experiment 1a versus (ii) the character that

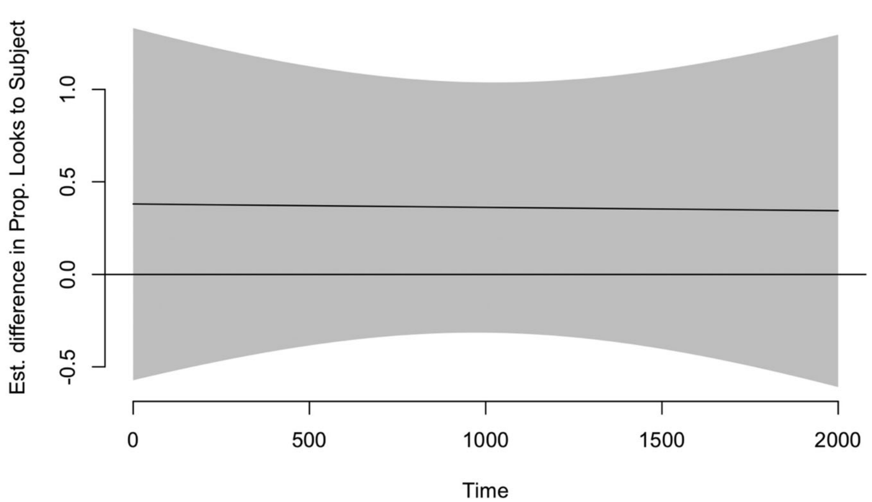

Figure 5. Difference plot from Experiment 1b, a non-linguistic inspection task, showing the mean proportion of looks to the character that *would have been the subject* in the SubjExp condition of Experiment 1a minus the mean proportion of looks in the character that *would have been the subject* in the ObjExp condition of Experiment 1a. Shaded areas showing the 95% CI suggest no significant differences between the SubjExp and ObjExp conditions in Experiment 1b.

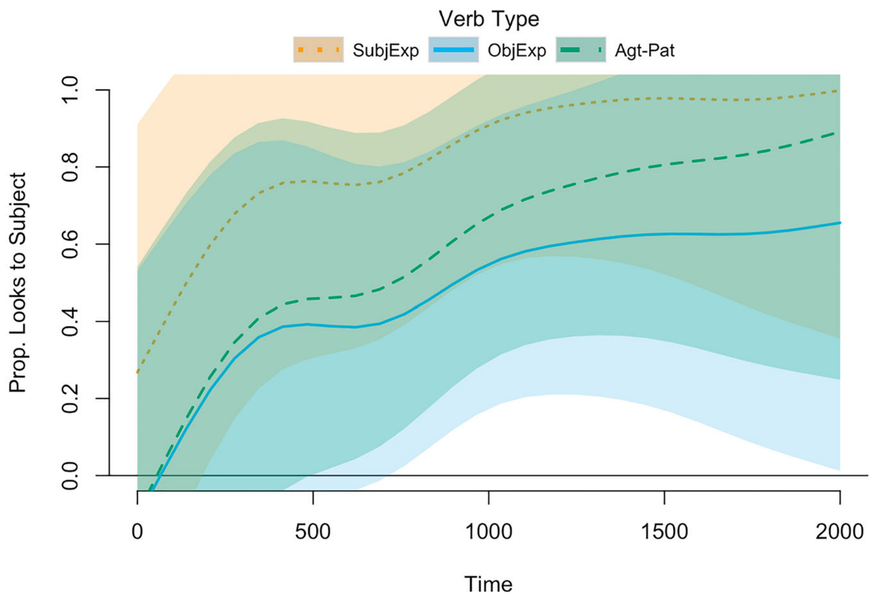

Figure 6. Proportion of looks to the would-be subject character pf each Verb Type condition in Experiment 1b, measured from critical image onset (0 ms) to roughly 2500 ms after image onset. Shading indicates ± 1 standard error.

would have been the subject in the SubjExp condition of Experiment 1a over time. In particular, the gray 95% confidence interval band encompasses zero across all times points, indicating that the proportion of looks to the would-be subject character of the ObjExp versus SubjExp conditions does not differ significantly during any point in the first 2000ms after image onset.

The results of our model are also consistent with what is shown in Figure 6, which plots the proportion of looks to the character that would have been the subject in Experiment 1a. Although the open-ended nature of the picture-inspection task means that the time-course of these eye-movements is difficult to interpret, the key point to note is that participants' eye-movements show a similar trajectory across the SubjExp passive (dotted orange), ObjExp passive (solid blue), and Agt-Pat passive (dashed green) conditions.

In other words, the results of Experiment 1b (with no language production task) show *no* evidence of a significant difference across any conditions and thus stand in stark contrast to the results of Experiment 1a (with a language production task), where we *did* find significant differences in speakers' eye-movements across the SubjExp and ObjExp conditions specifically during the time windows known to index processes associated with sentence formulation (Griffin & Bock, 2000).

3.2.1 Emotiveness

Participants in Experiments 1b were also asked to rate the emotiveness of each individual expression in the study. Those ratings were z-scored and analyzed using linear mixed regression models as before.

As in Experiment 1a, we found no reliable evidence that the Stimulus ($\bar{x} = 4.23$, SE = 0.08) was more emotive than the Experiencer ($\bar{x} = 3.90$, SE = 0.08) in the critical ObjExp condition ($\beta = -0.33$, SE = 0.31, $|t| = 1.06$, $p > 0.3$). Likewise, there was no difference in the emotiveness of the Experiencer ($\bar{x} = 3.40$, SE = 0.08) versus the Stimulus ($\bar{x} = 3.84$, SE = 0.08) in the SubjExp condition ($\beta = 0.3877$, SE = 0.3376, $|t| = 1.15$, $p > .2$). In the Agt-Pat condition, Agents ($\bar{x} = 3.95$, SE = 0.09) were again rated more emotive than Patients ($\bar{x} = 3.31$, SE = 0.09); this difference was significant ($\beta = -0.64$, SE = 0.10, $|t| = 6.26$, $p < .001$), echoing the results of Experiment 1a.

3.3 Discussion

Given the nature of the stimuli used in Experiment 1a, one concern was the possibility that our results (i.e. later fixations to the subject character in the ObjExp conditions compared to the SubjExp conditions) could be driven by visual properties associated with the emotiveness of the characters, rather than by processes associated explicitly with sentence planning. We reasoned that if this kind of non-linguistic emotiveness account is right – if eye-movements in Experiment 1a were ultimately driven by general differences in ways that the characters' various expressions were encoded (e.g. Fox et al., 2000; Pritsch, 2017; Humphrey et al., 2012) – then those same factors should also play a role in Experiment 1b, with no language production task. In fact, the effect of these visual factors should emerge even more strongly in the non-linguistic picture-inspection task used in Experiment 1b, where they cannot be masked by processes associated with language.

