Rethinking Probabilities: Why Corpus Frequencies Cannot Capture Speakers' Dynamic Linguistic Behavior

Santina Kemper (santina.simone.kemper@hu-berlin.de)

Humboldt Universität zu Berlin

Holly Jenkins (holly.jenkins@education.ox.ac.uk)

Department of Education, University of Oxford

Elizabeth Wonnacott (elizabeth.wonnacott@education.ox.ac.uk)

Department of Education, University of Oxford

Michael Ramscar (michael.ramscar@uni-tuebingen.de)

Department of Psychology, University of Tübingen

Abstract

Because information theory equates information with event occurrence probabilities, when applying its methods, language researchers typically take the information provided by words to be their relative frequencies in a corpus. This implicitly assumes words occur uniformly across contexts, however empirically, word distributions are bursty: the likelihood of most words appearing in most contexts is small, whereas the likelihood of a word recurring in context is much higher. In an elicitation study we examined whether speakers are sensitive to the dynamic word occurrence probabilities this implies. Consistent with proposals that prenominal adjectives increase noun predictability, participants produced numerous seemingly redundant adjectives prior to unambiguous nouns at first mention. However, despite receiving no feedback, they produced significantly fewer adjectives before subsequent mentions of the same nouns, indicating they had re-evaluated their probabilities. These results support the idea that prenominal adjectives facilitate efficient communication, and that speakers' representations of lexical probabilities are dynamic.

Keywords: information theory; burstiness; dynamic language modeling; language processing; speech production

Information theory and natural language

What is language, and how does it enable us to communicate our thoughts, beliefs, desires etc.? Traditional answers to this question have tended to assume that messages are built from acoustic/semantic 'atoms' whose contributions are then modified by syntactic rules and transferred from speakers to listeners. While these assumptions accord well with human intuition, the problems involved in operationalizing and formalizing them have led to claims that language is unlearnable and must be innate, which in turn has been undermined by developments in machine learning that have rendered claims about learnability somewhat implausible (Piantadosi, 2023). This pattern of scientific progression has led researchers to examine other ways of characterizing human communication, for example, framing human communication in terms of information theory (Gibson et al, 2019), and taking this to its logical conclusions, assuming that it involves structured semantic uncertainty reduction rather than compositionality and meaning transfer (Ramscar, 2021).

From this perspective, the simplest form of a communication system can be thought of as containing a source (of messages), a channel (the medium over which the messages are transmitted) and a receiver. Information theory (Shannon, 1948) is concerned with two basic issues concerning communication in such systems: how to define and quantify information, and how to best optimize the information communicated across a given channel.[1]

In answering the first of these questions, the theory takes the probabilities of the occurrence of events as a measure of their information (entropy): the information content of any part of a message is its probability of occurrence, such that the higher its probability, the less informative it is, whereas the smaller its probability, the more informative it is (Hartley, 1928; Shannon, 1948).

All communication channels have a capacity that limits the amount of information they can communicate at any point in time. In answer to the second question, Shannon (1948) showed how distributing information across sequences and organizing encoding in systems could allow codes that support communication close to the capacity of a channel to be defined. This idea is best understood by

[1] It is also concerned with dealing with channel noise, however this issue is somewhat orthogonal to the questions we address here.

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

considering the occurrence rates of English color adjectives shown in Table 1.

Table 1: The relative frequencies of 24 color adjectives in the 1.1 billion word corpus of American English (Davies, 2009), their probabilities, the information communicated by each adjective, and the information entropy of this distribution (the amount of information communicated on average).

	COCA Frequency	COCA Probability	Information Content (in bits)	Entropy (in bits)
WHITE	256270	0.245	2.027	0.497
BLACK	216258	0.207	2.272	0.470
RED	112362	0.108	3.216	0.346
GREEN	87871	0.084	3.571	0.300
BROWN	80732	0.077	3.693	0.286
BLUE	74391	0.071	3.811	0.272
GREY/GRAY	39832	0.038	4.712	0.180
YELLOW	32288	0.031	5.015	0.155
SILVER	27668	0.026	5.238	0.139
ORANGE	27657	0.026	5.239	0.139
PINK	21385	0.020	5.610	0.115
NAVY	19172	0.018	5.767	0.106
OLIVE	15415	0.015	6.082	0.090
PURPLE	11237	0.011	6.538	0.070
LIME	6045	0.006	7.432	0.043
TAN	5267	0.005	7.631	0.038
VIOLET	4251	0.004	7.940	0.032
TURQUOISE	1847	0.0018	9.143	0.016
MAROON	1174	0.0011	9.797	0.011
INDIGO	1092	0.0010	9.901	0.010
TEAL	909	0.0009	10.166	0.009
MAGENTA	635	0.0006	10.683	0.006
AQUAMARINE	267	0.0003	11.933	0.003
CYAN	195	0.0002	12.387	0.002
				3.34

