1

Experiment 2: Semantic predictability facilitates comprehension of degraded speech in a graded manner

1.1 Background

In the previous chapter, we showed that semantic predictability facilitates comprehension of degraded speech. There was a significant difference in word recognition accuracies between low and high predictability sentences (i.e., contextual facilitation), when listeners were instructed to attend to the entire sentence including the context. However, it was not clear how well listeners decoded the context word, i.e., if they correctly identified the verb in each degradation level.

In this chapter, we examine if semantic predictability facilitates comprehension of degraded speech in a graded manner; whether prediction is probabilistic, or if it is an all-or-none phenomenon. We first determine if listeners' are able to decode the context, at all instances, when they are attending to the entire sentence. However, context is neglected in measuring language comprehension in most of the existing studies (see Section X.X.X) Therefore, the metric of language comprehension in the following experiments takes into account the trials in which participants correctly identify the context evoking words.

1.1.1 Nature of predictability: Probabilistic or all-or-none?

There are two thoughts, or debate about the nature of predictability effects. One line of studies argue that prediction is all-or-none and deterministic [...CITE...]. For example, in garden path phenomenon, a parser first tries to predict the simplistic structure of a sentence. If this parsing is disconfirmed by the bottom-up input then the parser backs off and restarts – it reinterprets the context and predicts a new sentence structure. Therefore, according to XYZ, how a parser resolves garden path sentences is an evidence for all-or-none prediction. We use only one of many possible continuations in a sentence. XYZ argue that it is metabolically expensive to predict many parallel linguistic units when only most of them are going to be discarded. Therefore, evolution would not pick a language processing system that predicts multiple parallel linguistic units. Another line of studies show that predictions are probabilistic, or graded. Nieuwland et al. [1] showed in a large-scale replication study of DeLong, Urbach, and Kutas [2] that the N400 amplitude at the sentence-final noun is directly proportional to its cloze probability across a range of high- and low-cloze words. Heilbron et al. [3] also showed that a probabilistic prediction model outperforms a constrained guessing model, suggesting that linguistic prediction is not limited to highly predictable sentence endings, but it operates broadly in a wide range of probable sentence endings. The studies mentioned above were either reading studies, or were conducted with clean speech.

To our knowledge, only one study by Strauß, Kotz, and Obleser [4] has explicitly studied the nature of prediction in degraded speech comprehension. They presented listeners with severely degraded (4 channels noise vocoded), moderately degraded (8 channels noise vocoded), and clear (non-degraded) speech. Target word predictability was varied by manipulating its expectancy (i.e., how expected the target word is given the verb) and typicality (i.e., co-occurrence of target word and the preceding verb). They reported that at a moderate level of spectral degradation, N400 responses at strong-context, low-typical words and weak-context, low-typical words were largest. N400 responses at the latter two were not statistically different

from each other. However, the N400 response was smallest at highly predictable (strong-context, high-typical) words. The authors interpreted these findings as a facilitatory effect of sentence predictability which might be limited to only highly predictable sentence endings at a moderate level of spectral degradation. Based on these findings, Strauß, Kotz, and Obleser [4] proposed an 'expectancy searchlight model'. According to the expectancy searchlight model, listeners form 'narrowed expectations' from a restricted semantic space when the sentence endings are highly predictable. For less predictable sentence endings, listeners cannot preactivate those less predictable sentence endings in an adverse listening condition. Therefore, the expectancy searchlight model of Strauß, Kotz, and Obleser [4] is in line with all-or-none prediction account. These theoretical accounts have not been further explored in degraded speech comprehension. However, it has been shown that listeners rely on context when speech is moderately degraded.

1.1.2 Predictability effects in degraded speech comprehension

When the bottom-up perceptual input is less reliable, listeners rely more on top-down predictions by narrowing down their predictions to smaller sets of semantic categories or words, in adverse listening conditions like noisy environment with background noise [...CITE...], reverberation [...CITE...], and degraded speech [e.g., 5, 6].