Instead, contrary to what is predicted by the emotiveness account, the differences previously observed in Experiment 1a, where language was involved, disappeared entirely when processes associated with linguistically encoding a sentence were removed from the picture. Our results showed that participants' eye-movements were largely similar across conditions in Experiment 1b; we found no differences between participants' eye-movements in the critical SubjExp versus ObjExp conditions. This pattern also matches what we saw in Experiment 1b's (and Experiment 1a's) offline emotiveness measures – namely, that there was no reliable evidence of differences in the emotiveness of the Stimulus versus Experiencer arguments in either the SubjExp or ObjExp conditions.

So, while there may inevitably be *some* differences in the visual processing of the characters in the images, the findings of Experiment 1b suggest that non-linguistic factors associated purely with the processing of the emotional expressions depicted on our characters cannot account for the whole of our data in Experiment 1a; there are still linguistically-driven effects in Experiment 1a over and above any effects stemming from the emotiveness of the characters alone. Taken in conjunction with the results of our Emotiveness Rating questionnaire, which also showed no statistically significant differences between the emotional salience of characters in either the SubjExp or ObjExp conditions, the slower speech onset times and later fixations to the subject character during real-time production of ObjExp sentences were more likely driven by the challenge of having to map the less prominent Stimulus role onto the prominent subject position (and relegate the more prominent Experiencer to the less prominent object position) than by non-linguistic factors related purely to the emotiveness of our stimuli.

4. Experiment 2

To conceptually replicate and further test the hypothesis that sentence formulation is easier when speakers are able to map the most thematically prominent event role onto the syntactically prominent subject position, we asked participants in Experiment 2 to produce passivized ObjExp ("*Leslie$_{EXP}$ was scared/surprised/annoyed by Ann$_{STIM}$.*") and SubjExp ("*Leslie$_{STIM}$ was loved/feared/admired by Ann$_{EXP}$.*") sentences. In particular, if the ease of sentence formulation depends on the alignment of thematic to syntactic prominence, then the pattern of results shown in Experiment 1a should be reversed when speakers are asked to produce passivized sentences. This is because passivized ObjExp sentences allow the more prominent Experiencer to be realised as the subject – thereby allowing the thematically prominent element to occur in the syntactically prominent position, eliminating any misalignment. By contrast, in passivized SubjExp sentences, the less prominent Stimulus is what appears as the syntactic subject – in effect yielding a misalignment between thematic and syntactic prominence. So, whereas active ObjExp sentences should be more difficult to produce than active SubjExp sentences (Experiment 1a), passivized ObjExp sentences in Experiment 2 should be *easier* to produce than passivized SubjExp sentences. This experiment was pre-registered (https://osf.io/fng4m).

4.1 Methods

4.1.1 Participants

Thirty-two adult native speakers of American English who did not participate in either Experiment 1a or 1b were recruited from the University of Pennsylvania and given course credit for their participation.

4.1.2 Materials and design

The design and materials used in Experiment 2 were the same as in Experiment 1a with the following exceptions. First, images for two items were altered slightly because the Image Interpretability questionnaire indicated that these were especially difficult for participants to interpret. In addition, we also replaced "*worries*" with "*frightens*" because "*worries*" tended to yield a particularly high number of non-target structures in Experiment 1a (e.g. "*Mary worries about Ann.*", "*Mary worries.*")

Second, recall that prompt verbs in Experiment 1a were always shown in present tense to explicitly prevent passivization. In Experiment 2, we were specifically interested in the production of passives; verbs were thus presented in the WAS + PAST TENSE form (*was loved*, *was scared*, *was criticized*, etc.)

4.1.3 Procedure

The procedure and instructions in Experiment 2 were identical to Experiment 1a. That is, participants were simply instructed to create a sentence that used the prompt word in the exact form shown. They were free to produce any sentence they wished so long as they used the verb form shown on the screen (now adjusted to elicit passives). Given that the Emotiveness Rating questionnaires in Experiment 1a and 1b showed no difference between the emotiveness of the characters in the critical SubjExp and ObjExp conditions, participants in Experiment 2 were only asked to complete the Image Interpretability questionnaire.

Visual stimuli were presented on a 21.5-inch display at a resolution of 1920 × 1080 pixels and refresh rate of 60 Hz. Eye-gaze was recorded with an Eyelink 1000 eye-tracker at a sampling rate of 1000 Hz (SR Research). Participants were placed into the head support roughly 20 in. from the display monitor with just their foreheads touching the head rest. Eye-movements were calibrated using a 9-point procedure. The experimental session lasted roughly 30 min.

4.2 Predictions

In Experiment 2 we generally expected a reversal of the pattern seen in Experiment 1a: Participants should be significantly faster to begin speaking and they should preferentially fixate the subject character earliers when producing passivized ObjExp sentences than when producing passivized SubjExp sentences.

4.3 Results

Utterances were transcribed and coded by the first author. Using the same coding and exclusion criteria as in Experiment 1a, sentences where speakers used the wrong name, the wrong verb, or added extra words were excluded from analysis, affecting roughly 3% of the data. For the purposes of Experiment 2, short form passives (*"Leslie was loved."*), where the object in the by-phrase was omitted, were coded as errors.[8] This occurred predominantly in the ObjExp conditions, though the reasons for this are unclear. The remaining errors were "semantic role reversal" errors, which occurred in roughly 5% of Agt-Pat utterances, 8% of SubjExp utterances, and 3% of ObjExp utterances – approximately 16% of the data in total. Utterances which contained disfluences, which made up roughly 9% of utterances, were also excluded. Finally, outliers in the speech onsets times were detected using the Mad-Median Rule (Wilcox, 2012), affecting approximately 3% of utterances in the Agt-Pat condition, 4% in the SubjExp condition, and 2% in the ObjExp condition. (See Appendix B for additional details by condition.) In total, data from 420 passive sentences were submitted for analysis.

4.3.1 Speech onset times

Data from speech onset times were analyzed as before. We predicted that when asked to produce passives, the pattern observed in Experiment 1a should be reversed. Namely, speakers would be comparatively faster to begin producing passivized ObjExp sentences compared to passivized SubjExp sentences. This is because in the case of passivization, it is ObjExp verbs, not SubjExp verbs that exhibit the alignment between thematic and syntactic prominence. This predition was borne out (Table 6): There were no significant difference between the passivized Agt-Pat and the other two conditions ($\beta = -21.05$; SE = 86.09; $|t| = .24$; $p > .8$), but speech onset times were significantly faster in the ObjExp condition than in the SubjExp condition ($\beta = -375.87$, SE = 99.21, $|t| = 3.79$, $p < .01$).

Table 6. Mean Speech Onset Times, standard deviations, standard errors, and 95% CIs for in each condition in Experiment 2.

