Production of Syntactic Alternations Displays Accessibility But Not Informativity Effects

Emily Goodwin (goodwine@stanford.edu)
Department of Linguistics, 460 Jane Stanford Way
Stanford, CA 94305 USA

Judith Degen (jdegen@stanford.edu)
Department of Linguistics, 460 Jane Stanford Way
Stanford, CA 94305 USA

Abstract

This paper explores how speakers choose between two utterance alternatives with similar syntactic properties and distinct yet related meanings. We consider the interaction of two speaker pressures: to mention accessible lexical items early in the utterance and to mention informative content early in the utterance, the latter of which is explicitly predicted by an incremental Rational Speech Act (IRSA) model. In Exp. 1, we observed a significant effect of accessibility on utterance choice in an online spoken production task, which elicited descriptions of the relationship between two entities using a provided verb. We found that making entities more accessible via foregrounding led speakers to mention them earlier. In Exp. 2, an interactive production task, both informativity and foregrounding were manipulated. While IRSA predicts more informative content to be mentioned earlier in the sentence, we observed neither significant effects of informativity nor of accessibility. Consistent with recent work on Good-Enough theories of production, we conclude that even when two sentences are not entirely meaning-equivalent, production choices can be affected by lexical accessibility; the pressure to mention informative material early, however, should be investigated further.

Keywords: Language production; Syntactic alternations; Good-Enough Production; Informativity; RSA

Introduction

Language is flexible, often presenting speakers with the choice between many (near-)meaning-equivalent utterance alternatives. For example, how do speakers choose between the two forms given in (1)?

(1) a. Sally sprayed the fence with paint. (Location-first form)
b. Sally sprayed paint on the fence. (Substance-first form)

Rational Speech Act (RSA) models of production have been successfully used to capture whole-utterance production choices, including in the domain of scalar implicatures and redundant referring expressions (Degen et al., 2020; Frank & Goodman, 2012; Goodman & Frank, 2016). However, empirical work on production shows that speakers plan utterances incrementally. Much of this work has focused on speakers' preference to produce more accessible linguistic material earlier in meaning-equivalent sentences (Availability-Based Production, V. S. Ferreira & Dell, 2000; V. S. Ferreira & Griffin, 2003). The notion of accessibility is poorly defined, but typically refers to material that is easy to retrieve or produce. Building on this work, the Good-Enough Production account has emerged, under which speakers are taken to produce "good enough" utterances, balancing message alignment and word accessibility (Koranda et al., 2022). This theory predicts that when two alternative utterances have sufficiently similar meanings, speakers prefer to produce the one that places more accessible lexical items earlier.

Recent incremental RSA (IRSA) models have formalized the trade-off between producing informative vs. cheap material early (Cohn-Gordon et al., 2019; Futrell, 2023, 2024; Waldon & Degen, 2021). One under-explored aspect of incremental models is the grain-size of planning: the speaker model in Cohn-Gordon et al. (2019), for example, plans utterances one word at a time. This predicts that speakers prefer to produce more informative words first, and is at odds with the empirical observation that speakers tend to prefer recently-mentioned words before novel ones (the so-called "given-before-new" principle, Arnold et al., 2000; Gundel, 1988; Wasow, 2002). Since new words are likely to be more informative than recently-mentioned words, the IRSA model predicts the opposite of the given-before-new preference.

In this paper, we investigate whether speakers choosing between two utterances with slightly different meanings prefer to produce the form with more accessible or more informative material earlier, thus pitting Good-Enough Production and IRSA against each other. We begin by detailing the predictions of Good-Enough theories of production, then exemplify the predictions of an IRSA production model. In Exp. 1, we manipulate the accessibility of words, testing if speakers structure their sentences to permit early mention of more accessible nouns in a single-agent spoken production task. In Exp. 2, we manipulate both accessibility and informativity of the relevant nouns in an interactive spoken production task, testing whether speakers structure their sentences to permit early mention of more informative and/or more accessible nouns. We find accessibility effects in Exp. 1, but not in Exp. 2, and no effect of informativity.[1]

The *spray-load* alternation exemplified in (1) is a perfect test bed for studying the interaction of accessibility and informativity in language production for three reasons. First, unlike other English alternations (e.g., the dative alternation or voice alternations) the syntax of the two *spray-load* forms is similar and sentence length does not vary. Similarly, since the forms are identical until the first object is produced, the

[1]Code available at `https://github.com/emilygoodwin/holisticSprayLoad`.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

information content of the forms is also identical until the first object. This makes *spray-load* an excellent test-case for investigating incremental effects. Finally, the forms are associated with different (but similar) meanings: (1a), but not (1b), is said to convey a fence either entirely or more covered in paint (Anderson, 1971; Buck, 1993; Jeffries & Willis, 1984; Levin, 1993). This expands work on Good-Enough Production, by exploring the trade-off between message alignment and accessibility in whole utterance production.