According to the way the theory defines 'information', *White* can be considered to be the least informative English color adjective (it occurs most often, and so is most predictable), and *Teal* and *Magenta* the most informative English color adjectives. If we were to make a binary encoding of colors using 1 binary symbol for *White*, and 4 for *Teal* and *Magenta* it follows that the average number of symbols needed to communicate each color adjective will be much lower than if these adjectives were encoded using the same number of symbols (because 47% of all color adjectives communicated will be encoded using 1 symbol, whereas less than 6% of communicated color adjectives will be encoded using 4 symbols). Further, these coding probabilities will ensure that the average amount of information communicated across a channel at any point in time will be much closer to its capacity limit than if all of the color adjectives were equally likely to be communicated.

Information theory and natural language

Many of the properties that have been proposed for the formulation of efficient information theoretic codes have been shown to have their counterparts in natural language: for example, just as with the letters above, not only have the probabilities of words in each language have been shown to vary enormously, they do so in seemingly systematic ways (Estoup, 1916; Zipf, 1936). Further, and consistent with our discussion of efficient coding above, the forms of words that occur more often, and are more predictable, reliably tend to be shorter than more informative words (Piantadosi et al., 2021). Less informative words are also more likely to be phonetically reduced in production (Jurafsky et al., 2001; Bell et al., 2009; Seyfarth, 2014), while in a similar vein, optional function words (such as 'that') are more likely to be omitted when the phrase they precede is more predictable (Jaeger, 2010).

Given the many similarities between information theoretic codes and natural languages, it is perhaps unsurprising that researchers have proposed that information theory could provide a good model for understanding human communication (see e.g., Gibson et al, 2019), and that languages may actually be efficient in much the same way that information theoretic codes are, in that the distribution of lexical information across messages serves to keep the rate at which information is communicated in languages at a consistent average (a process known as 'smoothing'; Aylett & Turk, 2004; Jaeger & Levy, 2006).

Additionally, the explicitly predictive way in which information theory is framed has suggested functional roles for many aspects of language (e.g., grammatical gender) that have tended to defy traditional, logical approaches to language (Dye et al, 2017). This is because one particular matter that this perspective brings into focus is the potential problem posed by nouns (and proper names) to the ideas that communicative processes are efficient.

Measuring information in natural language

In common with most studies in the field (e.g., Piantadosi et al, 2021; Jurafsky et al., 2001; Bell et al., 2009; Seyfarth, 2014), Dye et al (2018)'s methods estimates of the uncertainty associated with adjectives and nouns were based on corpus frequencies. As in these other studies, Dye et al then used the probabilities they had estimated in order to determine the information provided by individual gendered articles and adjectives as well as the average information (entropy) of the distributions of the sets of articles and adjectives that might appear in a given context.

However, because this approach estimates the likelihood of occurrence of each noun and adjective as a function of the relative frequency with which it occurs across the range of contexts aggregated in a context, this methods of calculation the information conveyed by words implicitly assumes that words must occur across contexts in a relatively uniform manner. However, this assumption is incompatible with results that show that the actual occurrences of words in context are bursty (Katz, 1996). Most words – especially words with low average frequencies – typically occur in bursts: that is, they are likely to occur multiple times in a few texts, and never in most others. A consequence of this is that when most words first appear in context, their chances of reappearing in that context will increase (and typically this increase will be orders of magnitude greater than might be predicted from the base rate – i.e., average frequency – of the word, Katz, 1996). For example, it is likely that many readers of this paper will not have encountered the word 'bursty' prior to their doing so here (making its prior base rate effectively zero); yet if they keep on reading, they will encounter 'bursty' on several more occasions before they are through.