Verb type	Mean Onset Times (ms)	SD (ms)	SE (ms)	95% CI (ms)
Agt-Pat	2234	628	52	(2131, 2337)
SubjExp	2426	638	55	(2317, 2535)
ObjExp	2073	569	48	(1978, 2168)

As in Experiment 1a, a possible concern is that the difference in speech onset times between the SubjExp and ObjExp conditions could be due to differences in construction frequencies – that is, speakers in Experiment 2 may have been slower to begin speaking in the passive SubjExp condition simply because SubjExp verbs occur less frequently in the passive form than ObjExp verbs (e.g. Engelberg, 2018; Ferreira, 1994; Wilson & Dillon, 2020). To address this, we followed a similar approach as for Experiment 1a, but now focusing on the passive forms of the verbs: We extracted the passive voice frequency for each verb in Experiment 2 from COCA and compared a Frequency-Only model to the Verb Type model reported above.[9]

Here, as in Experiment 1a, the Frequency-Only model found no significant effect of word frequency ($\beta = -842.9$, SE = 577.2, $|t| = 1.46$, $p > .16$). In addition, AIC, BIC, and logLikelihood scores all suggest that the VerbType model provided a significantly better fit ($p < .01$) for our speech onset data. Fitting a separate model with both Frequency and VerbType as predictors revealed only a significant effect of VerbType ($\beta = -717.19$, SE = 214.34, $|t| = 3.35$, $p < .01$) with no significant main or interaction effects related to Frequency (p's >.08). Once again, then, we failed to detect any significant contribution of frequency in our speech onset data.

4.3.1.1 Experiment 1a vs experiment 2 comparison.

In addition, to directly compare the speech onset times from Experiment 1a to Experiment 2, we performed a separate analysis with both Verb Type and Experiment (Exp1a: Active vs Exp2: Passive) included as fixed effect factors. Verb Type was also included as a random effect in the by-subject and by-item random effects; Experiment was only included as a by-item random effect as this variable was manipulated between subjects. Consistent with

Table 7. Final model summary for Experiment 2. The results of this model are presented visually in Figure 7.

	Parametric Coefficients			
	Estimate	Standard Error	|*t*|-value	*p*-value
Intercept	0.45	0.04	10.66	<.001
OFVerbType2	−0.02	0.02	0.91	0.36
OFVerbType3	−0.02	0.02	0.71	0.48
	Smooth Terms			
	Edf	Ref.df	*F*-value	*p*-value
s(Time)	7.89	8.66	31.73	<.001
s(Time):OFVerbTypeSubjExp	6.80	7.99	5.69	<.001
s(Time):OFVerbTypeAgtPat	1.00	1.00	0.10	0.75
s(Time, Subject)	62.72	298.00	1.58	<.001
s(Time, Item)	15.90	234.00	0.11	<.01

the fact that speakers were generally slower to begin speaking in the context of passives (Exp 2), we found a significant main effect of Experiment ($\beta = 403.43$; SE = 76.97; $|t| = 5.24$; $p < .01$). Critically, in line with our predictions, we also detected a significant Verb Type x Experiment interaction specifically in the SubjExp vs ObjExp contrast ($\beta = -648.37$; SE = 73.59; $|t| = 8.81$; $p < .01$). This analysis thus confirms that differences between Experiment 1a and Experiment 2 are due, specifically, to speakers being significantly slower to begin speaking in ObjExp conditions than in SubjExp conditions of Experiment 1a, but faster to begin speaking in ObjExp conditions than in SubjExp conditions of Experiment 2.

4.3.2 Eye-movements

In Experiment 2, we analyzed eye-movements in the 0–2500 ms window after critical image onset; this was a larger window of analysis than in Experiment 1a because speakers were generally slower to begin speaking in Experiment 2. Data from eye-movements were analyzed using GAMMs with model specification and comparison done as before. Verb Type in Experiment 2 was again treatment coded with the passive ObjExp Verb Type set as the reference condition. An AR1 autocorrelation value of $\rho = 0.93$ was selected using `acf_resid()`. The model summary is reported in Table 7.

The parametric coefficients suggest that at critical image onset, the proportion of looks to the subject in the ObjExp condition did differ significantly from zero. However, there were no significant differences in the proportion of looks to the subject in either the SubjExp or Agt-Pat conditions, compared to the baseline ObjExp condition. In addition, the smooth terms again point to two difference smooths that differ significantly from zero. The first difference smooth, which refers to the smooth for the ObjExp reference condition, suggests that eye-movements showed a non-linear trend over time. Crucially, the model summary shows that a second difference smooth – namely, the smooth of the difference between the passivized ObjExp and passivized SubjExp conditions – also differs significantly from 0, suggesting that the eye-movement curves for these two conditions took significantly different trends over time. By contrast, the difference smooth for the passivized ObjExp versus passivized Agt-Pat curves did not appear to show significant differences over time.

As in Experiment 1a, while the model summary for Experiment 2 does show that the passivized ObjExp and passivized SubjExp curves differ, understanding the time periods when conditions differed must be done via visual inspection. As such, a *difference plot* of the proportion of looks to the subject character in the passivized ObjExp minus passivized SubjExp conditions is given in Figure 7. Unlike in Experiment 1a, where the difference plot represented active SubjExp minus active ObjExp, a positive difference score here indicates more looks to the subject in the passivized ObjExp condition compared to the passivized SubjExp condition while a negative score indicates more looks to the subject in the passivized SubjExp condition than the passivized ObjExp condition.

In Experiment 2, where participants were asked to produce passivized ObjExp and passivized SubjExp sentences, we predicted a pattern opposite the one seen in Experiment 1a: Unlike in Experiment 1a, where speakers preferentially fixated the subject of active SubjExp sentences earlier than in active ObjExp sentences, speakers in Experiment 2 should preferentially fixate the subject character earlier in passivized ObjExp sentences than in passivized SubjExp sentences. Figure 7 shows this to be the case: In contrast to Experiment 1a, where we found a clear preference for the subject in the *SubjExp* condition emerge during the 750–1050 ms window after image onset, Experiment 2 showed that speakers preferentially fixated the subject character of *ObjExp* conditions during a very similar (e.g. 700–1400 ms after image onset) time window (positive difference scores highlighted in red). Indeed, it is not until roughly 2000ms after image that we begin to see more looks to the subject (i.e. negative difference scores) emerging in the SubjExp passives than in the ObjExp passives, though these do not reach statistical significance during this time window.

This same pattern of results is also visible in Figure 8, which shows the proportion of looks to the subject character in the SubjExp passive (dotted orange), ObjExp passive (solid blue), and Agt-Pat passive (dashed green) conditions. Here we see that the proportion of looks to the subject reaches a peak of about 0.7 in the ObjExp passive condition roughly 1000 ms after image