Good-Enough Production

Studies investigating Availability-Based Production have shown that accessibility of lexical items or syntactic structures affects speakers' production preferences, such that speakers tend to order more accessible material earlier in sentences. Many factors affect accessibility, including a word's animacy (F. Ferreira, 1994; McDonald et al., 1993), imageability (Bock & Warren, 1985), givenness (Bock & Irwin, 1980), frequency (Oldfield & Wingfield, 1965), and subliminal salience (Gleitman et al., 2007).

Visual salience has also been shown to affect accessibility: for instance, Dutch speakers in a spoken-production study were more likely to use the active (agent-first) form when describing scenes with foregrounded rather than backgrounded agents (Vogels et al., 2013). Similarly, participants in a forced-choice task prefer *spray-load* structures that mention the foregrounded noun first (D'Elia, 2016). However, it is unclear whether these results are artifacts of the forced choice task or whether they generalize to spoken production.[2]

While most work testing Availability-Based Production has focused on meaning-equivalent alternatives, recent work has found that the pressure to mention more accessible material earlier can even lead to speakers producing incorrect or less than fully-informative utterances: Koranda et al. (2022) trained participants on novel words representing cardinal directions, then asked the participants to produce the novel words by typing instructions to characters in a game. They found that participants tended to type words that had been shown more often during training, even when they knew a less-frequent alternative that was more aligned with the target direction. Characterizing the nature of this trade-off is an important question for understanding sentence production, and the focus of the Good Enough theory of production.

Incremental Iterated Speech Production

RSA models have been used to model a number of production phenomena (Cohn-Gordon et al., 2019; Degen et al., 2020; Frank & Goodman, 2012; Goodman & Frank, 2016). These models assume that speakers and listeners reason about each other recursively: a pragmatic speaker S_1 produces an utterance u, given a world state w, with probability proportional to how likely a literal listener L_0 would arrive at the target world state given that utterance, minus the cost of producing utterance u:

$$P_{S_1}(u|w) \propto e^{\alpha(\log L_0(w|u) - cost(u))} \quad (1)$$

The literal listener L_0 assigns probability uniformly to all world states consistent with utterance u. For example, consider a listener choosing between three world states which each contain a different substance object (*paint, water* or *soap*), and all of which contain the same location object (*fence*). If the speaker produces the utterance *Someone sprayed the fence*, L_0 would assign $P(w) = 33\%$ to each of the three states, since all three contain a fence. However, if the utterance is *someone sprayed paint*, L_0 would assign a probability of 1 to one state and 0 to the others, since only this state is compatible with the utterance. Finally, the optimality parameter α models how rational a speaker is: higher values of α yields more rational speakers, i.e., that are more likely to choose the utility-maximizing utterance.

To extend this approach to capture incremental speech production, Cohn-Gordon et al. (2019) assume that speakers and listeners process utterances one word a time, updating their beliefs about the world with each new word. Speakers in this model select a word based on the world w and the list of words already produced c:[3]

$$S_1^{WORD}(word|c,w) \propto e^{\alpha(\log L_0^{WORD}(w|c,word) - cost(word))} \quad (2)$$

Applying this model to the *spray-load* alternation in (1), an incremental speaker S_1 with context $c =$ *Sally will spray the* must choose between the next words *paint* and *fence*. To see how informativity affects the speaker's choice, consider the three scenes in Fig. 1. Since each scene depicts a fence, but only the target scene depicts paint, the word *paint* is informative and the speaker is predicted to prefer (1b). Fig. 2 demonstrates this effect for a speaker selecting between an informative word w and a non-informative competitor. To operationalize the accessibility effect of foregrounding, we model backgrounded nouns as having a higher cost.