If lexical occurrences are indeed bursty (and the empirical evidence is that they are, Katz, 1996; Altmann et al, 2009; Slone et al, 2023), it follows that estimates of word probabilities made from averages of their occurrences of words across all contexts in a corpus must often be incorrect in many specific contexts. It also follows that if language users are sensitive to this property of language, such that they implicitly revise their probability of occurrence estimates for words after they have been encountered, it further follows that estimates of word probabilities made from corpus averages will often be poor predictors of linguistic behavior as well.

Intriguingly, there is evidence that language users actually may revise their expectations dynamically in the manner suggested above. Studies have shown novel reference phrases that are frequently reused become noticeably shorter after their first mention, while references of infrequently mentioned items are less likely to do so (Krauss & Weinheimer, 1964), while studies of production have shown that as referents are repeated in a context, the likelihood that their articulation will be reduced increases (Aylett, 2000). Accordingly, we set out to systematically examine whether, by eliciting spontaneous speech, we could establish whether language users are sensitive to the uncertainty associated with noun phrases, as findings about the way speakers use adjectives to make nouns more predictable would indicate (Dye et al, 2018), and at the same time, we sought to examine whether language users would modify their probabilistic behavior in context – producing less informative noun phrases as later mentions than first mentions in context – as the bursty nature of word occurrences and these empirical results would suggest.

However, while these effects are consistent with the idea that speakers' might implicitly revise their probability of occurrence estimates after words have been encountered, they are also amenable to other explanations. Krauss, & Weinheimer (1964)'s participants were collaborating in a reference task, and received feedback, such that the pattern of change in the reference phrases may reflect increased confidence in the likelihood of successful reference – convention building -- rather than increased expectancy for specific words themselves (Krauss, & Weinheimer, 1966), while the changes observed in articulation when the same word is repeated (Aylett, 2000) could also simply reflect local practice effects (i.e., learning in the articulatory system, Tomaschek et al, 2021).

Accordingly, in what follows we describe a study that was designed to examine whether Dye et al (2018)'s proposal that English speakers use apparently redundant prenominal adjectives to make the nouns in referential phrases more predictable (see also Rubio-Fernandez & Jara-Ettinger, 2020) could be observed at first mention in elicited spontaneous speech. It was then designed to elicit further mentions of the same referential phrases while controlling for the potential confounding effects of feedback and increased referential success. If our participants were using these apparently redundant prenominal adjectives to make the nouns in their referential phrases more predictable in the absence of any other information (and at between-mention durations longer than those associated with lexical priming, see e.g., Leinenger & Rayner, 2013), then it follows that if they are sensitive to burstiness their use of prenominal adjectives (or any other referential modifiers) should decrease, since the change in the recurrence probabilities of the already mention nouns would have changed the status of these adjectives from apparently redundant to actually redundant.

To achieve this, we designed a speech elicitation paradigm in which participants were presented displays containing 4 different, easily discriminable and nameable objects drawn from the set of objects in Table 2. The objects then moved around the display (this is described in more detail below), and after each movement, participants were asked to provide a set of verbal instructions that would allow someone presented with the original start state to recreate all of the subsequent movements. Because all of the objects were distinct and easily namable (Snodgrass and Vanderwart, 1980) it follows that any adjectives used to refer to them were, in theory, redundant: each object could be unambiguously referred to using a reference phrase that contained an article and the relevant noun. And because participants were asked to produce referential phrases that would be recorded and used at a later date, it follows that they received no feedback on the referential success of the phrases they produced from mention to mention (participants were not provided with any feedback at any point in the study).

A study of elicited reference phrases

Participants

64 native English speakers from the United Kingdom were recruited via Prolific (www.prolific.co) in September 2023. 1 was excluded due to a non-functioning microphone, 9 exceeded the one-hour time limit set for the experiment and were automatically excluded, and 3 chose to terminate

the experiment prematurely. Of the remaining 51 participants, 17 were excluded prior to transcription based on audio quality, missing recordings or failure to follow instructions, leaving us with complete datasets from 34 participants.

Table 2: The stimuli used, their SUBTLEX token frequencies (Center for Reading Research, 2013) and frequency ranks.