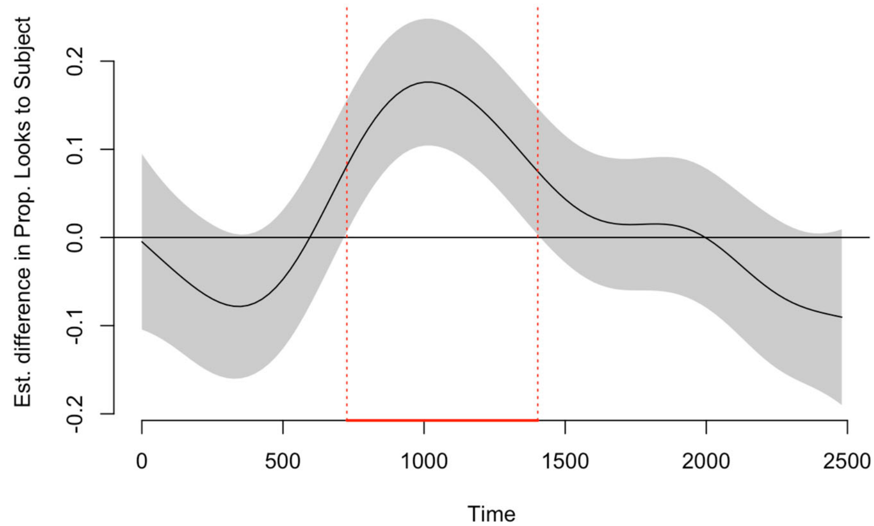

Figure 7. Difference plot with the mean proportion of looks to the subject in the ObjExp condition minus the mean proportion of looks in the SubjExp condition of Experiment 2. Shading represents 95% CI. Positive regions indicate more looks to the subject in the ObjExp condition than in the SubjExp condition while negative regions indicate more looks to the subject in the SubjExp than ObjExp conditions. Red dotted lines indicate time window where eye-movement differences in the ObjExp versus SubjExp conditions significantly differed from zero.

onset, while the proportion of looks to the subject character in the SubjExp passive condition hovers at roughly 0.5 during the entire 0–2500 ms time period. As predicted, this is again different from what we saw in Experiment 1a, where looks to the subject entity peak earlier in the active SubjExp condition than in the active ObjExp condition.

4.4 Discussion

In Experiment 2, speakers were asked to use passivized ObjExp (*"Leslie$_{EXP}$ was surprised/scared/annoyed by Ann$_{STIM}$."*) and SubjExp (*"Leslie$_{STIM}$ was loved/feared/admired by Ann$_{EXP}$."*) sentences to describe the same images as in Experiment 1. We reasoned that if the mapping between thematic and syntactic prominence could account for the differences in planning observed in Experiment 1a for the two verb classes, then we should be able to reverse that pattern by asking speakers to produce passivized sentences. Specifically, passivized ObjExp verbs should now be comparably *easier* to produce than passivized SubjExp sentences, because the passivized ObjExp structure allows the more thematically prominent Experiencer role to appear in the subject position, while the passivized SubjExp structure does not.

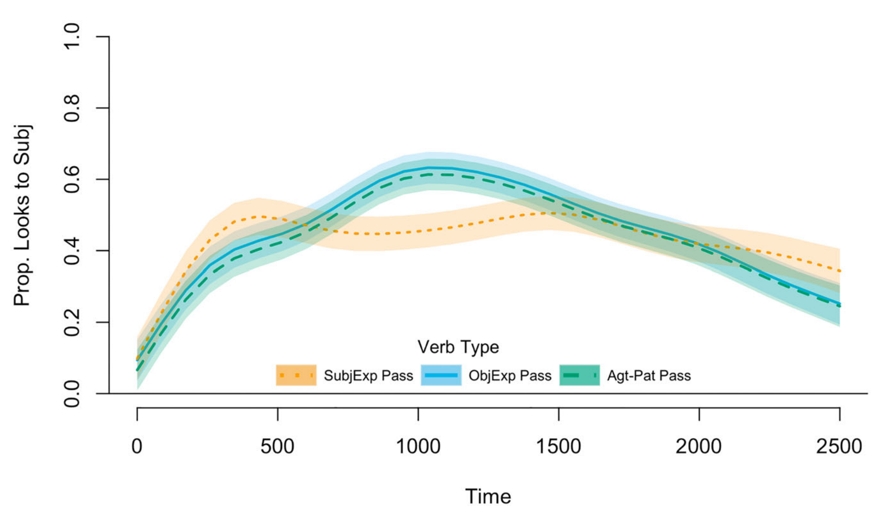

Figure 8. Proportion of looks to the subject from critical image onset (0 ms) to roughly 2500 ms after image onset in each Verb Type condition of Experiment 2. Shading indicates ± 1 standard error.

This prediction was confirmed in the speech onset times: Speakers in Experiment 2 were *faster* to begin speaking in the passivized ObjExp condition than in the passivized SubjExp condition. This is the opposite of what we observed in Experiment 1a. Likewise, the eye-movement data in Experiment 2 also showed a reversal of the pattern that we saw in Experiment 1a: Specifically, our results showed a clear preference to fixate the subject character in the passivized ObjExp conditions starting at roughly 700 ms after image onset. By contrast, looks to the subject character in the passivized SubjExp condition during this same time window remained at chance levels suggesting that in this latter condition, speakers were looking to the subject and object characters to a similar extent. In other words, production of a passivized ObjExp sentence elicited earlier fixations to the thematically prominent Experiencer, which is realised in the syntactically prominent subject position, whereas production of a passivized SubjExp sentence elicited competition between (i) the more thematically prominent Experiencer, which is now relegated to a position within the *by*-phrase versus (ii) the thematically less prominent Stimulus, which occurs in the syntactically prominent subject position. These different patterns of results across conditions held even though speakers were talking about similar types of events and producing the sentences with the same number of words and surface structure.

5. General discussion

In order to formulate a sentence speakers have to, among other things, map their conceptual representation of the event they are talking about onto a linguistic representation. To investigate how speakers navigate this transition from conceptual to linguistic structure, which occurs well before speakers actually begin uttering their sentences, we tracked speakers' speech onset times and eye-movements as they were preparing to produce active SubjExp sentences such as "*Leslie$_{EXP}$ loves Ann$_{STIM}$.*" versus active ObjExp sentences such as "*Leslie$_{STIM}$ scares Ann$_{EXP}$.*" (Experiment 1a) and as they were preparing to produce passive SubjExp sentences like "*Leslie$_{EXP}$ was scared by Ann$_{STIM}$.*" versus passive ObjExp sentences like "*Leslie$_{STIM}$ was loved by Ann$_{EXP}$.*" (Experiment 2). We wanted to see how sentence formulation unfolds when the the most prominent thematic role in the event structure, the Experiencer, either could or could not be assigned to the most prominent syntactic position in the sentence – i.e. the subject. Although the experiments here focused on SubjExp and ObjExp sentence types because they allowed us to investigate production in the context of an (at least superficially) minimal pair, the implications of our study extend to sentence formulation processes more generally.

Specifically, while prior work on these minimal pairs has shown that speakers typically map the most thematically prominent entity (i.e. the more highly ranked entity in the Thematic Hierarchy; Ferreira, 1994; Grimshaw, 1980; Jackendoff, 1972, 1987) onto the subject position, those studies were not able to point to a tangible cost to language production when speakers were unable to accomplish this mapping, at least among non-aphasic speakers (Thompson & Lee, 2009). As a consequence, the underlying question of *why* this tendency to align the most thematically and syntactic prominent entities exists among language producers remains an open question. Here, we used a fine-grained measure and found that there was, indeed, a cost associated with the inability to directly map the most thematically prominent role onto the subject position.