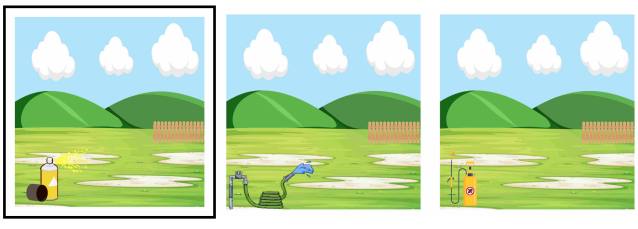


Figure 1: An example stimulus for Exp. 2, in the substance-foregrounded, substance-informative condition.

In sum, Good-Enough Production predicts that speakers produce utterances with more accessible items mentioned first. On the other hand, the IRSA model predicts more informative nouns to be mentioned first (with the preference

[2]There are various explanations for why foregrounded items are mentioned earlier. Vogels et al. (2013) argue that increased visual salience affects the global conceptual interpretation of the scene, while D'Elia (2016) attributes the effect to attention allocation.

[3]Note that Cohn-Gordon et al. (2019) do not include α in their equation, because they fix the optimality parameter for all their simulations. We vary α in our simulations, to show how increasingly rational speakers balance informativity and accessibility.

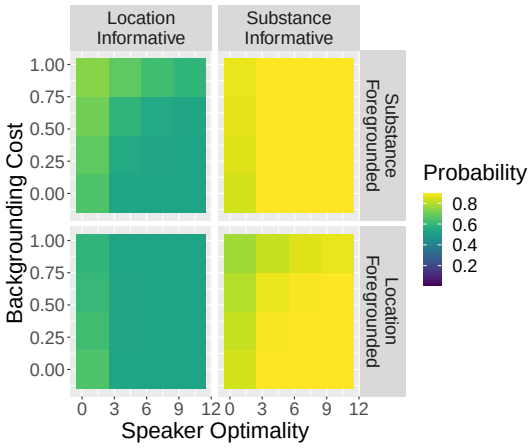
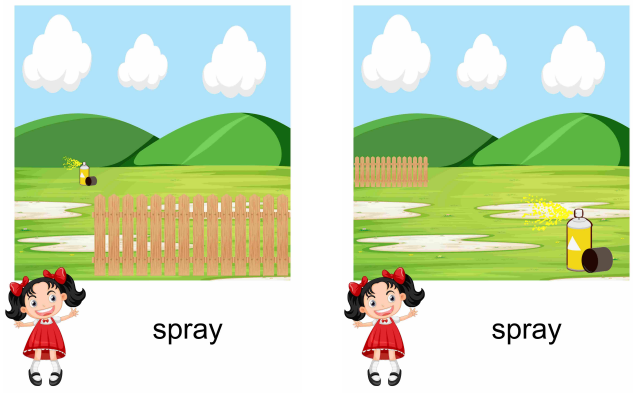


Figure 2: Predicted IRSA probability of producing the substance (*paint*) rather than the location (*fence*) first, based on its informativity (columns) and foregrounding (rows). Along the y-axis, different values for the cost of backgrounding the item; along the x axis, different speaker optimality values.



(a) Location foregrounded (b) Substance foregrounded

Figure 3: Example stimuli from Experiment 1 in each foregrounding condition.

modulated somewhat by accessibility, via a backgrounding cost). We tested these predictions in two spoken production experiments. Exp. 1 used a picture-description task targeting just the accessibility prediction. Exp. 2 used an interactive picture-description language game to test both the accessibility and informativity predictions.

Exp. 1: Non-Interactive Production

This experiment tested whether speakers describing visual scenes with *spray-load* verbs place nouns corresponding to foregrounded objects earlier.

Methods

Participants. We recruited 60 participants on Prolific. Five participants were excluded from the analysis because they did not report that their first language was English.

Materials and procedure. On each trial, participants were shown an image of an agent (Sally) next to a scene with a contextualizing background containing two objects (e.g., a fence and some paint in a field as in Fig. 3; a mushroom and some cheese in a kitchen). Additionally, they were shown a verb and asked to use it in a full sentence, describing the scene and mentioning Sally's role and both objects. Typical productions when shown the verb *spray* and a scene with a fence and can of spray paint are given in (1).

On every trial, one of the objects was placed in the background of the scene and one in the foreground. Each verb was shown with a foregrounded location object on two trials and a foregrounded substance object on two trials. An example stimulus in both of the conditions is shown in Fig. 3.

