

Chapter 7

Tell me something I don't know: Speaker salience and style affect comprehenders' expectations for informativity

✉ Vilde R. S. Reksnes, ✉ Alice Rees, ✉ Chris Cummins &

✉ Hannah Rohde

University of Edinburgh

A comprehender's estimate of what events or situations are typical in the world is distinct from their estimate of what a speaker is likely to report on. Comprehension and production studies have shown contradicting preferences for which types of estimates are used by comprehenders and by speakers: Typicality is favoured in comprehension (e.g. real-world typical content is associated with processing ease), whereas speakers' production choices favour the inclusion of surprising or informative content (i.e. easily inferable or typical content is disfavoured). We posit that comprehenders are aware of and make use of speakers' production preferences when anticipating upcoming content. In two studies, we elicit sentence completions as an index of comprehenders' expectations about upcoming material and evaluate the informativity of these completions (their object typicality, presence of modification or negation, and information theoretic entropy and relative entropy scores). Experiment 1 manipulated the salience of the speaker and found that increased emphasis on the speaker led to an increase in informativity, showing that the more aware comprehenders are made of an intentionally communicating speaker, the more their expectations favour upcoming words that would yield an informative utterance. Experiment 2 further tested the malleability of this informativity bias by familiarising participants with two speakers who differ in the informativity of their utterances. When completing utterances from the two speakers, comprehenders provide more informative completions for the high-informativity speaker, showing that comprehenders are able to adapt their expectations for informativity to individual speakers' communicative styles. This sensitivity to speakers' production preferences highlights a role for informativity-driven reasoning about the speaker in models of language processing.



1 Introduction

The question of how comprehenders generate expectations regarding upcoming content has received much attention in the field (see e.g. Kuperberg & Jaeger 2016 for a review). In a discourse context, expectations about upcoming content can be understood to reflect what comprehenders believe speakers will choose to talk about. On the one hand, speakers can choose to use language to talk about the world that they encounter, producing utterances that describe the kinds of situations that arise in their daily lives. On the other hand, speakers can choose to convey content that is worth talking about, producing utterances that they think a listener will find interesting and newsworthy.

The first choice characterises what we will call *transparent* language use, whereby speakers use language to directly narrate what happens to them. If speakers use language transparently, comprehenders ought to come to expect language that describes the kinds of situations that arise frequently and plausibly in the real world. The second choice characterises *filtered* language, whereby speakers limit their utterances to those whose content is novel and informative, inducing listeners to expect descriptions of situations which are interesting by virtue of their infrequency. If speakers do indeed filter available content and do so in ways to achieve informativity-driven communicative goals, and if comprehenders are aware of a speaker's role in this filtering, a prediction arises that a context that increases the emphasis on the speaker (in their role as a filter of potential content to convey) should increase comprehenders' expectations for content about infrequent situations. We will argue that prior work on comprehenders' expectations often focuses on expectations about transparent language use, missing an opportunity to find evidence of expectations that are driven by comprehenders' awareness of the speaker as someone engaged in intentional communication with informativity-driven goals.

A long-standing claim is that when a comprehender makes guesses about what words are coming next, they rely on their knowledge about the world. Studies have shown that knowledge of what is typical of real-world situations is active during language processing and thus typicality has been linked to processing ease in comprehension. Here, we define typicality as the frequency of a situation or event given the context (e.g. trains are typically present at a train station, a beach typically has sand). As an example of comprehenders' anticipation of upcoming content, Kamide et al. (2003) use a visual world eye-tracking methodology and show that participants expect utterances to convey real-world typical content. For example, for an image depicting a man, a motorbike and a carousel,

participants hear *The man will ride the...* and look to the motorbike before the noun is uttered, anticipating a continuation about the most real-world typical object in the scene (the motorbike). That work is in keeping with findings about the processing costs associated with encountering surprising events. For example, Kutas & Hillyard (1980) show that strong semantic incongruity caused a peak in the N400 relative to moderate incongruity (*taking a sip from the transmitter* vs *taking a sip from the waterfall*). Measures of sentence recall (Marks & Miller 1964) and reading with eye-tracking (Morris 1994) corroborate these effects.

There are studies showing that real-world typicality can be overridden. For example, Nieuwland & Van Berkum (2006) showed processing ease for a real-world non-typical event, namely a peanut singing about being in love, compared to a real-world typical event, namely a peanut being salted (see also Troyer & Kutas 2020). However, the important thing to note here is that comprehenders adjusted their expectations to fit the constraints of a fictional world, one where peanuts sing about their new girlfriends and their amorous feelings towards them. In other words, they still show a bias towards the typical, but this bias is governed by what is typical in the new “real world”. Overall, these studies suggest that comprehenders have a preference for typicality when they are processing language, i.e. a bias towards transparent language use.

Conversely, research on production shows that speakers do impose a filter on their language use: They tend to make production choices that favour the inclusion of surprising or otherwise informative content about a situation (e.g. Brown & Dell 1987, Lockridge & Brennan 2002). In doing so, they show a dispreference for real-world typical content: Speakers will often choose to omit altogether information that is inferable or typical, and instead highlight non-typical information. In Brown & Dell’s (1987) classic production study, participants read short stories that involved either a typical instrument or an atypical instrument, for example, a stabbing that happened either with a knife or an icepick. Participants were then asked to retell the stories to test whether they chose to mention the instrument in their event description. The results showed a preference to omit mention of the instrument if it was typical (knife being a typical stabbing instrument) and to include a mention of the instrument if it was atypical (icepick being an atypical stabbing instrument). In keeping with this pattern whereby inferable or typical content can be omitted or reduced, there is evidence that high frequency words are produced with shorter acoustic duration than low frequency words (Aylett & Turk 2004; see also Levy & Jaeger 2007, Jaeger 2010, Kravtchenko 2014, Lemke 2021). In other contexts, speakers are shown to include colour adjectives more often when referring to an object if the colour is atypical, such as a pink banana, than if it is typical (Sedivy 2003, Westerbeek et al. 2015, Rubio-Fernández

2016, Degen et al. 2020), and to mention the material of an object when atypical, such as a wool bowl compared to a ceramic bowl (Mitchell et al. 2013). Notably, Degen et al. (2020) found that speakers are less likely to mention the colour of an object if it is typical (e.g. a yellow banana) even when the colour is necessary for disambiguating the intended referent (i.e. when a brown banana is also present). The authors take this to indicate that the speakers are relying on the listener to infer that the most typical object was intended. In contrast, if the target object was an atypical colour (a blue banana, in this case), the colour was always mentioned when a competitor object was present.

The pictures that emerge from these two bodies of literature thus seem to conflict. Evidence from comprehension studies strongly suggests that comprehenders estimate upcoming content with a bias in favour of sentences about typical situations (transparent language use), while results from production studies demonstrate a preference for informativity or newsworthiness in content selection (filtered language use). According to this literature, it would therefore appear as if comprehenders' expectations are out of line with the predicted behaviour of speakers according to Gricean accounts; that is, speakers' contributions are expected to be appropriately informative and relevant (Grice 1975). However, if comprehenders are rational and sensitive to speakers' production preferences, these preferences should also bear upon their estimates regarding upcoming content: That is, comprehenders should expect the kind of content that cooperative speakers are likely to mention, rather than the kind of content that is likely to be the case in the real world. For example, you would probably expect there to be trains in a train station but generally not expect someone to tell you that fact out of the blue. However, if you receive a call from a friend who is at your local train station and they tell you the Hogwarts Express has just pulled up, you would be surprised by the fact that this particular train has made a visit to your local station, but likely not by the fact that your friend chose to communicate the information, given that the situation occurred.

On this view, a comprehender's estimate of what is typical in the world is distinct from their estimate of what the speaker is likely to say (Rohde et al. 2021). In principle, either of these estimates might be activated when a comprehender makes guesses about an upcoming utterance. In the current study, we examine the nature of comprehenders' predictions by using a sentence completion paradigm (a so-called Cloze task; Taylor 1953). We use this task to elicit specific sentence completions to assess what words comprehenders predict a speaker will produce, rather than testing whether comprehension is speeded/slowed by available content, as in previous work. Although the Cloze task involves a participant

producing words, it can also be understood as a comprehension-oriented process, in that the task probes what a comprehender thinks is coming next. This approach follows work on coreference that often uses sentence completion tasks to estimate comprehenders' interpretation biases and their expectations about next mention (e.g. Stevenson et al. 1994). In fact, the original Cloze study describes the task in comprehension terms as an "attempt to reproduce accurately a part deleted from a 'message' (any language product) by deciding, from the context that remains, what the missing part should be" and notably describes the participants as readers/listeners (Taylor 1953: 416).

In Experiment 1, we manipulate the emphasis on the presence of a speaker to test whether participants' estimates are modulated by how salient the speaker is. If this manipulation is sufficient to make participants consider the speaker's production process, their guesses about upcoming content should be influenced by their inferences about what the speaker's goals are and why they are choosing to speak. For example, completing the sentence *I'm at the train station, and there's _____* with the words *a train* yields a description of a likely situation, in that a scenario with a train would be a typical occurrence given the location. However, if comprehenders consider what a speaker would deem worth reporting on, they might not expect such an utterance, precisely given the high typicality of the scenario described. The presence of a train is so typical as to be inferable without mention, so a comprehender might assume that a speaker would not choose to mention it. Consequently, something more informative such as *street performers* or a similarly low-typicality occurrence might be a more likely completion when the presence of a speaker is emphasised.