In Experiment 1a we found that speakers were slower to begin speaking in active ObjExp conditions compared to active SubjExp conditions. Moreover, consistent with what has been identified by prior work as the time window typically associated with sentence formulation (e.g. Do & Kaiser, 2019; Ganushchak et al., 2014, 2017; Griffin & Bock, 2000; Konopka & Meyer, 2014), we also found that speakers were significantly slower to preferentially fixate the the subject character in the active ObjExp conditions than in the active SubjExp condition during the 750–1000ms time window after image onset. But, in Experiment 2, when speakers were asked, by contrast, to produce passivized SubjExp and passivized ObjExp sentences, the pattern seen in Experiment 1a was reversed – even though speakers were asked to describe the same set of images and types of events as in Experiment 1a. Specifically, Experiment 2 found that speakers were significantly *faster* to begin speaking and fixated the subject character *earlier* when producing passivized ObjExp sentences than when producing passivized SubjExp sentences.

In other words, our results show that language production is more difficult in precisely those situations where speakers are unable to map the most thematically prominent element of the event onto the subject position, namely in active ObjExp sentences (Experiment 1a) and passive SubjExp sentences (Experiment 2). Indeed, our findings go beyond prior work on the production of SubjExp and ObjExp sentences (e.g. Ferreira, 1994; Thompson & Lee, 2009) by demonstrating that there is an observable real-time cost – both in terms of speed and ease of production – associated with violating the preference to map the most thematically prominent role onto the syntactically prominent subject position. In doing so, our results point to a possible motivation underlying the preference – across languages and different types of events – to map more prominent entities in the Thematic Hierarchy to the role of the grammatical subject: In particular, they suggest that the preference to align thematic and syntactic prominence may be a consequence of speakers taking the "path of least resistance" during sentence formulation, as they turn their conceptual representations of an event into a linguistic form.

In addition, the reversal in both speech onset times and eye-movements found in Experiment 1a versus Experiment 2 suggests that while there may certainly be other differences between SubjExp and ObjExp verbs – for instance, differences in the temporal (i.e. aspectual) structure or differences in the way that SubjExp versus ObjExps encode causality (e.g. Brown & Fish, 1983; Corrigan, 1988; Grimshaw, 1990; Dowty, 1991; Pesetsky, 1995; Pylkkänen, 1999; Tenny, 1994; among others) – what appears to be critical *for the process of sentence formulation* is the mapping between thematic and syntactic prominence.

Our studies also show that it is the mapping between event and linguistic structures, specifically – and not, by contrast, other factors related to the non-linguistic processing of the images used in our work – that fundamentally drives the patterns that we saw in Experiment 1a and Experiment 2. First, participants in Experiment 1a and Experiment 2 saw and described the exact same images. If our results were driven by visual properties associated with our stimuli (e.g. one of our characters being more visually expressive, easier to interpret, etc.) or by uncertainty about the event structure, we should have seen a similar pattern of eye-movements across both studies, not the reversal that we actually found. Second, the Emotiveness Rating questionnaires administered after both experiments did not provide any evidence that participants detected any systematic differences in the emotiveness of the characters. Finally, contrary to what is predicted by an emotiveness account of our data, but consistent with a mapping-based account, differences between the SubjExp and ObjExp conditions that emerged during the course of sentence planning disappeared when the task of language production was no longer present (Experiment 1b), indicating that our results are specifically tied to processes associated with sentence planning.

Nor did we find any evidence that our data could be explained by effects stemming from the relative (in)frequency of ObjExp versus SubjExp verbs in active (Experiment 1a) versus passive (Experiment 2) constructions, respectively. And while we believe that frequency can certainly play a role in sentence production, our analyses did not provide any direct evidence for the possibility that the relative (in)frequency of active-voice ObjExp verbs or passive-voice SubjExp verbs in particular was fundamentally driving the pattern of results, at least in the results presented here. Given that we found no evidence for frequency effects, our results at least seem to suggest that the production of active ObjExp (Experiment 1a) and passive SubjExp (Experiment 2) was more difficult not because of reasons related to construction frequency, but because sentence formulation is harder when thematic and syntactic prominence are misaligned.

However, given that the present studies were not designed to explicitly understand the evolving nature of these relationships, further work is certainly necessary to fully investigate the relationship between broader, verb-based construction frequencies in language and processes related to sentence formulation. This is especially true in light of other studies suggesting that it may be processes in language production that ultimately drive the emergence of different construction frequencies in language (see e.g. Gennari & MacDonald, 2009; MacDonald, 1999 for one possible framework).

5.1 Beyond language production: the relationship between language and thought

Although our studies were focused on sentence production, the implications of our work can extend to other domains of language processing and to broader questions about the nature of the relationship between language and thought, more generally. In particular, our results suggest that the misalignment between thematic and syntactic prominence may also account for the behaviour of ObjExps across domains of comprehension and language acquisition. Prior work on Psych Verbs in the domain of comprehension (e.g., Brennan & Pylkkänen, 2010; Cupples, 2002; Gennari & MacDonald, 2009; Wilson & Dillon, 2020), for instance, has shown similar patterns to the ones we found. Very recent work by Wilson and Dillon (2020), for instance, used the self-paced reading paradigm to show that participants read significantly slower and answered sentence-final comprehension questions less accurately in active ObjExp than in active SubjExp conditions. They also showed that comprehenders were comparably slower and less accurate in the passivized SubjExp conditions than in passivized ObjExp conditions. Likewise, studies of adult language acquisition have shown that ObjExp verbs are more difficult to acquire than SubjExp verbs in English (e.g. Chen, 1996; White et al., 1999) and in a typologically different language (e.g. Spanish; Montrul, 1998). And, a similar pattern has been found among children acquiring Tagalog as a first language (De Guzman, 1992; but see Bowerman, 1990 and Hartshorne et al., 2015 for contrasting results in English). These studies found, in other words, that people comprehending and acquiring language performed significantly worse in precisely the same conditions where we found that the the misalignment between thematic and syntactic hierarchies could lead to slower speech onset times and delayed fixations to the subject character.

Taken in the broader context of these studies, then, our results add to the long-standing body of work (e.g. Landau & Gleitman, 1985; Gleitman, 1990) pointing towards a tight relationship between (i) the conceptual structures that we build when we encounter an event we intend to talk about and (ii) the linguistic structures that we use to communicate them (see Papafragou & Grigoroglou, 2019 for a relevant review). Thus, while there is some evidence suggesting that systemic mappings may not be strictly needed in order to acquire a language (e.g. Goldberg, 1995; 2006), our results suggest that they can at least make the job of language production easier.