The verbs for critical trials were taken from the list of *spray-load* alternating verbs in Levin (1993): *spray, spread, stuff,* and *load*. Verbs for control trials were listed as non-alternating: *drench, cover, put,* and *stash*. Each verb was shown with four scenes, each of which included a unique location and substance object. Each critical verb was paired with a control verb, such that each substance-location noun pair was shown twice: once with a *spray-load* verb, and once with a non-alternating verb. To avoid priming participants with the non-alternating verbs, each control verb was only shown after all four of the paired critical trials.

Each participant completed 16 critical trials (constructed from the *spray-load* verbs *spray, load, spread,* and *stuff*) and 16 control trials (constructed from the four non-alternating verbs *drench, stash, cover* and *put*). An additional four filler trials were constructed from two dative verbs, each of which was shown twice with a unique scene, for a total of 36 trials. The full list of the critical and control verbs, and their corresponding noun-pairs, are shown in Table 1.

Critical	Control	Substance∼Location
spray	drench	paint∼fence, water∼car, soap∼table, poison∼bush
spread	cover	frosting∼cupcake, honey∼pastry, butter∼toast, ketchup∼hotdog
stuff	put	cheese∼mushroom, cash∼envelope, paper∼shoe, rice∼bellpepper
load	stash	fruit∼plane, trash∼train, wood∼truck, hay∼wagon

Table 1: The *spray-load* verbs used on critical trials, corresponding non-alternating control verbs, and the substance∼location noun pairs used for each.

To familiarize participants with the objects and reduce any effects of noun frequency, participants were trained on the name of each object. This was done in two phases: In the familiarization phase, every object was shown with its corresponding noun. In the recall phase, participants were shown each object and asked to type the correct noun in a text field. After testing their microphone and completing the noun train-

ing, participants were trained on two examples of main experiment trials, each constructed with a non-alternating verb (*coat* and *spill*). They recorded themselves producing one-sentence spoken descriptions and saw an example answer for each. They then began the main phase of the experiment.

The experiment was programmed using jsPsych (de Leeuw et al., 2023) and utterances automatically transcribed using Faster-Whisper (https://github.com/SYSTRAN/faster-whisper). The utterances were coded by the authors as either location-first, substance-first, or as one of the to be excluded categories described next.

Exclusions We excluded trials for the following reasons: the audio was not sufficiently clear to transcribe; the participant self-corrected before finishing a complete sentence or did not mention both objects; the participant did not use the provided verb as the main verb of the sentence (e.g, *Use the paint to spray the fence*); the participant forced a non-alternating control verb into an ungrammatical structure (e.g, *drench paint on the fence*, but see the results section for an auxiliary analysis of these productions); the participant construed the event differently than intended, either because they interpreted the location and subject objects to play different roles (e.g, instead of *put paper in the shoe*, they said *put the shoe on the paper*), or reinterpreting the event to involve a location or subject object that was not one of the intended objects (e.g, instead of *spray poison on the bush* they said *spray poison on the field*). A total 1613 of the 1920 critical and control trials entered analysis; 805 critical and 808 control.

Norming

We ran a norming study to control for possible differences in baseline preferences for location-first utterances based on the meanings of the two *spray-load* forms. Recall that the location-first form is said to convey that the entire location is affected. Since our items might have differed in how likely a person engaging in the relevant event would be to bring about a state where the entire location is affected (e.g., the entire fence being covered in paint), we collected affectedness values by presenting a separate group of participants with the scenes from Exp. 1. Participants rated each item for how affected (i.e, how *full* or how *covered*) the location object would be after the event took place. We found that items differed in their average affectedness ratings, and therefore use the average affectedness rating for each stimulus as a predictor in the model of Exp. 1 productions. [4]

Results

Fig. 4A shows the proportion of location-first utterances with *spray-load* verbs in each foregrounding condition. Overall, participants preferred not to mention the location first. However, when the location was foregrounded, participants were slightly more likely to use the location-first form: 28.2% of the 401 location-foregrounded trials led to a location-first utterance, while only 21.0% of the substance-foregrounded trials led to a location-first utterance. This was confirmed by a mixed-effects logistic regression predicting location-first *spray-load* forms from centered fixed effects of foregrounding (reference level before centering: 'substance foregrounded') and location affectedness norms, as well as the maximal random effects structure justified by the design: by-participant and by-item random intercepts and slopes for the fixed effects. There was a significant effect of the foregrounding on the predicted log odds of producing a location-first *spray-load* form, such that foregrounded locations were more likely to be mentioned first ($\beta = 0.88, SE = .34, p < .01$). The effect of affectedness did not reach significance ($\beta = 0.02, SE = 0.03, p > .46$).