An emerging line of research shows that comprehenders do keep track of speakers' production preferences. Rohde et al. (2021) established the presence of informativity-driven effects in comprehension, showing that sentences containing newsworthy content were preferred in certain contexts over those that contained real-world typical content. In particular, they observed processing ease for atypical content in a natural dialogue setting with no contextual manipulations; participants were faster reading a newsworthy message about socks that cost \$100 than a message about more standardly priced \$2 socks. Rohde et al. take this to demonstrate that a communicative context with an intentionally communicating speaker behind an utterance is sufficient to induce a comprehension preference in favour of newsworthy content. Their studies show that comprehenders experience difficulty in integrating inexplicably uninformative content, but do not test what content comprehenders specifically expect or what factors might modulate such expectations. The current study aims to examine the nature of this predicted content in a more fine-grained way.

Corroborating the findings in Rohde et al. (2021), Rohde et al. (2022) conducted a series of forced-choice tasks in which participants were asked to estimate what a character in a story would say or what they might think; they found that the character's choice to produce an utterance unprompted, as opposed to responding to a question or thinking uncommunicated thoughts, cued participants to estimate a more atypical meaning, demonstrating that comprehenders favour sentences that convey newsworthy information. In fact, when comprehenders come across content that does not meet this standard for newsworthiness, they try to reconcile the absence of sufficiently informative content by recasting the utterance in a way that makes it an informative contribution (Kravtchenko & Demberg 2022). In Kravtchenko & Demberg's experiment, participants read short stories about stereotyped activities. In general, comprehenders' script knowledge about such activities allows speakers to omit typical or inferable information. Upon encountering an utterance about an easily inferable event, e.g. *Mary ate* in the context of going to a restaurant, participants made inferences that this generally typical action was nonetheless unusual for the protagonist. In other words, the study highlights cases in which a discrepancy emerges between the understanding of a speaker as having goals to be cooperatively informative and the uncooperative typicality of the situation being communicated. This discrepancy can be resolved if comprehenders change their prior beliefs about what was typical in this context, thereby ensuring that the utterance meets expectations for informativity.

Relatedly, Lemke et al. (2021b) make the point that script knowledge provides a way to quantify the likelihood of upcoming content as it may approximate extralinguistic knowledge, which is used when language users make predictions about upcoming events and, thereby, content. They argue for the importance of such knowledge when modelling anticipation, specifically for the production of ellipsis and fragments (Lemke et al. 2021a, Schäfer et al. 2021). The extralinguistic context is particularly relevant discourse-initially when linguistic context is not available to inform comprehenders' anticipations. Venhuizen et al. (2019) similarly argue that world knowledge has not been hitherto sufficiently incorporated in computational models of comprehension. They develop a model that derives a preference for informative content in processing. The studies presented here expand on such work by assessing the effect of speaker awareness, another aspect of extralinguistic knowledge, on comprehenders' expectations.

Awareness of the speaker can be understood to underlie comprehenders' guesses about the reason behind a speaker's production choice. In particular, there is a longer standing body of work that assesses comprehenders' guesses about

the reason behind a speaker's production choice, and in particular what referential expressions speakers choose when tasked with picking out a specific object among a set of possible referents (see e.g. Davies & Arnold 2019 for a review). Many of these studies show comprehenders reasoning specifically about the informativity of a speaker's utterances (e.g. Grodner & Sedivy 2011, Pogue et al. 2016, Ryskin et al. 2019). For example, if a speaker reliably uses adjectives contrastively (e.g. produces *click on the tall glass* when there is another glass present), participants are able to use that information to identify the referent more quickly; however, participants stop relying on the adjective to indicate contrast if a speaker habitually uses adjectives non-contrastively (e.g. produces *click on the tall glass* even when there is only one glass; Grodner & Sedivy 2011). Although this work shows comprehenders adjusting their expectations to different speakers with regards to informativity, these are specifically referential expectations; the pragmatic reasoning does not tap into expectations in the way we outline above, namely expectations for the selection of sentence content. Instead, this work tends to focus on tasks in which the communicative goal is referential success – the comprehender knows that the speaker has an intended referent they need to talk about and the speaker's choice lies in what forms to use for describing the referent. A comprehender can thus use their awareness of the speaker's referential communicative goal when interpreting the message (Sedivy 2003).

In contrast, the studies presented here test expectations for message content when an utterance is conveyed out of the blue, i.e. when content newsworthiness is more likely to be the goal rather than referential identification within a larger utterance. In this way, we aim to assess comprehenders' expectations for informativity more generally: What content do comprehenders expect cooperative speakers to mention? The current study also departs from the reference resolution studies discussed above in how the effect of speaker identity is manipulated. Past work on reference has examined this variable by contrasting two types of speakers, those who are pragmatically reliable and those who are unreliable. While our Experiment 2 implements an analogous manipulation to this, contrasting an informative and an uninformative speaker, our Experiment 1 investigates a more fundamental contrast by comparing the presence vs. absence of a speaker in order to test whether comprehenders' awareness of the speaker can itself influence their expectations about content newsworthiness.

There are current models of language use that, in line with the Gricean approach, take into account listeners' expectations for speakers to make informative contributions. For example, the Rational Speech Act (RSA) framework models a speaker who makes production choices based on a listener who in turn is

able to make inferences about the speaker and their communicative goals (see e.g. Goodman & Frank 2016 for an overview). The RSA model captures the interplay between speakers and listeners and how they reason about each other's linguistic choices. It follows that the speaker plays a key role when a listener processes an utterance. This point is tested explicitly by Kreiss & Degen (2020), who build an RSA model based on empirically elicited referring expressions from speakers to predict listeners' behaviour in a subsequent referential identification experiment with or without contrast objects present (e.g. another banana when the target is a yellow banana). They show that when listeners hear partial referring expressions (e.g. *click on the yellow*) they make inferences about how likely a speaker is to use a modifier based not only on the presence of a contrast object, which is what other contrastive inference studies have consistently found, but also on the typicality of a competitor (e.g. a yellow strawberry). Their RSA model accurately predicts this listener behaviour. In other words, listeners identify the intended referent based not only on linguistic or cognitive factors related to the referents themselves, but also based on the listener's reasoning about the speaker's production probabilities. Note that again, this study is one where referential success is the goal of the communicative interaction. The current work is consistent with the principles of RSA, and adds to this framework by explicitly testing the salience of the speaker and its effect on comprehenders' expectations for content.

In sum, whereas a range of prior studies have emphasised the importance of (real-)world typicality, an emerging body of work aims to bring attention to the role informativity plays in processing. Our study adds to this informativity-driven account of processing, and shows that expectations for informativity also matter for the estimates comprehenders make regarding upcoming content. In anticipating what someone will say next, comprehenders may be tracking several distributions of probabilities about the speaker, such as the probability of particular situations that a speaker might encounter, the likelihood of choosing to formulate an utterance about that situation (as opposed to staying silent), and even the choice among available formulations for expressing that meaning. On this view, anticipating upcoming content must be understood to rely on (at least) two components: One concerns the situation and one concerns the speaker and how and whether they will filter the scene. We argue that in order to accurately capture the processes at play, there needs to be a role for informativity-driven reasoning about the speaker in models of language processing.

2 Experiment 1: Awareness of speaker intentions in comprehension

Our main thesis, in keeping with the findings discussed above, is that comprehenders can and do take into account speakers' production preference to be informative during processing. This in turn means that certain contexts may emphasise the relevance of speakers' production preferences because they highlight the presence of the speaker themselves. The prediction is that increasing the salience of an intentionally communicating speaker should lead to an increase in comprehenders' expectations for informative content. Experiment 1 tests this claim using a Cloze task paradigm to elicit sentence completions as an indicator of what the participant expects a speaker would say. Specifically, we ask participants to complete statements about what is present at a particular location, e.g. a train station. A total of 20 different locations were used.¹ In a between-participants design we manipulate the salience of the speaker across four conditions. As shown in Figure 1, all conditions use the *there's ____* prompt. The BARE condition only mentions the location (e.g. *At the train station, there's ____*); the THIRD PERSON condition describes someone at the location (*They're at the train station, and there's ____*); the FIRST PERSON condition directly mentions the speaker (*I'm at the train station, and there's ____*); and the VISIBLE SPEAKER condition uses first person and adds a photograph of a person speaking, with the text prompt being identical to the FIRST PERSON condition. The manipulation is intended to vary the perceived communicative intent of the context, such that our most speaker-salient condition, VISIBLE SPEAKER, is the one where communication is most strongly foregrounded by emphasising the presence of a speaker who has uttered the sentence. The expectation is that such contexts will elicit more completions about the presence of non-typical or otherwise unexpected entities in the target location.

If speaker salience influences comprehenders' guesses, the least speaker-salient condition is predicted to reveal comprehenders' reliance on real-world knowledge² of typical entities or situations that are likely to be present at the different locations, in keeping with prior comprehension studies demonstrating comprehenders' reliance on real-world knowledge (e.g. *a train* for the train station location). We in turn expect an increase in informative completions across the four

¹The locations are: Bakery, library, forest, living room, beach, petrol station, cinema, office, playground, mountain, bathroom, park, bedroom, post office, train station, garden, golf course, restaurant, hospital, and stationery shop.

²Barring any construction of a fictional world, we assume comprehenders tend to rely on their real-world knowledge for their typicality estimates.



Figure 1: Example target sentences for the train station location in each of the four conditions

conditions, with the most informativity expected in the VISIBLE SPEAKER condition (e.g. *a celebrity*). The possibility that the less speaker-salient conditions will be distinguished from each other is predicted by work in embodied cognition showing that people are sensitive to the use of different pronouns when taking perspective; people adopt an actor's perspective when first person is used, and an external perspective with the third person (Borghi & Cimatti 2010). Our goal in including a gradation of speaker salience conditions is to test whether – and via what properties – an emphasis on the speaker can be achieved. These manipulations themselves are less of interest than the possibility that, if a speaker-oriented perspective can be achieved, comprehenders' guesses of what that speaker will say next may favour utterances conveying more informative content. We will be using a set of measures to capture different senses of informativity (see measures and data analysis below). An increase in informativity would indicate that rather than simply invoking typicality when they anticipate upcoming content, comprehenders make use of their awareness of speakers' production preferences and filter the possible options through the lens of the speaker – in that way estimating what content a cooperative speaker would consider worth uttering.