Against this backdrop, then, an especially important question is what is it about the Experiencer that makes it more prominent than the Stimulus in the conceptual representation of SubjExp and ObjExp events? The present study only investigated events described via Psych Verbs in "out of the blue" contexts. But, one can imagine that a number of factors – including (i) the level of granularity with which an event is represented or the way that an event is construed (e.g. Gleitman et al., 2007), (ii) the salience of the causal relationships between the entities participating in an event (e.g. Au, 1986; Brown & Fish, 1983; Corrigan, 1988; Grimshaw, 1990; Pesetsky, 1995; Rudolf & Forsterling, 1997; Stevenson et al., 1994), (iii) the temporal structure of the event (e.g. Rohde, Kehler, & Elman, 2006), and/or (iv) its relationship to surrounding events – might also have consequences which event role may be considered more conceptually prominent and by extension, which entity may ultimately selected as the subject of the sentence. Yet, compared to the many studies on language production showing how the individual characteristics of an entity can affect subjecthood assignment (Bock et al., 1992; Bock & Warren, 1985; Ferreira & Yoshita, 2003; Prat-Salá & Branigan, 2000; Rissman et al., 2018; among others), far fewer have looked at how the "characteristics" of an event might also affect what is considered prominent and how that conceptual prominence is later encoded in language. For the moment, these questions are beyond the scope of the present work, though our hope is that the studies presented here might motivate future research on these issues.

5.2 Formal representations of psych verbs

Finally, our work also has implications for formal theories of linguistics – in particular, our results appear to pose a challenge for accounts of SubjExp and ObjExp verbs, such as Baker's (1997) Uniformity of Theta Assignment Hypothesis (UTAH), Experiencer and Stimulus roles are both understood as "proto-agents": the Experiencer because it is necessarily animate and volitional, and the Stimulus because it is the cause of an event (e.g. Dowty, 1991; but see Brown & Fish, 1983; Corrigan, 1988). Under these accounts, ObjExp verbs – like SubjExp and Agt-Pat verbs – involve no mismatch between syntactic and semantic prominence and so ObjExp verbs are expected to behave just like SubjExp and Agt-Pat verbs. The results presented here, however, provide evidence to the contrary: Speakers were slower to begin uttering ObjExp sentences and slower to select a starting point for sentence formulation relative to SubjExp and Agt-Pat sentences. Thus, our data demonstrate that an account which treats the Stimulus and the Experiencer as "equally good" candidates for agentivity (and by extension, for subjecthood) is unlikely to be on the right track.

Whether our results provide direct evidence for alternative theories of ObjExp verbs is an open question. Under well-established movement approaches, like that of Belletti and Rizzi (1988), ObjExp verbs are analyzed as SubjExp verbs which have undergone syntactic movement. Evidence for this latter account comes largely from Italian where (i) the Stimulus argument behaves like a moved subject rather than a base-generated one and (ii) the Experiencer argument appears to be a direct object of the verb, but ultimately does not behave like an object in movement operations. In our work, the slower time course of sentence formulation in the case of active voice ObjExp sentences may be construed as evidence for these movement-based accounts (cf. Bock, 1982). At the same time, though, our results may also be consistent with other accounts pointing towards semantic differences between SubjExp and ObjExp verbs. According to some, a fundamental difference between SubjExp and ObjExp verbs is the causal structure of the latter (e.g. Pesetsky, 1995; Pylkkänen, 1999). If so, one possibility could be that speakers in our study had an easier time formulating passive ObjExp sentences (Experiment 2) because the passive structure allowed them to place the thematic causer of the event in the by-phase position – an option that was not available to them when producing sentences like "*Leslie scares Ann.*" (Experiment 1a).

A formal account of Psych verbs being beyond the aims of the current studies, though, we leave it to future work to fully investigate the precise the nature of the relationship underpinning these two verb types. We hope that future work may continue viewing linguistic phenomena through the lens of language production not only because production comprises a naturally core part of language but also because doing so can shed new light on open issues in current theories of linguistics.

6. Conclusion

We asked speakers to produce active (Experiment 1a) and passive (Experiment 2) sentences containing SubjExp (*e.g. loves/fears/admires*) or ObjExp (e.g. *surprised/scared/annoyed*) verbs. In both experiments, we found that speakers were slower to begin speaking and preferentially fixated the subject character later when they were unable to map the most thematically prominent role in the event (i.e. the Experiencer; Grimshaw, 1980; Jackendoff, 1972, 1987) to the subject position. Our studies shed light on how the mapping from conceptual representations to linguistic structures unfolds in real time. They also suggest that sentence formulation is easier when there is a tight correspondence between the conceptual structures people formulate for events and the linguistic structures they use to talk about them.

Notes

1. This misalignment between thematic and syntactic prominence that exists with ObjExp verbs also poses challenges for theoretical linguistics analyses. In particular, an open question is whether the syntactic representation of ObjExp verbs is isomorphic with the surface representation (e.g., Baker, 1988, 1997; Pylkkänen, 1999; Pesetsky, 1995) or whether ObjExp verbs are syntactically derived from their SubjExp counterparts (e.g., Belletti & Rizzi, 1988). The goal of our work is not to provide a formal account for Psych verbs. As will become clear, our work hinges only on the *presence* of a misalignment between thematic and syntactic prominence, and crucially, not on the reason for that mismatch. Our results can be interpreted irrespective of questions of syntactic representation. We return to this issue in the General Discussion.
2. To further ensure that the Agent-Patient were conceptually/thematically distinct from Experiencer-Stimulus verbs, we also used the test of adjective derivations (Brown & Fish, 1983; Levin, 1993). According to this test, Agent-Patient verbs do not typically lend themselves to *-able* endings that sound particularly "natural" (e.g. *?blamable, ?criticizable, ?congratulable*, etc.), while affixation of *-able* to Experiencer-Stimulus verbs appears comparatively better (e.g. *loveable, detestable, admirable*, etc.). Similarly, to ensure that Agent-Patient verbs are conceptually/thematically distinguished from Stimulus-Experiencer verbs, we used the animacy test set forth by Brown and Fish (1983): Agent-Patient are verbs which "name *voluntary* [emphasis added] actions (pg. 242)" and Stimulus-Experiencer verbs are those which "name *involuntary* [emphasis added] mental states, affective, sensorial or cognitive (pg.242)." Thus, when the subject of an Agent-Patient judgement verb is inanimate (i.e. unable to perform a voluntary action), these sentences no longer appear felicitous (e.g. "**The painting blames/criticizes/congratulates/etc. Leslie*"). By contrast, Stimulus-Experiencer verbs seem perfectly acceptable (e.g. "*The painting scares/confuses/amazes Leslie.*").
3. In addition to the image interpretability and Emotiveness questionnaires, we also asked participants to complete the ASQ Questionnaire (Baron-Cohen et al., 2001) at the very end of the experiment session. Here, there were no specific predictions; however, we thought it might be possible to use the ASQ to investigate whether individual differences in communication/emotional intelligence might affect the way that speakers would plan utterances related to psychological states/emotions. We found no significant results involving the overall ASQ or its subscales; for this reason, we do not discuss it further.
4. A separate analysis was conducted where trials which were marked "unclear" in the Image Interpretability Questionnaire were excluded. These showed the same pattern as the results reported below.
5. Additional analyses related to the rates of errors across Verb Type conditions was also performed. They are available through the first author's OSF and discussed in further detail in those documents.
6. In fact, we also performed a separate analysis, which excluded items which were marked "unclear". This analysis yielded a similar pattern of results as the ones reported here.
7. Due to experimenter error, two different items were inadvertently excluded from each of the three lists: Participants in List 1 did not see the *praises* or *hate* items; participants in List 2 did not see the *praises* or *confuses* items; and participants in List 3 did not see the *hates* or *confuses* items.
8. A separate analysis which did include short passives was conducted. This yielded the same pattern of results.
9. In the COCA corpus analyses, we only included long passive occurrences such as "was loved/frightened by" for two reasons: (1) We only analyzed long passive utterances in our data and (2) We did not want to include in our corpus counts utterances such as "He was hated and controversial in his last years … ", because these adjectival passive constructions are syntactically and semantically different from the verbal passive constructions that participants were asked to produce in our study (e.g., Wasow, 2020).