The foregrounding effect numerically went in the predicted direction for each verb (see Fig. 4B). However, although the tested verbs are all categorized as alternating by Levin (1993), the rate of location-first use varies widely between them, from 51.6% (*stuff*, location-foregrounded condition) to 2.80% (*spread*, substance-foregrounded condition).[5]

To investigate the verb-specific effects, the proportions of location-first utterances produced by each participant with each verb are shown in Fig. 4C. Five participants who did not produce complete *spray-load* utterances on at least three of the four trials per verb were excluded from this analysis. There was variation in location-first productions: 44 participants preferred the substance-first form, including five who produced only the substance-first form, while only six participants preferred the location-first form.

Individuals thus differed in how likely they were to use the location-first form, but followed a general trend with respect to what verbs are used with the location-first form: participants who produced *spray* in location-first forms were more likely to produce *stuff* in this form as well; similarly, the participants who used *load* in the location-first form were more likely to use *spray* and *stuff* in this form as well. To our knowledge, this by-verb pattern has not been reported in production; it is interesting to note, however, that a parallel implicational hierarchy has been reported in the cross-linguistic distribution of which verbs alternate (Kim, 1999).

Control verbs Recall that on control trials, participants saw the same stimuli as with *spray-load* verbs, but with the non-alternating verbs *cover* and *drench* (which only take the location-first form) and *put* and *stash* (which only take the substance-first form). Participants overwhelmingly produced grammatical structures with the control verbs. However, on 11 trials, participants forced the control verbs into the alternate, ungrammatical structure, including *cover frosting (all over/ onto) the cupcake; stash the wagon with the hay;* and *drench paint on the fence*. Nine of the 11 coerced productions put the foregrounded item first. This might therefore be ex-

[4] Norming materials and analyses are in the project repository.

[5] It is possible that self-priming or priming from control trials affected participants' productions. Our initial analyses suggest there may have been some priming from control trials, but priming is not sufficient to explain the pattern of results. See the project repository for a summary of these analyses.

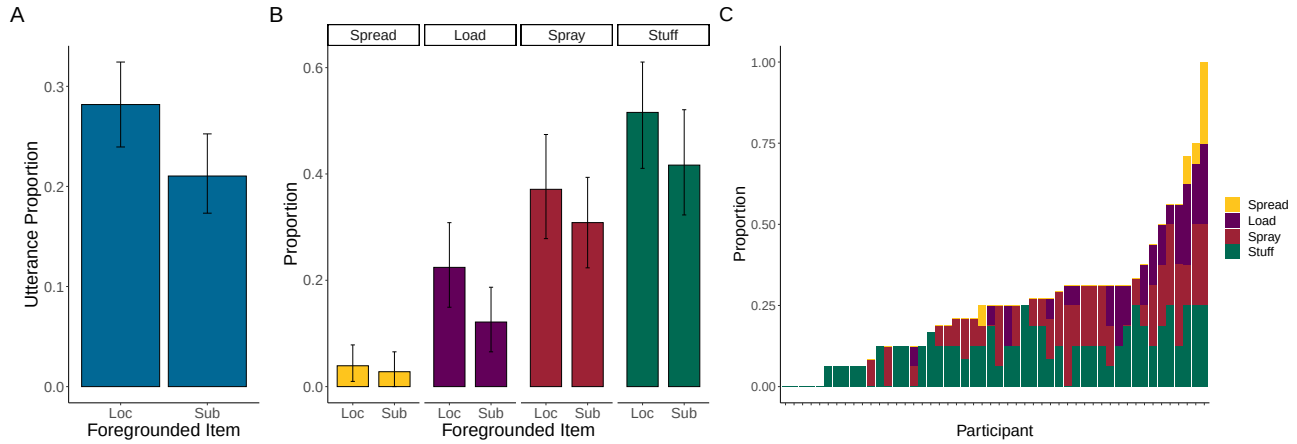


Figure 4: A. Proportion of location-first *spray-load* utterances when location or substance was foregrounded. Error bars indicate bootstrapped 95% CIs. B. Proportion of location-first *spray-load* utterances when location or substance was foregrounded, by verb. C. By-participant proportions of trials with location-first *spray-load* utterances by verb (only including participants who produced full *spray-load* utterances on at least three critical trials).

plained as an instance of Good-Enough Production: speakers produce an ungrammatical utterance because of the increased availability of the noun referring to the foregrounded item.