Noun	FR	Freq	Noun	FR	Freq
house	1	135288	zebra	21	991
car	2	55649	caterpillar	22	924
hand	3	55065	ruler	23	864
ball	4	42777	kangaroo	24	840
heart	5	40164	cigar	25	788
door	6	36644	ribbon	26	765
book	7	32652	snowman	27	673
dog	8	29680	windmill	28	644
eye	9	27123	flute	29	605
bed	10	26282	harp	30	506
table	11	25504	screwdriver	31	440
tv	12	22923	waistcoat	32	343
hair	13	21445	hanger	33	269
sun	14	20693	thimble	34	193
train	15	19163	pincers	35	182
glass	16	16716	grasshopper	36	162
clock	17	15610	seahorse	37	125
cat	18	13608	raccoon	38	74
cloud	19	11742	doorknob	39	32
bus	20	11117	nail file	40	12

Materials and methods

40 object stimuli originally created by Snodgrass and Vanderwart (1980) to be readily nameable were used as targets (in their colored versions: Face Recognition Lab, 2004). Namability was confirmed in a pretest. Because the lexical frequencies of the targets vary systematically (as is inevitably the case with natural language stimuli) they were split into two groups, comprising the most and least frequent stimuli (see Table 2) to control for any possible influence of these differences (although given the fact that the objects were selected to be easily namable, the possibility of the influence was not considered to be especially likely).

Each trial comprised five sequences. The experiment started with one of four items inside a circular boundary, then three more items appeared on the outside of the boundary. In the next 3 sequences, the remaining objects moved into the boundary. In the last 2 sequences, an object moved its position inside the boundary (Figure 2). After each sequence participants were asked to provide instructions that would allow a partner who only had knowledge of each objects' initial position (Figure 1) to recreate what they had seen. Because this design meant that participants observed and reported on the movements of each of the other objects prior to the final 2 movements, it also served to minimize the effects of lexical priming on our participants' behavior, given that the durations between mentions in this paradigm would be far longer than the timescales over which priming is usually observed (see e.g. Leinenger & Rayner, 2013).

Every participant completed two elicitation trials, counterbalanced so that one group saw high frequency items in the first trial and low frequency items in the second trial, or vice versa.

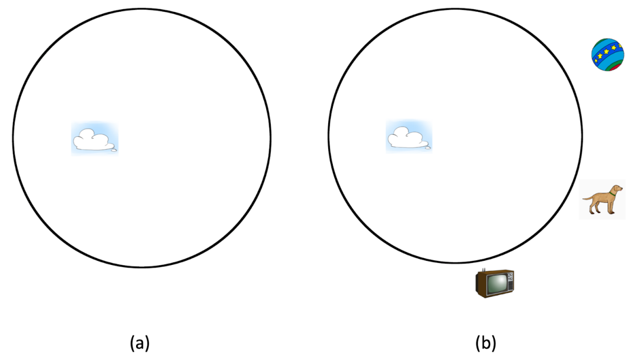

Figure 1: In the initial scene (a) an object appears in the boundary, followed by the appearance of 3 objects outside it. Panels (a) and (b) together thus comprise the total knowledge of the partner who must recreate the later moves.

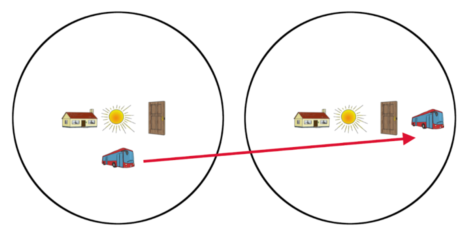

Figure 2: Movement of objects within the boundary: the bus moves from below the sun to the right of the door.

Results

Frequency

The verbal descriptions provided by each participant were then transcribed prior to analysis.

To examine whether the frequency of the targets had affected the reference phrases used to describe them, the reference phrases produced by each participant in each condition were extracted and compared. Note that although Dye et al (2018) report that on average low frequency English nouns are more likely to be modified by an adjective than high frequency English nouns, the fact that all of the targets were selected so as to be easily named by all of our participants gave us reason to doubt that this would be the case in this instance. The average mean length of the reference phrases produced by our participants is plotted in Figure 3, and as can be seen, there was no effect of frequency on reference phrase length ($t(66) = 0.2709$, $p = 0.79$).