2.1 Methods

2.1.1 Participants

400 participants with English as their first language, no known language impairments, location in the US or the UK, and of minimum 18 years of age ($M = 36.09$, $SD = 11.73$, range = 18–78), were recruited on Prolific (<https://www.prolific.com/>). We tested an equal number of participants in each of the four conditions, i.e. 100 participants per condition, each providing 20 data points. Participants who failed two or more attention checks or reported another language than English as their first language were excluded and more participants were recruited to reach the desired number per condition. Participants were paid £7.12/hour on average.³ In addition, 22 participants were recruited through the University of Edinburgh's Experiment Volunteer Panel for a pre-test, used to assess the typicality of a variety of objects in the item locations (see measures). These participants were students who received course credits as compensation.

2.1.2 Design

Experiment 1 tests whether speaker salience influences comprehenders' expectations about sentence completions. In order to implement this speaker salience manipulation, each of the four conditions increasingly highlight the speaker, as exemplified in Figure 1. The **BARE** condition only mentions the location; the **THIRD PERSON** condition invokes a speaker talking about someone at the location; the **FIRST PERSON** condition directly mentions the speaker; and the **VISIBLE SPEAKER** condition uses first person and adds a photograph of a person speaking, with the sentence embedded in a speech bubble. This last condition is intended as the one where communicative intent and therefore awareness of speakers' production preferences is most emphasised.

Participants in the **BARE**, **THIRD PERSON** and **FIRST PERSON** conditions are not given any information about the communicative context, whereas participants in the **VISIBLE SPEAKER** condition are told that each utterance is the beginning of a phone call where the speaker has called someone to tell them something. In other words, the speaker has not been prompted by a question and is speaking out of the blue.

³The details of the cost were established by Prolific based on the median time taken to complete the experiment per submission for each group.

2.1.3 Materials

The 20 target items each mentioned a different location (see Footnote 1). Locations were chosen to be places that would have an adequate number of typical and therefore expected objects present (e.g. trains, platforms, passengers, etc. for train station) but also allowed for the possibility of less typical entities being mentioned without raising doubts about the speaker's reliability.

Each target sentence followed the same format within a condition, mentioning a location in the first clause and providing a sentence-final blank space for the participant to fill. The target sentence templates for each condition were as follows: BARE: *At the [location], there's ____*, THIRD PERSON: *They're at the [location], and there's ____*, FIRST PERSON and VISIBLE SPEAKER: *I'm at the [location], and there's ____* (all exemplified with *train station* in Figure 1). The VISIBLE SPEAKER condition additionally displayed the target sentence in a speech bubble alongside a picture of a person talking on the phone. A total of 20 different speaker images were used, 10 portraying female speakers and 10 portraying male speakers. Speaker images were allocated to specific items, but participants only saw each speaker for a single target item (e.g. the man in Figure 1 is seen talking about the train station location). The speaker images were reused in the fillers, so that a given list in the VISIBLE SPEAKER condition contained the same speaker image a total of three times. The experiment included 40 fillers in each condition. For BARE, THIRD PERSON and FIRST PERSON conditions, fillers were adapted Cloze task items from Altarriba et al. (1996) with the blank space appearing either initially, in the middle or at the end of the sentence. For VISIBLE SPEAKER, fillers were created that were more natural-sounding as conversation-initial utterances. Two fillers in each condition served as catch trials as they had only one or two likely completions, e.g. *I'm going to the swimming ____ this afternoon* in VISIBLE SPEAKER condition, where *pool* is the expected completion. See OSF for a full list of materials for the VISIBLE SPEAKER condition.⁴

2.1.4 Procedure

The experiment was presented using Qualtrics (<https://www.qualtrics.com/>). For BARE, THIRD PERSON and FIRST PERSON conditions, participants were told they would read sentences where a part was missing and their task was to type in the word or words they thought should be in the sentence. We suggested an upper limit of 3 words in the instructions at the beginning of the experiment with the aim of discouraging participants from expanding the sentence beyond a

⁴<https://osf.io/7h5qs/>

single target noun phrase. For VISIBLE SPEAKER, the instructions also emphasised the communicative intent of the sentences by telling participants that they were seeing someone who had just called up a friend and their task was to fill in the missing part with what they thought the speaker could have said. They were also instructed that what they saw was the first thing said in the conversation, to further highlight that these are unprompted utterances. Items were presented in fully randomised order. Participants were presented with one item at a time with a text field below the item stating *Fill in the blank*, in which they typed in their completions. They then clicked an arrow at the bottom right of the screen to proceed to the next item. Any one participant saw only one condition, i.e. a total of 60 items (20 of which were target items). The median completion times in each condition were between 10 and 15 minutes.

2.1.5 Measures and data analysis

Prior to data analyses, responses were standardised, i.e. manually cleaned to have matching case, punctuation and spelling, initial articles (*a/an/the*) were removed, and any spelling differences were collapsed (e.g. *doughnut/donut*). Where participants listed more than one response (e.g. *doctors/nurses*), only the first item mentioned was included. For each response, the main noun was identified as well as any modification and negation used. A total of 10 continuations were non-meaningful and therefore excluded from analysis. The continuations included ranged from 1 word in length to 11 words, with a mean of 1.61.

The pre-test was a non-communicative task in which a separate group of participants ($N = 22$) were asked to list a minimum of 3 and maximum of 10 things they would be likely to find at each of the locations mentioned in the target sentences. The pre-test provides a measure of what is considered typical at the different locations. Main nouns were extracted from responses and collated into ranked lists, ranging from highly expected/common (e.g. *train* for *train station*) to less expected/common (e.g. *bustle, musician, pigeon*). Each noun in a given location was then given a typicality score estimated as the proportion of pre-test participants who mentioned that noun (e.g. for the train station location, 0.91 and 0.00 for *train* and *delay*, respectively).

The informativity of participants' completions is assessed with five measures. Each of these measures is intended to capture a different sense of how a continuation provided in the experiment may be informative:

1. Variability of responses (mean entropy score per condition to compare consistency versus unpredictability of responses)

2. Kullback-Leibler (KL) divergence of responses (i.e. relative entropy of responses with BARE condition as baseline)
3. Inclusion of modification (which may make otherwise typical content like a train more newsworthy, e.g. *steam train*)
4. Inclusion of negation to mark the absence of something (often something typical, e.g. *no train*)
5. Typicality of objects mentioned (compared to responses elicited in the pre-test; specifically, how many participants in the pre-test listed object X for location Y)

Entropy is an information-theoretic measure for quantifying the variation in a distribution of outcomes. As such, entropy provides a measure of the average amount of information needed to represent an outcome drawn from a given distribution. In a condition where participants' sentence continuations are strongly biased to a few possible outcomes, the entropy for that distribution of outcomes will be low (i.e. a given outcome is, on average, not very informative as it is highly predictable). Conversely, in a condition with more unique outcomes or more variability across outcomes, the entropy for the distribution will be high (i.e. a given outcome is, on average, very informative as it is difficult to predict).

Entropy scores were calculated using participants' full text responses.⁵ First, an entropy score was calculated for each location in each condition (e.g. for train station in VISIBLE SPEAKER condition), from which we derived a mean entropy score per condition. By testing an equal number of participants (=100) per condition, we avoid a concern that the probability computations which contribute to the entropy scores would be distorted by unequal sample sizes. Such distortion could arise, for example, if one condition had fewer responses than another, such that the singleton responses (those produced only once) would be assigned higher probability in the small-sample-size condition than singleton responses in a condition with more responses.

KL divergence was also calculated using participants' full text responses. Relative entropy provides a measure of comparison between two probability distributions. Comparing the distribution of responses from each condition relative to the BARE condition can therefore be seen as a proxy for comparing comprehenders' estimates of what objects or events they expect a speaker to mention versus what objects or events are expected to be present in a location. Relative

⁵Where x is a particular response provided by one or more participants (e.g. *a train*), we computed entropy as follows for each location and each condition: Entropy = $\sum_x p(x) \log(p(x))$.

entropy complements the entropy measure described above, which only characterises a single distribution rather than drawing a comparison between the response distribution between conditions. Also, using the entropy measure alone may risk missing potential differences between distributions; two distributions could have similar entropy but very different properties (e.g. the BARE condition for the train station location might favour *train* with the same probability that another condition favours *delay*; in that case, if the remaining alternatives have similar probabilities, the two distributions would have similar entropy and any differences in distributions would remain undetected). Since all responses did not appear in all conditions, responses with a zero probability in BARE condition (i.e. responses that were not provided by any of the participants in this condition but did occur in one or more of the other conditions) were given a +1 smoothing. We opted to calculate relative entropy over the top 4 most frequent responses and bin the remaining responses in *Other*. This method avoids giving undue weight to any of the responses that were not observed in, for example, the BARE condition (i.e. responses that were smoothed), which may only have a very low probability of being observed in, for example, the VISIBLE SPEAKER condition. That is, the *Other* bin ensures that no non-occurring response would be given the same probability as a low-probability response.

A speaker's use of modification and negation in a continuation provides a way of adding more information. For the analyses of modification and negation, the presence of either element was manually coded as present (1) or not (0).

Typicality is about what is surprising and atypical for a given situation. For the typicality measure, the main noun in each response was checked against main nouns extracted in the pre-test data set and assigned a score based on the popularity rank of that response in the pre-test.