The small but robust accessibility effect on the choice of *spray-load* form is predicted by Good-Enough Production and compatible with IRSA. Next we test the informativity effect predicted by IRSA.

Exp. 2: Interactive Production

Recall that the IRSA model predicts speakers should mention informative nouns earlier in the utterance. To test this prediction we extended the design of Exp. 1, manipulating both the foregrounding and the informativity of the substance and location nouns. To manipulate informativity, directors were asked to describe a scene from a set of three to a matcher, whose task was to click on the described target scene. Informativity was manipulated by varying how often the location or substance object occurred across scenes.

Methods

Participants. We recruited 152 participants on Prolific and assigned them to pairs. After excluding participants who did not report English as their first language, did not record usable audio, were not judged to be first-language English speakers by the authors, or whose partner left the game experiment, 63 pairs remained. We additionally excluded individual trials according to the same criteria as in Exp. 1, for a total of 800 critical utterances to analyze.

Materials and procedure. Participants in each pair were randomly assigned to the director or matcher role. During the main experiment phase participants saw three scenes on each trial. Directors also saw a black square placed around the target scene, a verb, and instructions that read "*Make a sentence with the verb [**verb**], to tell your partner what Sally will do.*" Matchers saw the same three scenes, shuffled to be presented in a potentially different order, and instructions that read "*Please select the picture that shows what Sally will do today, based on your partner's description*". Both participants received feedback on whether the matcher had selected the correct target, and then advanced to the next trial.

Participants first completed the familiarization phase (but not the recall phase) of Exp. 1, to become acquainted with the names of all of the experimental objects. They then completed two practice trials, each constructed with a non-alternating verb (*cover* and *fill*). In the practice phase, directors were shown an example stimulus and verb, and then shown a possible example sentence; guessers were shown an example stimulus and sentence, and then which of the scenes was the intended target. After the practice phase, participants connected to an audio call with their partner and began the main portion of the experiment. They remained in the audio call for the rest of the experiment. The experiment was programmed using the virtual lab platform Empirica (Almaatouq et al., 2021), with audio recording using the 100ms API (https://www.100ms.live). Utterances were automatically transcribed as in Exp. 1.

As in Exp. 1, each scene included a foregrounded and a backgrounded object. The location was foregrounded in all three scenes in the location-foregrounded condition, and vice versa in the substance-foregrounded condition. We crossed this factor in a 2x2 design with informativity: in the location-informative condition, the three scenes each included a different location but the same substance; and vice versa in the substance-informative condition. An example substance-foregrounded /-informative scene is shown in Fig. 1.

The verbs and target scenes for Exp. 2 were taken from Exp. 1. As in Exp. 1, each verb was shown with four target scenes, each of which included a unique location and substance object. Distracter scenes were generated using the other target scenes for the same verb, so that items were al-

ways shown on an appropriate background. Exp. 2 also used the same trial order randomization as Exp. 1, so that participants saw each non-alternating control verb after all four of its matched *spray-load* trials were shown. Each participant pair saw 16 critical trials (constructed from four *spray-load* verbs), 16 control trials (constructed from four non-alternating verbs), and four filler trials (constructed from two dative verbs), for a total of 36 trials.

Results

The proportion of location-first utterances in each of the conditions is shown in Fig. 5. As in Exp. 1, participants overall preferred not to mention the location first: 33.17% of utterances on critical trials were location-first ($n = 100$).