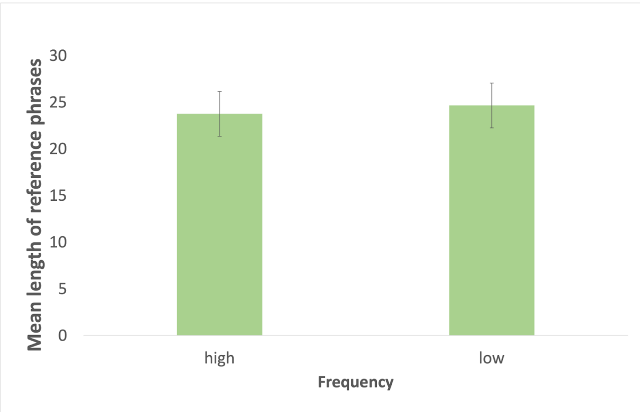

Figure 3: The average length of reference phrases produced in the high and low frequency conditions (error bars = SEM).

Further analyses, showed that there was no difference in the mean length between the high frequency trials of condition A and condition B ($t_{hf}(17) = 0.6036$, $p = 0.55$), or between the low frequency trials of condition A and condition B ($t_{lf}(17) = 0.8776$, $p = 0.39$), see Figure 4.

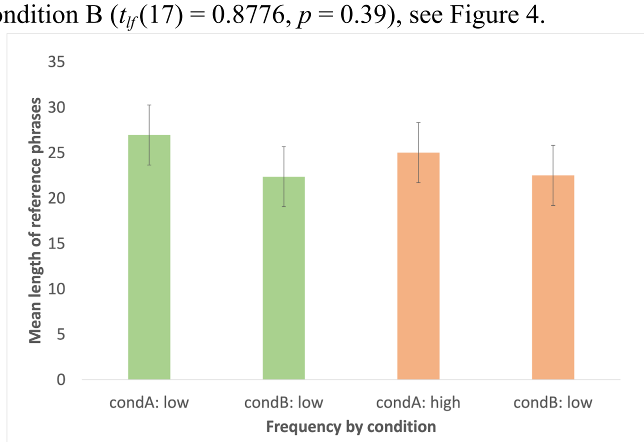

Figure 4: The average length of reference phrases produced for the high and low frequency trials in each condition (error bars = SEM).

Effects of first versus subsequent mention

We then examined whether, as hypothesized, participants would produce modified nouns more often at first mention as compared to second and later mentions. Because of the design of the movement sequences, we expected that most of the target nouns would be mentioned at least twice across each elicitation trial but that further mentions would occur rather less often. As this is what transpired, the relatively sparse data for these later mentions were collapsed for the purposes of analysis. Further, because of the results of our frequency analysis, the frequency differences between the targets were ignored.

The referential noun phrases extracted from the transcriptions of each participant's audio recording were again examined, and the probability of modification for each noun at first, second, and later mention calculated. Based on whether a noun was modified or not, we assigned each noun the value *vmod* = 1 for modified and *vnonmod* = 0 for non-modified. To calculate the probability of being modified we divided the sum of *vmod* by the sum of nouns per mention *n*.

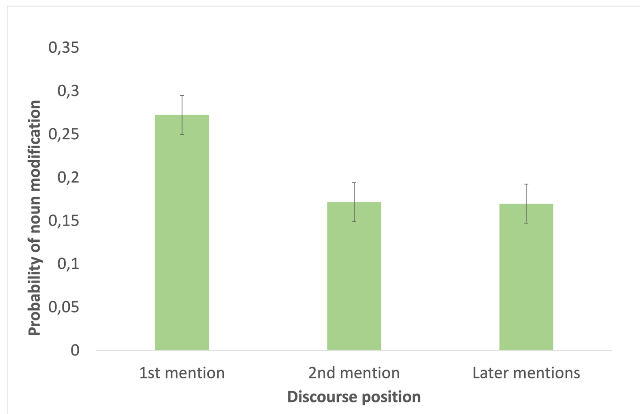

Figure 5: The probabilities of participants producing a modified noun phrase at first, second and subsequent mentions of each target noun (error bars = SEM).