2.2 Results

As predicted, VISIBLE SPEAKER yielded the most informative completions on all five measures, as shown in Figures 2 to 5. For an illustration of the variation in responses across the 4 conditions, see OSF for figures displaying the distribution of responses in the train station location.⁶

The mean number of distinct responses per location in each of the 4 conditions were as follows: VISIBLE SPEAKER: 70.5, FIRST PERSON: 47.1, THIRD PERSON: 49.0, and BARE: 39.6. Entropy scores were analysed with paired Wilcoxon signed-rank tests across the items. Completions showed higher entropy for VISIBLE SPEAKER

⁶<https://osf.io/7h5qs/>

(mean: 4.00), FIRST PERSON (mean: 3.14) and THIRD PERSON (mean: 3.23) conditions compared to the baseline BARE (mean: 2.80; $p < 0.001$ for all three pairwise comparisons), and also the entropy of VISIBLE SPEAKER condition was higher than that of FIRST PERSON ($p < 0.001$), as seen in the left panel of Figure 2. Although there is a numeric difference between FIRST PERSON and THIRD PERSON conditions in the opposite direction to the prediction, this difference was not significant ($p = 0.17$). However, the difference between VISIBLE SPEAKER and the numerically adjacent THIRD PERSON is significant ($p < 0.001$). In order to test whether the difference in entropy was independent of use of modification or negation in participants' responses, we also conducted the analysis on the subset of responses not containing any modification and negation.⁷ The above pattern is somewhat altered, with VISIBLE SPEAKER still showing the highest entropy (mean: 2.70) but the remaining conditions showing a different descending order across THIRD PERSON (mean: 2.60), followed by BARE (mean: 2.40), and then lowest entropy in FIRST PERSON condition (mean: 2.33). VISIBLE SPEAKER differs significantly from BARE and FIRST PERSON, but only numerically from THIRD PERSON. There is a significant difference between BARE and FIRST PERSON, but none between BARE and THIRD PERSON.

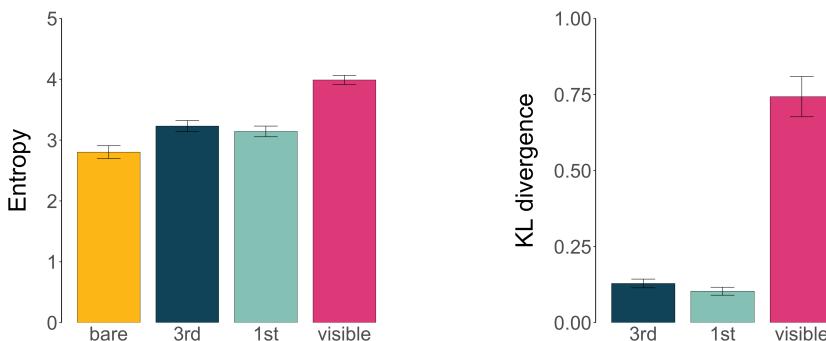


Figure 2: The left panel shows mean entropy for each condition; the right panel shows KL divergence as a comparison between the distribution in the listed condition and that of the baseline BARE condition.

⁷Thank you to an anonymous reviewer for suggesting this analysis. We note that while the subset analysis eliminates the matching number of responses across conditions, it eliminates more responses from the VISIBLE SPEAKER condition than any other, where the most use of modification and negation was registered. If higher entropy were to emerge simply due to the presence of more singleton unique responses, the subset analysis reduces the chance of seeing such an increase in VISIBLE SPEAKER. Hence, we do not believe this introduces a confound where analysing this subset would inadvertently yield higher entropy in the predicted VISIBLE SPEAKER condition.

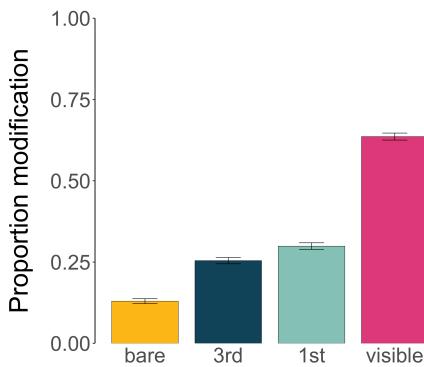


Figure 3: Proportion of use of modification across conditions

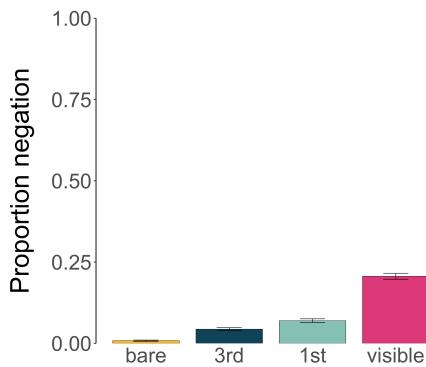


Figure 4: Proportion of use of negation across conditions

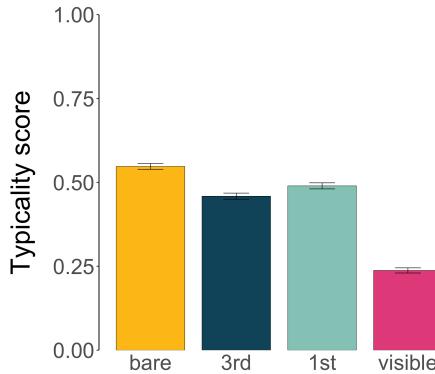


Figure 5: Mean typicality score of responses across conditions

The results for relative entropy confirm a role for speaker salience, as seen in the right panel of Figure 2⁸: Across items, the KL divergence between BARE and VISIBLE SPEAKER conditions (mean: 0.74) is greater than that between BARE and FIRST PERSON (0.10, paired *t*-test: $p > 0.001$) or between BARE and THIRD PERSON (0.13, $p > 0.001$). The comparisons between BARE and FIRST PERSON and between BARE and THIRD PERSON did not show a significant difference. As with entropy above, we also analysed the subset of responses that did not contain any modification or negation. In this case, the gradation of results stays the same: The KL divergence for BARE and VISIBLE SPEAKER (mean: 0.44) is higher than BARE and

⁸See OSF for examples of a list of the binned responses for each of the locations across conditions.

FIRST PERSON ($0.08, p > 0.001$), as well as than BARE and THIRD PERSON ($0.12, p > 0.001$). However, we also see a significant difference in the comparison between BARE and FIRST PERSON and between BARE and THIRD PERSON ($p > 0.05$). In sum, the distribution of participants' completions in VISIBLE SPEAKER condition is the one that is most clearly different from the distribution in BARE condition.

The binary outcomes of modification and negation were analysed with generalised mixed effects models⁹ (GLMM: Jaeger 2008) using the lme4 package (Bates et al. 2015) in R (R Core Team 2022), treating condition as a fixed effect, with random slopes and intercepts of condition for participants and locations (Barr et al. 2013). In order to test whether we observe the predicted gradual increase in use of modification and negation across conditions, we forward coded the condition prior to analysis. This allows us to compare each condition to the adjacent one: BARE vs. THIRD PERSON, THIRD PERSON vs. FIRST PERSON, and FIRST PERSON vs. VISIBLE SPEAKER. To achieve convergence for the modification and negation models with maximal random effects structure, we used the optimiser optimx with the method bobyqa which led to convergence. For each fixed effect, we report a *p*-value as generated by the lmer logistic regression which is based on the Wald Z statistic (Agresti 2003). As seen in Figure 3, modification rates were highest in VISIBLE SPEAKER condition. VISIBLE SPEAKER was significantly higher than FIRST PERSON ($\beta = 2.81, SE = 0.38, z = 7.46, p < 0.001$), and THIRD PERSON was higher than BARE ($\beta = 2.46, SE = 0.47, z = 5.26, p < 0.001$). Although there is a difference between FIRST PERSON and THIRD PERSON in the predicted direction, this difference was not significant ($\beta = -0.13, SE = 0.41, z = -0.32, p = 0.75$). For negation (Figure 4), VISIBLE SPEAKER shows significantly higher proportions than FIRST PERSON ($\beta = 2.63, SE = 0.29, z = 9.01, p < 0.001$) and FIRST PERSON is higher than THIRD PERSON ($\beta = 3.78, SE = 0.85, z = 4.44, p < 0.001$); however, although numerically the predicted pattern is present for THIRD PERSON compared to BARE, this difference was not significant ($\beta = 11.18, SE = 7.60, z = 1.47, p = 0.142$).¹⁰

⁹modification~condition + (1 + condition | subjectID) + (1 + condition | location)

negation~condition + (1 + condition | subjectID) + (1 + condition | location)

¹⁰Using bobyqa produced a warning, but did not stop the models from converging. If we simplified the models by removing random slopes of condition for participant and for location, we achieve convergence without bobyqa and encountering no warnings. For modification the results stay the same (VISIBLE SPEAKER higher than FIRST PERSON ($\beta = 2.56, SE = 0.35, z = 7.29, p < 0.001$), no difference between FIRST PERSON and THIRD PERSON ($\beta = 0.19, SE = 0.36, z = 0.52, p < 0.60$), and THIRD PERSON higher than BARE ($\beta = 1.67, SE = 0.38, z = 4.44, p < 0.001$)), for negation we see significance in all three comparisons (VISIBLE SPEAKER higher than FIRST PERSON ($\beta = 2.08, SE = 0.30, z = 7.03, p < 0.001$), FIRST PERSON higher than THIRD PERSON ($\beta = 0.93, SE = 0.35, z = 2.66, p < 0.01$) and THIRD PERSON higher than BARE ($\beta = 2.54, SE = 0.72, z = 3.53, p < 0.001$)).