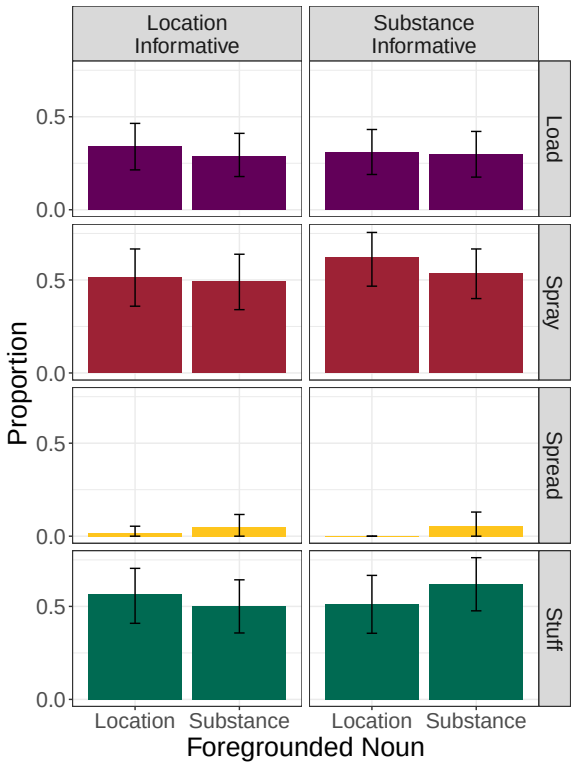


Figure 5: Proportion of location-first *spray-load* utterances by verb. Error bars indicate bootstrapped 95% CIs.

To test whether there was an effect of foregrounding or informativity on location-first mention, we ran a mixed-effects logistic regression predicting location-first mention from centered fixed effects of foregrounding, informativity (reference levels before centering: 'substance foregrounded', 'substance informative'), their interaction, and the independently-obtained affectedness values for each stimulus. The random effects structure included by-item and by-participant random intercepts. There was no significant effect of foregrounding ($\beta = 0.24, SE = 0.23, p > .30$) or informativity ($\beta = -0.11, SE = .21, p > .60$), nor of their interaction ($\beta = 0.39, SE = 0.41, p > 0.35$). The effect of affectedness was marginally significant, indicating that participants were perhaps more likely to use the location-first form with stimuli that were independently judged to show a higher degree of affectedness ($\beta = .04, SE = .02, p < .06$). This is the direction expected by the semantic theory, where location-first forms convey an entirely affected location object.

General Discussion

This paper tested two pressures on speakers' incremental production choices: accessibility of lexical items and a predicted pressure to be mention informative material early.

Foregrounding objects in a visually presented scene made them more accessible and increased their rate of early mention, but only when the speaker did not directly address a listener (Exp. 1). When the speaker produced utterances in a conversation with a partner (Exp. 2), we did not find a significant effect of foregrounding. These results are consistent with Good-Enough Production, whereby speakers balance message alignment and word accessibility. One possible explanation for the accessibility effect arising only in Exp. 1 is that speakers did not have a salient target audience. Given the subtle nature of the meaning distinction between *spray-load* forms, speakers may have been less committed to producing a particular meaning: both utterances were consistent with possible outcomes of "what Sally will do today." In contrast, participants in Exp. 2 conversed with another human: in this context, speakers may have been more careful to produce an utterance in line with their intended meaning. The effect of object affectedness in Exp. 2 provides supporting evidence for this possibility. Moreover, participants in Exp. 2 had a specific communicative goal for which foregrounding was not relevant. Lastly, foregrounding may simply be more salient when viewing one scene instead of three.

We also tested whether speakers produced more informative words earlier in their utterances. We did not find evidence for such an effect despite being predicted by an IRSA model (Cohn-Gordon et al., 2019). These results indicate that, to capture the empirical facts of incremental production, IRSA models may need to plan over larger chunks, e.g., at a grain of entire verb phrases. Under such a model, both utterance alternatives are equally informative in our two "informativity" conditions, predicting no condition effect. This approach is compatible with work suggesting that speakers plan utterances in larger units than single words (Lee et al., 2013; Martin et al., 2010; Momma & Phillips, 2018).

The takeaways from this work are two-fold. First, even when two sentences are not entirely meaning-equivalent, production choices can be affected by lexical accessibility, echoing recent work on Good-Enough theories of production. Second, extending RSA models to capture incremental speech production may have to take into account that informativity-based utterance planning is likely not performed one word at a time but in larger chunks. However, given that this conclusion is based on a null result in Exp. 2, more empirical work is needed to assess incremental informativity considerations in sentence planning.