As can be seen in Figure 5, the probability that a participant would produce a modified noun phase differed from first to second mention, and then the same lower likelihood of modification was observed for subsequent mentions ($F(1.66, 109.47) = 7.35$, $p = 0.002$). A pairwise t-test applying the Bonferroni correction for multiple comparisons revealed a significant effect for the comparison between the first and second mention ($t(66) = 3.3982$, $p_{adjusted} < 0.01$). Additionally, the comparison between the first and third mention showed a significant trend for an adjusted *p*-value $p_{adjusted} < 0.05$ for an assumed α value of 0.05. However, the difference between the third and second mention was not statistically significant ($p = 0.55$).

Linear mixed models were then applied to the data to corroborate these findings. The initial model was constructed with noun modification as the dependent variable and mention as the independent variable while accounting for participants as a random effect. The fixed effects results suggest that mention is indeed significantly associated with noun modification. The coefficient estimate exhibits a negative trend in the slope (-0.05), indicating an average decrease of 0.05 units in probability for each increase in mention. The *t*-value of $t(133) = -3.534$ confirmed this effect was significant ($p < 0.0001$).

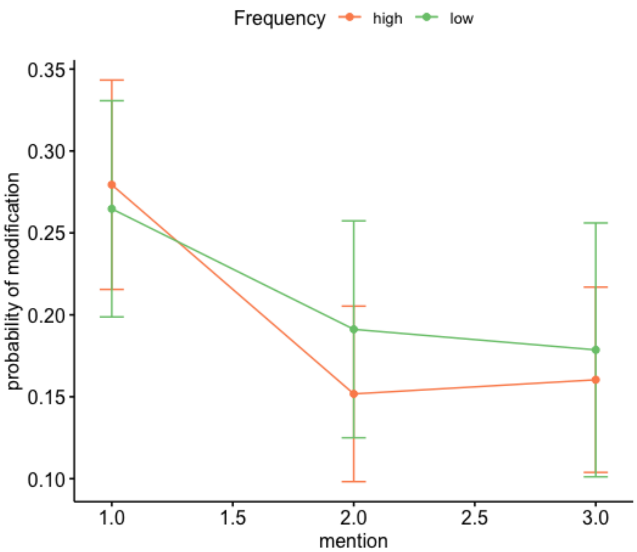

Figure 6: Differences in noun modification probability by the lexical frequencies of the targets (high vs low) and mention (error bars = SEM).

Finally, a secondary model was fitted for noun modification as the dependent variable, incorporating mention and frequency as the independent variables, while assuming an interaction between mention and frequency. We included participants as a random intercept, with an estimated standard deviation of 0.2216, showing a wider difference between the assumed intercept and the subjects. The results of the fixed effects reveal a downward trend for mention, indicating a significant effect of mention on noun modification with an average decrease in probability of 0.05 units ($t(66) = 2.630$, $p < 0.05$). Frequency and the interaction term of frequency and mention were not significant ($t_{freq}(85.22) = 0.168$, $p = 0.87$; $t_{freq:mention}(66) = 0.422$, $p = 0.67$), see Figure 6.

Do speakers dynamically adjust their probability models in speech?

Information theory (Hartley, 1938, Shannon, 1948) is a popular model for theorizing about human communication, and its methods have been widely applied in studies of speech and language (Gibson, 2019). As we noted earlier, in an analogy to the way information theory defines information in terms of the probabilities with which events such as symbols in a code occur, language researchers have tended to assume the information provided by words can be defined in terms of their relative likelihood of occurring in a corpus (i.e., in terms of their frequencies). However, this way of estimating the information in words implicitly assumes that words occur at a uniform rate across contexts. This assumption is empirically undermined by numerous observations that indicate that empirically, word occurrences are bursty (Katz, 1996, Altmann et al, 2009, Slone et al, 2013): for most words, their likelihood of occurrence in most contexts is very, very small; however, a given word does appear in a context, the likelihood of its recurrence is much, much higher.

Although previous studies have found results that are consistent with the idea that speakers may be subject to the dynamic probabilities of words across contexts, for example, in studies of reference phrase and speech production (Krauss, & Weinheimer, 1964, Aylett, 2000, Hawkins et al, 2020) these results do not provide strong evidence for it because they results can be explained by other, perhaps simpler discourse factors. Accordingly, we conducted an elicitation study to examine whether speakers are sensitive to the dynamic probabilities that a sensitivity to burstiness would imply. Consistent with this proposal our participants produced numerous seemingly redundant adjectives at first mention. Given that the objects/nouns we used in our task were entirely unambiguous, it seems reasonable to assume that their initial use of these seemingly redundant adjectives was because far from being redundant, these adjectives actually made these nouns more predictable (Dye et al 2018; Rubio-Fernandez & Jara-Ettinger, 2020) prior to their having been mentioned.