For the measure of typicality, we calculated a score for each main noun based on the number of pre-test participants who mentioned that object for the relevant location (meaning that a response given in Experiment 1 that was not provided in the pre-test receives a score of 0). Results are shown in Figure 5. To analyse these typicality scores we again forward coded the condition (comparing each condition to the adjacent one; BARE vs. THIRD PERSON, THIRD PERSON vs. FIRST PERSON, FIRST PERSON vs. VISIBLE SPEAKER)¹¹ and used a linear mixed effects model¹² with condition as a fixed effect and random slopes and intercepts of condition for participants and locations (Barr et al. 2013), as above. The significance of the fixed effect of condition was determined via a likelihood ratio test comparing the fit of the model to one with the same random effects structure but no fixed effect. Again, to achieve convergence with the models with maximal random effects structure, we used the optimiser `optimx` with the method `bobyqa`. We see a main effect of condition in the model comparison ($p < 0.001$), showing that the model with condition as a fixed effect significantly improved the model fit. Typicality was significantly lower in VISIBLE SPEAKER ($\beta = -0.25$, SE = 0.03, $t = -7.93$) compared to FIRST PERSON, and THIRD PERSON was lower than BARE ($\beta = -0.09$, SE = 0.03, $t = -2.86$). The difference between FIRST PERSON and THIRD PERSON was not significant ($\beta = 0.03$, SE = 0.03, $t = 1.20$).

2.3 Discussion

The results of Experiment 1 show that participants' expectations about informativity, as elicited by an utterance completion task, vary according to how and whether the speaker is made salient. In the conditions which were intended to make the speaker more salient, participants' completions score higher in informativity as shown by the five measures: Higher entropy, greater KL divergence for VISIBLE SPEAKER condition compared to the BARE condition, more use of modification and of negation, and fewer typical entities provided in the VISIBLE SPEAKER condition compared to the other conditions. We take this increase in informativity to reflect an increased awareness of the speaker, inducing participants to consider the speaker's production preferences when estimating possible utterance completions.

We find only partial support for the hypothesis that continuations gradually increase in informativity across the four conditions: The use of modification and

¹¹Since the observed means do not pattern with the predicted order, we conducted an additional analysis setting the forward contrast coding to order the conditions to match the observed order; BARE vs. FIRST PERSON, FIRST PERSON vs. THIRD PERSON, THIRD PERSON vs. VISIBLE SPEAKER (as seen in Figure 5). The result remains the same.

¹²typicality~condition + (1 + condition | subjectID) + (1 + condition | location)

negation follows the expected pattern numerically with a gradual increase from the least to most speaker-salient conditions, but there is no significant difference between THIRD PERSON and FIRST PERSON conditions. For entropy, KL divergence and typicality, the pattern is somewhat less clear: The results pattern as predicted except for in THIRD PERSON and FIRST PERSON conditions, which are mirrored numerically compared to the predicted direction. However, these differences are also not significant. Overall, we interpret this to show that the manipulation generally works in the intended way; it affects a participant's probability of using the second estimate discussed above, i.e. one that takes into account a speaker's production preference to convey informative content. The gradual pattern observed in the data might thus reflect individuals' varying sensitivity to the manipulation: The VISIBLE SPEAKER condition most strongly cues participants to consider the speaker's production process, but evidently some participants seem to already be engaging their awareness of the speaker in the less speaker-salient conditions. The manipulation seems to affect the measures differently, suggesting that the mechanism for comprehenders to change perspectives in this way may be more complex than our experimental design can account for. Nevertheless, it seems that whatever the process may be that generates candidate expressions when anticipating content in a sentence completion task, this process is substantially modulated by comprehenders' awareness of the speaker.

3 Experiment 2: Awareness of speaker style in comprehension

Having established in Experiment 1 that comprehenders appear to use their awareness of speakers' production biases in generating expectations about what a speaker will say next, Experiment 2 tests whether such expectations also reflect properties of the speaker.¹³ We ask whether comprehenders have fixed expectations for speakers to convey newsworthy content, or whether such expectations are malleable. This study is different from the studies on referential expectations discussed in the introduction, where the communicative goal is identifying the intended referent. Experiment 2 instead aims to test speaker-specific expectations for informativity regarding upcoming message content where the communicative goal is unspecified and likely to be communicative interest. For example, comprehenders may estimate that different speakers have different thresholds for what counts as an informative contribution. Comprehenders may in turn allow for variation in the way a speaker delivers newsworthy content depending on what they know about the speaker.

¹³This experiment was preregistered: <https://osf.io/r2h7d>

Studies on speaker adaptation in a range of linguistic domains show that comprehenders use cues from a speaker's previous discourse when anticipating upcoming content and when interpreting it. For example, a speaker's accent or dialect can lead comprehenders to shift their perception of phonemes (Hay et al. 2006, Hay & Drager 2010); when encountering temporarily ambiguous words comprehenders make anticipatory looks to a more frequently mentioned referent based on the speaker's previous behaviour (Creel et al. 2008); in the early stages of processing, comprehenders are influenced by a speaker's either literal or non-literal style when disambiguating metaphorical polysemous words (Davies et al. 2022); and comprehenders adapt their interpretations of scalar quantifiers such as *some* and *many* following exposure to different speakers (Yildirim et al. 2016). There are also formalised accounts of, e.g. phonetic adaptation (Kleinschmidt & Jaeger 2015) and semantic/pragmatic adaptation (Schuster & Degen 2020) which capture comprehenders' ability to track subtle characteristics of different speakers and adjust their interpretations accordingly.

Using the VISIBLE SPEAKER condition from Experiment 1, our second experiment tests whether speaker-specific effects are observable in comprehenders' expectations for the informativity of upcoming content. Two different speakers with different communication styles were introduced in an exposure phase: One who produces a mix of informative and uninformative utterances and one who produces only informative utterances. Following the exposure phase, participants completed sentences purportedly uttered by these two speakers. If comprehension reflects speaker-specific expectations for informativity, comprehenders are predicted to provide more informative completions for the HIGH-INFORMATIVITY speaker compared to the LOW-INFORMATIVITY speaker, showing that they can modify their expectation about what someone is going to say next after learning about the specific production preferences of the individual speakers. Such a finding would indicate that in anticipating upcoming content, we are able to use awareness of production dynamically as we get to know the communication style of particular speakers.

3.1 Methods

3.1.1 Participants

200 participants were recruited on Prolific with the same criteria as in Experiment 1: English as first language, no known language impairments, location in the US or the UK, and minimum 18 years of age ($M = 36.07$, $SD = 14.20$, range = 18–76). Additionally, people who had taken part in Experiment 1 were blocked

from participating in Experiment 2. As in Experiment 1, participants who failed two or more attention checks or reported another language than English as their first language were excluded and replaced by other participants to reach the desired number of 100 participants per condition. Participants were paid £7.60/hour on average.¹⁴



Figure 6: Example of a filler item for the **LOW-INFORMATIVITY** speaker in exposure phase. All items in exposure consisted of two parts, shown here in two boxes. Image from Freepik (<https://freepik.com>).

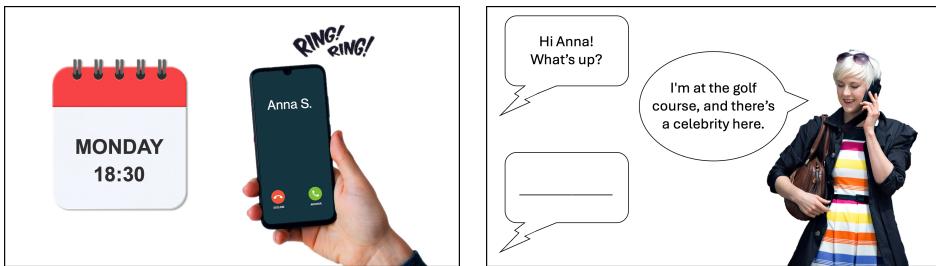


Figure 7: Example of a critical item for the **HIGH-INFORMATIVITY** speaker in exposure phase. All items in exposure consisted of two parts, shown here in two boxes. Image from Pixabay (<https://pixabay.com>).

3.1.2 Design

Experiment 2 tests whether comprehenders adapt their expectations for informativity based on the speaking style of two different speakers: One who produces a mix of informative and uninformative utterances and one who produces informative utterances only. The experiment has two conditions (**HIGH-INFORMATIVITY** vs. **LOW-INFORMATIVITY** speaker style), manipulated within participants. The task

¹⁴The details of the cost were established by Prolific based on the median time taken to complete the experiment per submission for each group.

consisted of two phases; an exposure phase and a test phase. In the exposure phase, participants were familiarised with two different speakers, Suzy and Anna, and their differing communication styles (see an example for each of the speakers in Figures 6 and 7). The test phase is a near replication of the VISIBLE SPEAKER condition in Experiment 1 in which participants write completions for the two different speakers (examples in Figures 8 and 9). We designed this paradigm to encourage participants to feel like they were engaged in a series of phone calls. In the exposure phase, the participant's task is to type in utterances that contribute their own turns in the dialogue. In the test phase, on the other hand, they are instructed to fill in parts of the caller's utterances. The exposure phase thus ensures that they witness the kind of utterances produced by the two callers, whereas the test phase asks them to indicate what they think each of the callers would be likely to say.



Figure 8: Example of a critical item for the LOW-INFORMATIVITY speaker in test phase

Completions for the HIGH-INFORMATIVITY speaker are expected to be more informative than those for the LOW-INFORMATIVITY speaker. Such an observed difference in the measures between LOW-INFORMATIVITY speaker condition and HIGH-INFORMATIVITY speaker condition would reflect a change in participants' evaluation of what counts as informative for each of the speakers, in line with our thesis that comprehenders are able to dynamically use their knowledge of speakers' communicative styles in their estimates of likely upcoming content.



Figure 9: Example of a critical item for the HIGH-INFORMATIVITY speaker in test phase

3.1.3 Materials

Materials were constructed to depict a series of beginnings of phone calls, with a picture of the speaker (the person calling), and speech bubbles showing different turns of the conversation. The critical items (training items in the exposure phase and test items in the test phase) are in the form of *I'm at the [location], and there's _____*, as in Experiment 1, using the same 20 locations split across the two phases.