Acknowledgements

We are extremely grateful to Ana Besserman for creating the images used in these experiments. We also thank Toben Mintz, Stefan Keine, Andrew Simpson, Rand Wilcox, Michael Wilson, and our reviewers for their thoughtful feedback on this work. Finally, we thank the audiences of the 10th meeting of the International Workshop on Language Production and the 93rd Meeting of the Linguistic Society of America, where portions of this work were presented.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Declaration of interests statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Au, T. K. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*(1), 104–122. https://doi.org/10.1016/0749-596X(86)90024-0 doi:10.1016/0749-596X(86)90024-0

Baker, M. (1988). *Incorporation: A theory of grammatical function changing*. PhD dissertation. University of Chicago.

Baker, M. (1997). Thematic roles and syntactic structure. In L. Haegeman (Ed.), *Elements of grammar: Handbook of generative syntax* (pp. 73–137). Kluwer.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01 doi:10.18637/jss.v067.i01

Becker, D. V., Anderson, U. S., Mortensen, C. R., Neufeld, S. L., & Neel, R. (2011). The face in the crowd effect unconfounded: Happy faces, not angry faces, are more efficiently detected in single- and multiple-target visual search tasks. *Journal of Experimental Psychology: General*, *140*(4), 637–659. https://doi.org/10.1037/a0024060 doi:10.1037/a0024060

Belletti, A., & Rizzi, L. (1988). Psych verbs and q-theory. *Natural Language and Linguistic Theory*, *6*(3), 291–352. https://doi.org/10.1007/bf00133902 doi:10.1007/BF00133902

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contribu- tions to sentence formulation. *Psychological Review*, *89*(1), 1–47. https://doi.org/10.1037/0033-295X.89.1.1 doi:10.1037/0033-295X.89.1.1

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning & Verbal Behavior*, *19*(4), 467–484. https://doi.org/10.1016/S0022-5371(80)90321-7

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

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

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*(1), 150–171. https://doi.org/10.1037/0033-295X.99.1.150 doi:10.1037/0033-295X.99.1.150

Boersma, P., & Weenink, D. (1992–2011). Praat: Doing phonetics by computer. http://www.praat.org/

Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics*, *28*(6), 1253–1289. https://doi.org/10.1515/ling.1990.28.6.1253

Brennan, J., & Pylkkänen, L. (2010). Processing psych verbs: Behavioural and MEG measures of two different types of semantic complexity. *Language and Cognitive Processes*, *25*(6), 777–807. https://doi.org/10.1080/01690961003616840

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*(3), 237–273. https://doi.org/10.1016/0010-0277(83)90006-9 doi:10.1016/0010-0277(83)90006-9

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

Chen, D. (1996). *L2 acquisition of English psych predicates by French and Chinese native speakers*. Unpublished PhD dissertation, McGill University, Montreal.

Clark, Herbert H. (1965). Some structural properties of simple active and passive sentences. *Journal of Verbal Learning and Verbal Behavior*, *4*(5), 365–370. https://doi.org/10.1016/S0022-5371(65)80073-1

Clark, H. H., & Begun, J. S. (1971). The Semantics of sentence subjects. *Language and Speech*, *14*(1), 34–46. https://doi.org/10.1177/002383097101400105 doi:10.1177/002383097101400105

Corrigan, R. (1988). Who dun it? The influence of actor-patient animacy and type of verb in the making of causal attributions. *Journal of Memory and Language*, *27*(4), 447–465. https://doi.org/10.1016/0749-596X(88)90067-8 doi:10.1016/0749-596X(88)90067-8

Cupples, L. (2002). The structural characteristics and on-line comprehension of experiencer-verb sentences. *Language and Cognitive Processes*, *17*(2), 125–162. https://doi.org/10.1080/01690960143000001

Davies, M. (2008). The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/

De Guzman, V. (1992). *Acquisition of voice affixes in psychological verbs in Tagalog*. Paper presented at the 3rd International symposium on Pan-Asiatic Linguistics, January.

Do, M. L., & Kaiser, E. (2019). Subjecthood and linear order in linguistic encoding: Evidence from the real- time production of wh-questions in English and mandarin Chinese. *Journal of Memory and Language*, *105*, 60–75. https://doi.org/10.1016/j.jml.2018.11.001

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619. https://doi.org/10.2307/415037 doi:10.1353/lan.1991.0021

Engelberg, S. (2018). The argument structure of psych-verbs: A quantitative corpus study on cognitive entrenchment. In Boas, H. C. & Ziem, A. (Eds.), *Constructional approaches to syntactic structures in German*, 47–84. De Gruyter Mouton. https://doi.org/10.1515/9783110457155-002

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

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

Fillmore, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1-88). Holt, Rinehart and Winston.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic boostrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 143–149. https://doi.org/10.1002/wcs.17

Fox, E., Lester, V., Russo, R., Bowles, R.J, Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently?. *Cognition & Emotion*, *14*(1), 61–92. https://doi.org/10.1080/026999300378996

Ganushchak, L. Y., Konopka, A. E., & Chen, Y. (2014). What the eyes say about planning of focused referents during sentence formulation: A cross-linguistic investigation. *Frontiers in Psychology*, *5*, 1124. https://doi.org/10.3389/fpsyg.2014.01124 doi:10.3389/fpsyg.2014.01124

Ganushchak, L. Y., Konopka, A. E., & Chen, Y. (2017). Accessibility of referent information influences sentence planning: An Eye-tracking study. *Frontiers in Psychology*, *8*, 250. https://doi.org/10.3389/fpsyg.2017.00250 doi:10.3389/fpsyg.2017.00250

Gennari, Silvia P, & MacDonald, Maryellen C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23. https://doi.org/10.1016/j.cognition.2008.12.006

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the give- and-take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. https://doi.org/10.1016/j.jml.2007.01.007 doi:10.1016/j.jml.2007.01.007

Goldberg, A. (1995). *Construction grammar: a construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work: The nature of generalization in language*. Oxford, UK: Oxford University Press.

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

Grimshaw, J. (1980). *Argument Structure* . Cambridge, MA: The MIT Press.

Grimshaw, J. (1990). *Argument structure*. the MIT Press.

Hartshorne, J., Pogue, A., & Snedeker, J. (2015). *Love* is hard to understand: The relationship between transitivity and caused events in the acquisition of emotion verbs. *Journal of Child Language*, *42*(3), 467–504. https://doi.org/10.1017/S0305000914000178 doi:10.1017/S0305000914000178

Hirsh-Pasek, K, & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA.: The MIT Press.