However, despite their receiving no feedback on whether their initial referential phrases had succeeded communicatively, our participants then proceeded to produce significantly fewer seemingly redundant adjectives at second and subsequent mentions. If we accept that their initial use of the seemingly redundant adjectives was to make these nouns more predictable, then given that they received no other information from the task context, and given that it seems reasonable to infer that their decreased use of adjectives – which initially served to increase the likelihood of the target nouns – is because the likelihood probabilities of the targets in context have been revised. This serves to make the targets more likely, which, of course, now makes the adjectives more redundant. The most obvious explanation here is that our participants revised their estimations of the likelihood that the target nouns would recur after their first appearance. Which of course, is exactly what we predicted that speakers sensitive to burstiness would do.

While the burstiness of lexical distributions has long been understood (Katz, 1996) its potential impact on linguistic probability estimates has not been appreciated in current probabilistic approaches to language processing. The results we report not only highlight the importance of considering this factor in analyses of the informativity of noun phrases, but rather they suggest that burstiness and the dynamic probability models used by language users ought to be considered as factors that could potentially influence behavior – and hence analysis – in any area of language processing (and likely add another theoretical explanation for why contextual diversity has often been shown to be a better prediction of lexical behavior than relative frequency, see e.g., Johns et al, 2016).

References

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLOS one*, 4(11), e7678

Aylett, M. (2000). Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech. In *Sixth International Conference on Spoken Language Processing* (pp. 646-649). ISCA

Aylett, M., & Turk, A. (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

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92-111.

Davies, M. (2009). The 385+ million word corpus of contemporary american english (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistic*s, 14(2), 159–190

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2017). A functional theory of gender paradigms. *In Perspectives on morphological organization*(pp. 212-239). Brill.

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2018). Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. *Topics in cognitive science*, 10(1), 209-224.

Estoup, J. B. (1916) *Gammes Stenographiques*. Institut Stenographique de France, Paris

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language*. Trends in cognitive sciences,* 23(5), 389-407.

Hartley, R. V. L. (1928). Transmission of Information. *Bell System Technical Journal*. 7 (3): 535–563. doi:10.1002/j.1538-7305.1928.tb01236.x.

Hawkins, R. D., Frank, M. C., & Goodman, N. D. (2020). Characterizing the dynamics of learning in repeated reference games. *Cognitive Science*, 44(6), e12845.

Jaeger, T., & Levy, R. (2006). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, 19.

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic bulletin & review*, 23, 1214-1220.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological studies in language*, 45, 229-254

Katz, S. M. (1996). Distribution of content words and phrases in text and language modelling. *Natural language engineering*, 2(1), 15-59.

Kohler, K. J. (1998). The disappearance of words in connected speech. *ZAS Papers in Linguistics*, 11, 21-33.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3), 343–346

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science,* 1, 113-114.

Leinenger, M., & Rayner, K. (2013). Eye movements while reading biased homographs: Effects of prior encounter and biasing context on reducing the subordinate bias effect. *Journal of Cognitive Psychology*, 25(6), 665-681.

Piantadosi, S. (2023). *Modern language models refute Chomsky's approach to language*. Lingbuzz Preprint, lingbuzz, 7180.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526-3529.

Ramscar, M. (2021). How children learn to communicate discriminatively. *Journal of Child Language*, 48(5), 984-1022.

Rubio-Fernandez, P., & Jara-Ettinger, J. (2020). Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences,* 117(24), 13399-13404.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140-155

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379-423

Slone, L. K., Abney, D. H., Smith, L. B., & Yu, C. (2023).The temporal structure of parent talk to toddlers about objects. *Cognition*, 230, 105266

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, 6(2), 174.

Thompson, D. (1995). *The concise Oxford dictionary of current English*. Oxford: Oxford University Press

Tomaschek, F., Arnold, D., Sering, K., Tucker, B. V., van Rij, J., & Ramscar, M. (2021). Articulatory variability is reduced by repetition and predictability. *Language and speech*, 64(3), 654-680.

Zipf, G. (1936). *The Psychobiology of Language*. London: Routledge