In the exposure phase, participants see the start of a phone call and a contribution from either the HIGH-INFORMATIVITY speaker or the LOW-INFORMATIVITY speaker. For the critical items in the exposure phase, the HIGH-INFORMATIVITY speaker only utters training sentences about non-typical situations (e.g. *I'm at the golf course, and there's a celebrity here.*), whereas the LOW-INFORMATIVITY speaker utters some training sentences about non-typical situations as well as some about typical situations (e.g. *I'm at the cinema, and there's popcorn.*). For the critical training items in the exposure phase, the speakers' utterances are taken from responses in Experiment 1: The HIGH-INFORMATIVITY speaker has 2 critical training items, both using completions from the VISIBLE SPEAKER condition, whereas the LOW-INFORMATIVITY speaker has 8 critical training items, 2 with completions from the VISIBLE SPEAKER condition and 6 from the BARE condition. Note that this setup ensures that both speakers provide the same number of interesting contributions overall, to avoid participants inferring that one speaker encounters more non-typical situations than the other. The different communication styles of the

speakers is then emphasised by the fact that the **LOW-INFORMATIVITY** speaker additionally reports on 6 typical situations. In addition to the critical training items, the exposure phase includes 18 filler items; 6 for the **HIGH-INFORMATIVITY** speaker and 12 for the **LOW-INFORMATIVITY** speaker. This setup means that overall the participant answers calls from the **LOW-INFORMATIVITY** speaker noticeably more often than the **HIGH-INFORMATIVITY** speaker (20 vs. 8 items). Filler items were either utterances requesting information or conveying information. These were constructed to match the communication style of the speakers, such that the fillers for the **HIGH-INFORMATIVITY** speaker targeted newsworthy information (e.g. *Do you know when Lisa is arriving next week?*), whereas the **LOW-INFORMATIVITY** speaker called about mundane events (e.g. *What did you have for dinner?*). Across the critical and filler items in the exposure phase, participants were encouraged to engage in the task and to see the setup as a communicative context, by requiring them to provide either a starting utterance for the conversations or a response to the speaker's utterance.

All the items in the exposure phase consisted of two images (see Figures 6 and 7). The first shows a calendar with a day and a time next to a hand holding a ringing phone with the caller's name clearly visible. The day and time was included to reinforce the manipulation that the **LOW-INFORMATIVITY** speaker calls more often; they call every day and occasionally several times a day, whereas the **HIGH-INFORMATIVITY** speaker calls a maximum of once a day and not every day. The second image shows the speaker and a series of speech bubbles indicating the first, second and third turn of the dialogue. The speaker's speech bubble, which is always the second turn, is always fully visible in order to expose the participant to each speaker's communicative style, whereas the speech bubbles for the callee (the participant) contain fill-in-the-blank spaces in either the first or the third turn.

For the test phase, the speakers' utterances include a fill-in-the-blank space to elicit participants' guesses about what each of the speakers is likely to say. All items in the test phase were similar to those in the **VISIBLE SPEAKER** condition of Experiment 1, showing the speaker and a speech bubble with an utterance that includes a blank; here in Experiment 2, the test items additionally depict a ringing phone and the callee's greeting in a speech bubble appearing to one side (see examples Figures 8 and 9). A participant sees an equal number of test items from each speaker (5 from the **LOW-INFORMATIVITY** speaker and 5 from the **HIGH-INFORMATIVITY** speaker), plus 10 filler items from each speaker as well. Filler items were the same as in Experiment 1, including the two that served as attention checks.

Critical training and test items (i.e. the 20 locations from Experiment 1) were counterbalanced across lists such that a location that appeared in the exposure phase in one list appeared as a test item in another list. Similarly, participants saw the HIGH-INFORMATIVITY (HI) and LOW-INFORMATIVITY (LI) speaker depicted as one of the two speaker images, counterbalanced across participants, to avoid any potential bias associated with the appearance of the speakers. Counterbalancing items and pictures of speakers resulted in 4 lists:

- LI speaker: dark-haired, HI speaker: blonde + exposure items: target set A
- LI speaker: blonde, HI speaker: dark-haired + exposure items: target set A
- LI speaker: dark-haired, HI speaker: blonde + exposure items: target set B
- LI speaker: blonde, HI speaker: dark-haired + exposure items: target set B

Under this counterbalancing, a given item in the exposure phase was always associated with the same speaker style (e.g. *train station* was always uttered by the LOW-INFORMATIVITY speaker in exposure), and likewise a given item in the test phase was always associated with the same speaker style (e.g. *restaurant* was always uttered by the HIGH-INFORMATIVITY speaker in test).

3.1.4 Procedure

The experiment was presented using Qualtrics. In the exposure phase, participants were instructed to imagine answering calls from each of two speakers, filling in an empty speech bubble from the participant's side of the conversation in each call. Each trial consisted of two parts (example trials in Figures 6 and 7). First, participants see a day and time and a phone showing the caller's name; second, they see the caller talking into a phone and the dialogue in speech bubbles, and a text box for filling in the missing content. The items varied between the participant having to complete the very first utterance (the greeting) or the response to the caller's first utterance. This variation was intended to ensure participants stayed engaged and read the utterances fully, as well as to distract them from the potentially unnatural sounding utterances from the caller by allowing them to influence the dialogue by contributing their own turns. Items in the exposure phase were presented in a fixed order, so that every participant within a list saw the same item order. As in Experiment 1, participants were presented with one item at a time with a text field below the item where they typed their completions, and they proceeded to the next trial by clicking the arrow on screen.

In the test phase, participants were again instructed that they would be seeing the beginnings of a series of phone calls from the same two speakers, but this time they would be completing the callers' utterances. The test phase was essentially a replication of the VISIBLE SPEAKER condition in Experiment 1, with the only difference being the edits explained above. The test phase used simpler visualisations of the phone calls compared to the exposure phase, as shown in Figures 6 and 7 versus 8 and 9; this simplification was intended to make the second phase of the experiment less tiresome for the participant by reducing the number of images they had to click through. Presentation of items in the exposure phase was fully randomised.

After the test phase, there were three questions intended to assess whether participants had paid attention and were sensitive to the different speaker styles. The first of these was a speaker line-up where participants had to choose which of five speakers they had talked to. The speaker line-up consisted of the images of the two speakers flipped horizontally, and three images of other speakers taken from Experiment 1. The second question asked if the participant noticed a difference between the two speakers, and if so, what that difference was. The third asked them to rate how interesting they found each of the two speakers on a scale from 1 (not at all interesting) to 7 (very interesting). The median completion times in each list were between 17 and 19 minutes.

3.1.5 Measures and data analysis

For Experiment 2, we used the same coding procedure and the same measures as in Experiment 1: Entropy as a measure of variability in participant responses, KL divergence to measure how the distribution of responses in the two conditions here compares to that of a baseline condition (the BARE condition from Experiment 1), modification and negation as measures of participants' enhancement of their responses, and typicality of the responses' main noun as compared to the nouns elicited in the Experiment 1 pre-test. A total of 2 continuations were non-meaningful and therefore excluded from analysis. The included continuations ranged from 1 word in length to 6 words, with a mean of 1.55.

3.2 Results

The mean number of distinct responses in the HIGH-INFORMATIVITY condition was 58.0, compared to 44.7 in the LOW-INFORMATIVITY condition. As predicted for the entropy measure, the HIGH-INFORMATIVITY speaker yielded a distribution of responses with higher entropy (mean: 3.60) compared to the LOW-INFORMATIVITY

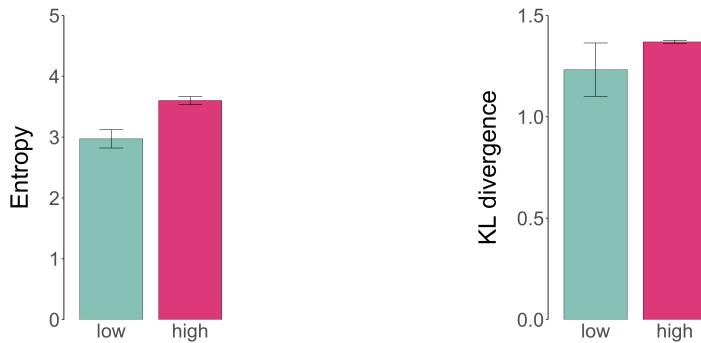


Figure 10: The left panel shows mean entropy for each speaker informativity condition; the right panel shows KL divergence as a comparison between the distribution in the listed condition and that of the baseline BARE condition.

speaker (mean: 3.00; Wilcoxon signed-rank test: $p < 0.005$), as seen in the left panel of Figure 10. For KL divergence, there was a numeric difference between HIGH-INFORMATIVITY vs. BARE (1.37) and LOW-INFORMATIVITY vs. BARE (1.23), but the difference was not significant (right panel, Figure 10). As in Experiment 1, we also calculated the entropy of the subset of responses not containing any modification or negation. For entropy, the pattern still holds, with HIGH-INFORMATIVITY (mean: 2.80) being significantly higher than LOW-INFORMATIVITY (mean: 2.20, $p < 0.05$). For KL divergence, the pattern above remained numeric only (HIGH-INFORMATIVITY vs. BARE (1.14) and LOW-INFORMATIVITY vs. BARE (1.10)).

As in Experiment 1, the binary variables of modification and negation were analysed with generalised mixed effects models treating condition (here, HIGH-INFORMATIVITY vs. LOW-INFORMATIVITY) as a fixed effect and with random slopes and intercepts of condition for participants and items. For the model for modification, we again used the optimiser `optimx` with the method `bobyqa` to achieve convergence with the full random effects structure. As predicted and as can be seen in Figure 11, modification was used more frequently in completions of utterances from the HIGH-INFORMATIVITY speaker compared to the LOW-INFORMATIVITY speaker ($\beta = -0.88$, SE = 0.19, $z = -4.57$, $p < 0.001$). The model for negation converged with full random effects structure and the default optimiser; however, although numerically the predicted pattern is observable (see Figure 12), the difference between HIGH-INFORMATIVITY/LOW-INFORMATIVITY speakers was not significant ($\beta = -0.58$, SE = 0.57, $z = -1.02$, $p = 0.306$).