Acknowledgements

We are grateful for extensive and helpful feedback from Beth Levin, Jiayi Lu, Brandon Waldon, Penny Pan, Shota Momma, and members of the Stanford ALPs Lab. This research received funding support from the Stanford Institute for Human-Centered Artificial Intelligence (HAI).

References

Almaatouq, A., Becker, J., Houghton, J. P., Paton, N., Watts, D. J., & Whiting, M. E. (2021). Empirica: a virtual lab for high-throughput macro-level experiments. *Behavior Research Methods*, *53*(5), 2158–2171. doi: 10.3758/s13428-020-01535-9

Anderson, S. R. (1971). On the Role of Deep Structure in Semantic Interpretation. *Foundations of Language*, *7*(3), 387–396.

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language*, *76*(1), 28. doi: 10.2307/417392

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

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

Buck, R. A. (1993). Affectedness and Other Semantic Properties of English Denominal Locative Verbs. *American Speech*, *68*(2), 139. doi: 10.2307/455674

Cohn-Gordon, R., Goodman, N., & Potts, C. (2019). An Incremental Iterated Response Model of Pragmatics. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019* (pp. 81–90). (Type: 10.7275/cprc-8x17) doi: 10.7275/cprc-8x17

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review*, *127*(4), 591–621. doi: 10.1037/rev0000186

de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jspsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351. Retrieved from `https://doi.org/10.21105/joss.05351` doi: 10.21105/joss.05351

D'Elia, S. C. (2016). *The Spray/Load and Dative Alternations: Aligning VP Structure and Contextual Effects.* Thesis (Doctor of Philosophy (PhD)), University of Kent.

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

Ferreira, V. S., & Dell, G. S. (2000). Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology*, *40*(4), 296–340. doi: 10.1006/cogp.1999.0730

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological Influences on Lexical (Mis)Selection. *Psychological Science*, *14*(1), 86–90. (Publisher: SAGE Publications Inc) doi: 10.1111/1467-9280.01424

Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998. doi: 10.1126/science.1218633

Futrell, R. (2023). Information-theoretic principles in incremental language production. *Proceedings of the National Academy of Sciences*, *120*(39), e2220593120. doi: 10.1073/pnas.2220593120

Futrell, R. (2024). An Information-Theoretic Account of Availability Effects in Language Production. *Topics in Cognitive Science*, *16*(1), 38–53. doi: 10.1111/tops.12716

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of memory and language*, *57*(4), 544–569.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic Language Interpretation as Probabilistic Inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. doi: 10.1016/j.tics.2016.08.005

Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond, E. A. Moravcsik, & J. Wirth (Eds.), *Typological Studies in Language* (Vol. 17, p. 209). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/tsl.17.16gun

Jeffries, L., & Willis, P. (1984). A return to the spray paint issue. *Journal of Pragmatics*, *8*(5-6), 715–729. doi: 10.1016/0378-2166(84)90007-9

Kim, M. (1999). A cross-linguistic perspective on the acquisition of locative verbs. *ProQuest Dissertations and Theses*, 232. (Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2023-02-22)

Koranda, M. J., Zettersten, M., & MacDonald, M. C. (2022). Good-Enough Production: Selecting Easier Words Instead of More Accurate Ones. *Psychological Science*, *33*(9), 1440–1451. doi: 10.1177/09567976221089603

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

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago press.

Martin, R. C., Crowther, J. E., Knight, M., Tamborello Ii, F. P., & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*(2), 177–192. doi: 10.1016/j.cognition.2010.04.010

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

Momma, S., & Phillips, C. (2018). The Relationship Between Parsing and Generation. *Annual Review of Linguistics*, *4*(1), 233–254. doi: 10.1146/annurev-linguistics-011817-045719

Oldfield, R. C., & Wingfield, A. (1965). Response Latencies in Naming Objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273–281. doi: 10.1080/17470216508416445

Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes*, *28*(9), 1323–1349. (Publisher: Routledge) doi: 10.1080/01690965.2012.682072

Waldon, B., & Degen, J. (2021). Modeling cross-linguistic production of referring expressions. In *Proceedings of the Society for Computation in Linguistics* (Vol. 4). (Publisher: University of Massachusetts Amherst) doi: 10.7275/VSFN-T057

Wasow, T. (2002). *Postverbal behavior* (No. no. 145). Stanford, Calif: CSLI.