Humphrey, K., Underwood, G., & Lambert, T. (2012). Salience of the lambs: A test of the saliency map hypothesis with pictures of emotive objects. *Journal of Vision*, *12*(1), 1–15. https://doi.org/10.1167/12.1.22 doi:10.1167/12.1.22

Hwang, H., & Kaiser, E. (2015). Accessibility effects on production vary cross-linguistically: Evidence from English and Korean. *Journal of Memory and Language*, *84*, 190–204. https://doi.org/10.1016/j.jml.2015.06.004 doi:10.1016/j.jml.2015.06.004

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. MIT Press.

Jackendoff, R. S. (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry*, *18*(3), 369–411.

Konopka, A. E. (2019). Encoding actions and verbs: Tracking the time-course of relational encoding during message and sentence formulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(8), 1486–1510. https://doi.org/10.1037/xlm0000650 doi:10.1037/xlm0000650

Konopka, A. E., & Kuchinsky, S. E. (2015). How message similarity shapes the timecourse of sentence formulation. *Journal of Memory and Language*, *84*, 1–23. https://doi.org/10.1016/j.jml.2015.04.003 doi:10.1016/j.jml.2015.04.003

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

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press.

Lee, Eun-Kyung, Brown-Schmidt, Sarah, & Watson, Duane G. (2013). Ways of looking ahead: Hierarchical planning in language production. *Cognition*, *129*(3), 544–562. https://doi.org/10.1016/j.cognition.2013.08.007

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

Levin, B. (1993). *English verb classes and alternations*. University of Chicago Press.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language*. Mahweh, New Jersey: Lawrence Erlbaum Associates Publishers.

Matuschak, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001 doi:10.1016/j.jml.2017.01.001

Montrul, S. (1998). The L2 acquisition of dative experiencer subjects. *Second Language Research*, *14*(1), 27–61. https://doi.org/10.1191/026765898668810271 doi:10.1191/026765898668810271

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. (2015). Word order affects the time- course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*(9), 1187–1208. https://doi.org/10.1080/23273798.2015.1006238.

Osgood, C. E. (1971). Where do sentences come from? In D. Steinberg (Ed.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 497–529). Cambridge University Press.

Papafragou, A., & Grigoroglou, M. (2019). The role of conceptualization during language production: Evidence from event encoding. *Language, Cognition and Neuroscience*, *34*(9), 1117–1128. https://doi.org/10.1080/23273798.2019.1589540 doi:10.1080/23273798.2019.1589540

Pesetsky, D. (1995). *Zero syntax*. MIT Press.

Porretta, V., Kyröläinen, A-J., van Rij, J., & Järvikivi, J. (2018). Visual world paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski et al. (Eds.), *Intelligent decision technologies 2017* (pp. 268-277).

Springer International Publishing. https://doi.org/10.1007/978-3-319-59424-8_25

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

Prentice, J. L. (1967). Effects of cueing actor vs. Cueing object on word order in sentence production. *Psychonomic Science*, *8*(4), 163–164. https://doi.org/10.3758/BF03331600 doi:10.3758/BF03331600

Pritsch, C., Telkemeyer, S., Mühlenbeck, C., & Liebal, K. (2017). Perception of facial expressions reveals selective affect-biased attention in humans and orangutans. *Scientific Reports*, *7*(7782), 1–12. https://doi.org/10.1038/s41598-017-07563-4

Pylkkänen, L. (1999). On stativity and causation. In L. Carol Tenny, & J. Pustejovsky (Eds.), *Events as grammatical objects: The converging perspective of lexical semantics and syntax* (pp. 417–444). CSLI Publications.

R Core Team. (2016). *R: A language and environment for statisticalcomputing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rissman, L., Woodward, A., & Goldin-Meadow, S. (2018). Occluding the face diminishes the conceptual accessibility of an animate agent. *Language, Cognition and Neuroscience*, *34*(3), 273–288. https://doi.org/10.1080/23273798.2018.1525495 doi:10.1080/23273798.2018.1525495

Rohde, H., Kehler, A., & Elman, J. (2006). Event structure and discourse coherence biases in pronoun interpretation. In R. Sun & N Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rudolf, U., & Forsterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, *121*(2), 192–218. https://doi.org/10.1037/0033-2909.121.2.192 doi:10.1037/0033-2909.121.2.192

Sauppe, S., Norcliffe, E., Konopka, A. E., van Valin, R. D., & Levinson, S. (2013). Dependencies first: Eye-tracking evidence from sentence production in tagalog. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, 1265–1270.

Savage, R., Lipp, O. V., Craig, B. M., Becker, S. I., & Horstmann, G. (2013). In search of the emotional face: Anger versus happiness superiority in visual search. *Emotion*, *13*(4), 758–768. https://doi.org/10.1037/a0031970 doi:10.1037/a0031970

Schumacher, P., Roberts, L., & Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German *er* and *der*. *Lingua. International Review of General Linguistics. Revue internationale De Linguistique Generale*, *185*, 25–41. https://doi.org/10.1016/j.lingua.2016.07.004

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*(4), 519–548. https://doi.org/10.1080/01690969408402130 doi:10.1080/01690969408402130

Tenny, C. (1994). *Aspectual Roles and the Syntax-Semantics Interface*. Dordrecht: Kluwer.

Thompson, C. K., & Lee, M. (2009). Psych verb production and comprehension in agrammatic Broca's aphasia. *Journal of Neurolinguistics*, *22*(4), 354–369. https://doi.org/10.1016/j.jneuroling.20.08.11.0 03

van Rij, J. (2015). Overview GAMM analysis of time series data. http:// www.sfs.uni-tuebingen.de/~jvanrij/Tutorial/GAMM.html

van Rij, J., Wieling, M., Baayen, R., & van Rijn, H. (2020). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 2.4.

Wasow, T. (1977). Transformations and the lexicon. In P. Culicover, A. Akmajian, & T. Wasow (Eds.), *Formal Syntax* (pp. 327–360). New York: Academic Press.

White, L., Brown, C., Bruhn-Garavito, J., Chen, D., Hirakawa, M., & Montrul, S. (1999). Psych verbs in second language acquisition. In E. Klein, & G. Martohardjono (Eds.), *The development of second language grammars. A generative approach* (pp. 173–199). John Benjamins.

Wilcox, R. R. (2012). *Modern statistics for the social and behavioral sciences: A practical introduction*. CRC Press.

Wilson, M., & Dillon, B. (2020). *Thematic roles' alignment with grammatical functions facilitates sentence processing*. Poster presented at the 26th Architectures and Mechanisms for Language Processing (AMLaP) conference. September 3-4. University of Massachusetts, Amherst.

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, growth curve analysis and Generalized Additive modeling. *Journal of Language Evolution*, *1*(1), 7–18. https://doi.org/10.1093/jole/lzv003

Wood, S. (2017). *Generalized Additive models: An introduction with R* (2nd ed). Chapman and Hall/CRC.