Typicality was analysed with the same approach as in Experiment 1, and again the predicted pattern was confirmed (Figure 13). We constructed a linear mixed

effects model with condition (HIGH-INFORMATIVITY VS. LOW-INFORMATIVITY) as a fixed effect and random slopes and intercepts of condition for participants and items, with the optimiser `optimx` combined with the method `bobyqa` to achieve convergence. The model confirms that typical main nouns were used less often for the HIGH-INFORMATIVITY speaker compared to LOW-INFORMATIVITY speaker ($\beta = -0.13$, SE = 0.06, $t = -2.32$). To derive a p -value, we used model comparison via a likelihood ratio test and found a main effect ($p < 0.05$), showing that the model with condition as a fixed effect provides a significantly better fit than the model with same random effects structure but no fixed effect.

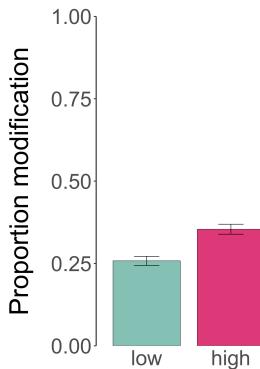


Figure 11: Proportion of use of modification across conditions

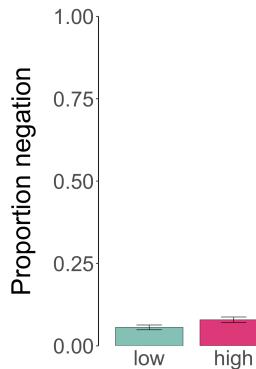


Figure 12: Proportion of use of negation across conditions

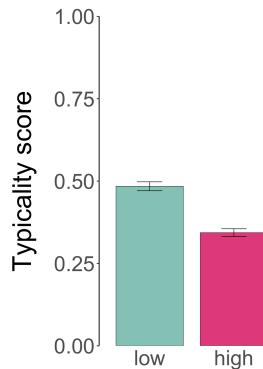


Figure 13: Mean typicality score of responses across conditions

3.3 Discussion

In line with the claim that comprehension reflects speaker-specific expectations for informativity, the results of Experiment 2 show that participants' completions were higher in informativity for the HIGH-INFORMATIVITY speaker compared to the LOW-INFORMATIVITY speaker. The HIGH-INFORMATIVITY speaker completions were more variable as shown by the higher entropy, and elicited more modification and lower typicality, showing that participants adapt their expectations to the individual speakers' communication styles as predicted. For KL divergence, there was a numeric difference in the predicted direction, but this was not significant. The results for the rate of negation were numerically in line with the prediction that the HIGH-INFORMATIVITY speaker is more likely to comment on the *lack* of something typical compared to the LOW-INFORMATIVITY

speaker, but this effect was also not significant. This may reflect the fact that none of the critical items in the exposure phase explicitly demonstrate how the two speakers use negation, whereas some of both filler and critical items in the exposure phase did use modification.

4 General discussion

The goal of this study was to investigate the effect of comprehenders' awareness of a speaker's production preferences and communication style on their anticipation of upcoming content. Previous work on comprehension and production tends to show contradicting preferences for comprehenders and speakers, such that typicality is favoured in comprehension (e.g. being associated with processing ease), whereas speakers tend to make production choices that favour the inclusion of surprising or otherwise informative content (meaning easily inferable or typical content is disfavoured). However, we posited that comprehenders are aware of and make use of speakers' production preferences in their expectations for upcoming content, such that they have a bias towards filtered language use rather than a simple expectation of transparent language use. The two Cloze task studies presented here tested this by tapping into the comprehender's expectations, when the comprehender does not have access to the scene being described.

Experiment 1 addressed the question of whether increasing the emphasis on the speaker would lead to an increased expectation for informative content. The higher entropy of completions in the VISIBLE SPEAKER condition and also FIRST PERSON and THIRD PERSON compared to BARE, show that participants' completions were more variable and thus less predictable in the conditions where the speaker was more salient. Furthermore, KL divergence showed that the distribution of participants' completions in VISIBLE SPEAKER condition was the most clearly different from the distribution in BARE condition. We take this as evidence that participants are tuning in to speakers' production preferences and expecting speakers to tell them interesting things by providing more unique completions in these conditions compared to the BARE condition. For example, while *train* was a common completion for the train station location in the BARE condition, it was much less frequent in the VISIBLE SPEAKER condition, suggesting that participants are taking into account what a speaker might consider worth talking about when they provide their completions.

Our predictions were upheld for the other three measures. Rather than looking at the overall variability in participants' completions, these measures aim to capture properties of the content of individual completions. Modification and

negation are taken to be strategies that participants can employ to make something that might otherwise be considered typical for a location more non-typical and therefore worthy of reporting. Participants made use of modification such as descriptive adjectives (e.g. *a cute dog*), quantifying expressions (e.g. to communicate unusual amounts of an entity: *so much choice*) or modification to specify a less typical subtype of an entity (e.g. *steam train*). Similarly, negation can be a way to mark the absence of something that one would typically expect to find at a location, such as a train at the train station, and reporting this absence (e.g. *no trains*) would therefore likely be considered informative. As for typicality, the results showed that participants produced completions mentioning typical entities most often in the **BARE** condition, and significantly less often in the **VISIBLE SPEAKER** condition. These measures and the progressive differences across conditions are taken to reflect comprehenders' sensitivity to the experimental manipulation of speaker salience and, more broadly, to reveal their expectation for speakers to convey non-typical and informative content.

These findings have potential repercussions for the use of Cloze tasks in other psycholinguistic research since the kinds of completions that participants provide are evidently malleable. Given the differences in response distributions that we observed across conditions, researchers using Cloze tasks may need to tune their experimental tasks to their precise research goals. This might require using standard Cloze task phrasing and procedures to elicit completions that reflect participants' estimates of what is typical in the world or using more situated tasks that depict a communicative context in order to elicit completions that reflect participants' estimates of what a speaker is likely to talk about.

Experiment 2 addressed the question of whether this expectation is further malleable depending on properties of the speaker, specifically whether comprehenders are sensitive to different communication styles. Using completions provided in Experiment 1, participants were exposed to two speakers whose utterances were either high or low in informativity. Similar to the speaker salience manipulation in Experiment 1, the manipulation of speaker properties in Experiment 2 allows us to test whether comprehenders take into account properties of the speaker in guessing upcoming content (transparent language use vs. filtered language use). Participants showed a bias to expect the **HIGH-INFORMATIVITY** speaker to produce utterances conveying more informative content than the **LOW-INFORMATIVITY** speaker. Entropy was higher for the **HIGH-INFORMATIVITY** speaker, showing more variability in completions. Although the KL divergence measure did not show a significant difference between the two conditions (when the distribution of responses in each speaker condition is compared to that of the **BARE** condition in Experiment 1), there was a numeric difference in the predicted

direction. Modification was higher for the HIGH-INFORMATIVITY speaker, and although not significant, we also saw an increase in use of negation with the HIGH-INFORMATIVITY speaker. Lastly, mention of typical entities was more frequent for the LOW-INFORMATIVITY speaker. We take this to show that comprehenders are able to estimate that different speakers have different thresholds for what counts as an informative contribution and adjust their expectations accordingly.

A challenge when discussing how we anticipate upcoming content is to clearly distinguish the relevant concepts involved. Typicality, plausibility and (im)possibility all seem interwoven, and although they all likely play a role in anticipating what someone is going to say next, we have focussed on typicality in this study. We have defined typicality as the frequency of an event or situation; the frequency with which something occurs is conceptually and empirically relatively straightforward to measure. As we observed in this study, participants expect utterances to convey less typical content when the speaker is emphasised. Importantly, however, no participant seemingly contributed responses that were so extremely non-typical that they crossed the threshold into being impossible (e.g. a steam train in the kitchen). One could imagine scenarios where utterances about highly non-typical situations would be felicitous (e.g. in descriptions of fictional or dream worlds; Foy & Gerrig 2014, Troyer & Kutas 2020), and perhaps even the most expected in some contexts, but there seems to be an intuition that utterances should stay within the realm of what is plausible.

One way of conceiving of the interplay between typicality and plausibility could be that the latter provides a range within which situations can happen, whereas whether or not those situations are typical will depend on their frequency. For instance, taking a sip from a transmitter is implausible; drinking from a waterfall, however, certainly is plausible, but significantly less typical than drinking from a tap. Since continuations in our current study all seemed to fall within the range of plausible utterances, our data cannot speak to the role of plausibility in processing and anticipating upcoming content. There is work attempting to tease apart the effects of impossibility and implausibility on processing (see e.g. Warren & McConnell 2007) and other work that makes a distinction between real-world plausibility and word-predictability (e.g. Albu et al. 2023). As future work looks at more fine-grained processing in relation to anticipation of upcoming content and speaker salience, it could become important to make clear distinctions between all these concepts; typicality, plausibility and possibility.

As discussed above, previous work has emphasised the importance of typicality in processing (e.g. Marks & Miller 1964, Kutas & Hillyard 1980, Morris 1994, Kamide et al. 2003). When looking at the psycholinguistic literature, one might

think that the default approach is to see sentences describing implausible or non-typical situations as anomalies that need to be reparsed (see e.g. Cai et al. 2022). However, such work sidesteps the role of pragmatic reasoning, failing to incorporate a comprehender's awareness of the possibility that such an utterance has been produced precisely because its content is surprising, in particular in contexts where the communicative goal of the interaction is unspecified. Of course, it is reasonable to assume that most language users will expect a contribution to stay within a certain range of plausibility – if a contribution is too implausible, one would expect processing to suffer because it may be hard for comprehenders to reconcile what the speaker has said with what they know about the world. However, the results presented here suggest that comprehenders do have expectations that speakers will talk about interesting and non-typical things, highlighting a role for informativity-driven reasoning about the speaker in models of language processing.

Acknowledgements

We thank editor Robin Lemke and two anonymous reviewers for very helpful comments.

References

- Agresti, Alan. 2003. *Categorical data analysis*. 2nd edn., vol. 482 (Wiley Series in Probability and Statistics). Somerset: Wiley.
- Albu, Elena, Carolin Dudschg, Tessa Warren & Barbara Kaup. 2023. Does negation influence the choice of sentence continuations? Evidence from a four-choice cloze task. *Quarterly Journal of Experimental Psychology*. 17470218231158109. DOI: 10.1177/17470218231158109.
- Altarriba, Jeanette, Judith F. Kroll, Alexandra Sholl & Keith Rayner. 1996. The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition* 24(4). 477–492. DOI: 10.3758/BF03200936.
- Aylett, Matthew & Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1). 31–56. DOI: 10.1177/00238309040470010201.

- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68(3). 255–278. DOI: 10.1016/j.jml.2012.11.001.
- Bates, Douglas, Martin Mächler, Ben M. Bolker & Steve C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of statistical software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.
- Borghi, Anna M. & Felice Cimatti. 2010. Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia* 48(3). 763–773. DOI: 10.1016/j.neuropsychologia.2009.10.029.
- Brown, Paula M. & Gary S. Dell. 1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19. 441–472. DOI: 10.1016/0010-0285(87)90015-6.
- Cai, Zhenguang G., Nan Zhao & Martin J. Pickering. 2022. How do people interpret implausible sentences? *Cognition* 225. 105101. DOI: 10.1016/j.cognition.2022.105101.
- Creel, Sarah C., Richard N. Aslin & Michael K. Tanenhaus. 2008. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106. 633–664. DOI: 10.1016/j.cognition.2007.03.013.
- Davies, Catherine & Jennifer E. Arnold. 2019. Reference and informativeness: How context shapes referential choice. In Chris Cummins & Napoleon Katsos (eds.), *The Oxford handbook of experimental semantics and pragmatics*, 474–493. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198791768.013.6.
- Davies, Catherine, Vincent Porretta, Kremena Koleva & Ekaterini Klepousniotou. 2022. Speaker-specific cues influence semantic disambiguation. *Journal of Psycholinguistic Research* 51(5). 933–955. DOI: 10.1007/s10936-022-09852-0.
- Degen, Judith, Robert D. Hawkins, Caroline Graf, Elisa Kreiss & Noah D. Goodman. 2020. When redundancy is useful: A Bayesian approach to “overinformative” referring expressions. *Psychological Review* 127(4). 591–621. DOI: 10.1037/rev0000186.
- Foy, Jeffrey E. & Richard J. Gerrig. 2014. Flying to Neverland: How readers tacitly judge norms during comprehension. *Memory & Cognition* 42(8). 1250–1259. DOI: 10.3758/s13421-014-0436-8.
- Goodman, Noah D. & Michael C. Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences* 20. 818–829. DOI: 10.1016/j.tics.2016.08.005.
- Grice, Herbert Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics: Speech acts*, 41–58. New York: Academic Press.

- Grodner, Daniel & Julie C. Sedivy. 2011. The effect of speaker-specific information on pragmatic inferences. In Edward A. Gibson & Neal J. Pearlmuter (eds.), *The processing and acquisition of reference*, 239–272. Cambridge: The MIT Press. DOI: 10.7551/mitpress/9780262015127.003.0010.
- Hay, Jennifer & Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4). 865–892. DOI: 10.1515/ling.2010.027.
- Hay, Jennifer, Aaron Nolan & Katie Drager. 2006. From fush to feesh: Exemplar priming in speech perception. *Linguistic review* 23(3). 351–379. DOI: 10.1515/TLR.2006.014.
- Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446. DOI: 10.1016/j.jml.2007.11.007.
- Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61(1). 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.
- Kamide, Yuki, Gerry T. M. Altmann & Sarah L. Haywood. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49(1). 133–156. DOI: 10.1016/S0749-596X(03)00023-8.
- Kleinschmidt, Dave F. & T. Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 122(2). 148–203. DOI: 10.1037/a0038695.
- Kravtchenko, Ekaterina. 2014. Predictability and syntactic production: Evidence from subject omission in Russian. *Proceedings of the annual meeting of the Cognitive Science Society* 36. 785–790.
- Kravtchenko, Ekaterina & Vera Demberg. 2022. Informationally redundant utterances elicit pragmatic inferences. *Cognition* 225. 105159. DOI: 10.1016/j.cognition.2022.105159.
- Kreiss, Elisa & Judith Degen. 2020. Production expectations modulate contrastive inference. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. 259–265.
- Kuperberg, Gina R. & T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31. 32–59. DOI: 10.1080/23273798.2015.1102299.
- Kutas, Marta & Steven A. Hillyard. 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207. 203–205. DOI: 10.1126/science.7350657.

- Lemke, Robin. 2021. *Experimental investigations on the syntax and usage of fragments* (Open Germanic Linguistics 1). Language Science Press. DOI: 10.5281/zenodo.5596236.
- Lemke, Robin, Ingo Reich, Lisa Schäfer & Heiner Drenhaus. 2021a. Predictable words are more likely to be omitted in fragments: Evidence from production data. *Frontiers in Psychology* 12. 662125–662125. DOI: 10.3389/fpsyg.2021.662125.
- Lemke, Robin, Lisa Schäfer & Ingo Reich. 2021b. Modeling the predictive potential of extralinguistic context with script knowledge: The case of fragments. *PLoS ONE* 16. DOI: 10.1371/journal.pone.0246255.
- Levy, Roger & T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. In Bernhard Schölkopf, John Platt & Thomas Hofmann (eds.), *Advances in neural information processing systems*, 849–856. Cambridge: The MIT Press. DOI: 10.7551/mitpress/7503.003.0111.
- Lockridge, Calion B. & Susan E. Brennan. 2002. Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review* 9. 550–557. DOI: 10.3758/BF03196312.
- Marks, Lawrence E. & George A. Miller. 1964. The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior* 3. 1–5. DOI: 10.1016/S0022-5371(64)80052-9.
- Mitchell, Margaret, Ehud Reiter & Kees Van Deemter. 2013. Typicality and object reference. *Proceedings of the Annual Meeting of the Cognitive Science Society* 35. 3062–3067.
- Morris, Robin K. 1994. Lexical and message-level sentence context effects on fixation times in reading. *Journal of Experimental Psychology: Learning, Memory & Cognition* 20. 92–103. DOI: 10.1037/0278-7393.20.1.92.
- Nieuwland, Mante S. & Jos J. A. Van Berkum. 2006. When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18(7). 1098–1111. DOI: 10.1162/jocn.2006.18.7.1098.
- Pogue, Amanda, Chigusa Kurumada & Michael K. Tanenhaus. 2016. Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology* 6. 2035–2035. DOI: 10.3389/fpsyg.2015.02035.
- R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. <https://www.R-project.org/>.
- Rohde, Hannah, Richard Futrell & Christopher G. Lucas. 2021. What's new? A comprehension bias in favor of informativity. *Cognition* 209. 104491. DOI: 10.1016/j.cognition.2020.104491.

- Rohde, Hannah, Jet Hoek, Maayan Keshev & Michael Franke. 2022. This better be interesting: A speaker's decision to speak cues listeners to expect informative content. *Open Mind* 6. 118–131. DOI: 10.1162/opmi_a_00058.
- Rubio-Fernández, Paula. 2016. How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology* 7. DOI: 10.3389/fpsyg.2016.00153.
- Ryskin, Rachel, Chigusa Kurumada & Sarah Brown-Schmidt. 2019. Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science* 43(8). e12769. DOI: 10.1111/cogs.12769.
- Schäfer, Lisa, Robin Lemke, Heiner Drenhaus & Ingo Reich. 2021. The role of UID for the usage of verb phrase ellipsis: Psycholinguistic evidence from length and context effects. eng. *Frontiers in Psychology* 12. 661087–661087. DOI: 10.3389/fpsyg.2021.661087.
- Schuster, Sebastian & Judith Degen. 2020. I know what you're probably going to say: Listener adaptation to variable use of uncertainty expressions. *Cognition* 203. 104285. DOI: 10.1016/j.cognition.2020.104285.
- Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32. 3–23. DOI: 10.1023/A:1021928914454.
- Stevenson, Rosemary J., Rosalind A. Crawley & David Kleinman. 1994. Thematic roles, focus and the representation of events. *Language and Cognitive Processes* 9(4). 519–548. DOI: 10.1080/01690969408402130.
- Taylor, Wilson L. 1953. "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly* 30(4). 415–433. DOI: 10.1177/107769905303000401.
- Troyer, Melissa & Marta Kutas. 2020. Harry Potter and the Chamber of What?: The impact of what individuals know on word processing during reading. *Language, Cognition & Neuroscience* 35(5). 641–657. DOI: 10.1080/23273798.2018.1503309.
- Venuizen, Noortje J., Matthew W. Crocker & Harm Brouwer. 2019. Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes* 56(3). 229–255. DOI: 10.1080/0163853X.2018.1448677.
- Warren, Tessa & Kerry McConnell. 2007. Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review* 14(4). 770–775. DOI: 10.3758/BF03196835.
- Westerbeek, Hans, Ruud Koolen & Alfons Maes. 2015. Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology* 6. DOI: 10.3389/fpsyg.2015.00935.

Vilde R. S. Reksnes, Alice Rees, Chris Cummins & Hannah Rohde

Yildirim, Ilker, Judith Degen, Michael K. Tanenhaus & T. Florian Jaeger. 2016.
Talker-specificity and adaptation in quantifier interpretation. *Journal of Memory and Language* 87. 128–143. DOI: 10.1016/j.jml.2015.08.003.