In their studies, Obleser and colleagues [7–9], used sentences of two levels of semantic predictability (high and low) and systematically degraded the speech signal by passing it through various numbers of noise vocoding channels ranging from 1 to 32 in a series of behavioral and neuroimaging studies. They found that semantic predictability facilitated language comprehension at a moderate level of speech degradation (at 4 channels, or 8 channels noise vocoding). That is, participants relied on the sentence context when the speech signal was degraded but intelligible enough. Accuracy of word recognition was found to be higher for highly predictable target words than for lowly predictable target words at such moderate levels of speech degradation [8]. For the extremes, i.e., when the speech signal was highly

degraded or when it was clearly intelligible, the word recognition accuracy was similar across both levels of sentence predictability, meaning that predictability did not facilitate language comprehension. The conclusion of these findings is that at moderate levels of degradation, participants rely more on the top-down prediction generated by the sentence context and less on the bottom-up processing of unclear, less reliable, and degraded speech signal [10]. Reliance on prediction results in higher word recognition accuracy for target words with high-cloze probability than for the target words with low-cloze probability. In the case of a heavily degraded speech signal, participants may not be able to understand the sentence context and, therefore, be unable to predict the target word; or their cognitive resources may already be occupied by decoding the signal, leaving little room for making predictions [. . . CITE. . . look Ryskin2021]. Thus, there is no differential effect of levels of sentence predictability. On the other extreme, when the speech is clear and intelligible (at the behavioral level, i.e., when the participants respond what the target word of the sentence is), participants recognize the intelligible target word across all levels of sentence predictability. Hence, no differential effect of levels of predictability of target word can be expected.

1.2   Procedure

1.2.1   Participants

We recruited two groups of participants via Prolific Academic and assigned them to one of the two groups: 'unpredictable channel context' (n=48; $\bar{x} \pm$ SD = 24.44 $\pm$ 3.5 years; age range = 18-31 years; 16 females) and 'predictable channel context' (n=50; $\pm$ SD = 23.6 $\pm$ 3.2 years; age range = 18-30 years; 14 females). All participants were native speakers of German residing in Germany. Exclusion criteria for participating in this study were self-reported hearing disorder, speech-language disorder, or any neurological disorder. All participants received monetary compensation for their participation. The study was approved by Deutsche

Gesellschaft für Sprachwissenschaft (DGfS) Ethics Committee, and the participants provided consent in accordance with the Declaration of Helsinki.

1.2.2 Stimuli

We used the same stimuli described in Section X.X.X in Chapter X.X. The stimuli were digital recordings of 360 German sentences spoken by a female native speaker of German in a normal rate of speech. All sentences were in present tense consisting of pronoun, verb, determiner, and object (noun) in the Subject-Verb-Object form. We used 120 unique nouns to create three catetoreis of sentences – high predictability sentences (HP sentencs), medium predictability sentences (MP sentences) and low predictability sentences (LP sentences) – that differed in cloze probability of sentece final target words. (See Appendix A for examples.) Their mean cloze probabilities were 0.022 ± 0.027 ($\bar{x} \pm$ SD; range = 0.00 - 0.09) for LP sentences, 0.274 ± 0.134 ($\bar{x} \pm$ SD; range = 0.1 - 0.55) for MP sentences, and 0.752 ± 0.123 ($\bar{x} \pm$ SD; range = 0.56 - 1.00) for HP sentences. The distribution of cloze probability across LP, MP and HP sentences are shown in Figure X.X.

In the *unpredictable channel context*, each participant was presented with 120 unique sentences: 40 HP, 40 MP and 40 LP sentences. Channel condition was also balanced across each sentence type, i.e., in each of HP, MP, and LP sentences, ten sentences passed through each noise vocoding channels – 1, 4, 6, and 8 – were presented. This resulted into 12 experimental lists. The sentences in each list were pseudo-randomized, that is, not more than 3 sentences of same noise condition (i.e., same noise vocoding channel), or same predictability condition appeared consecutively. This randomization confirmed the uncertainty of next-trial speech quality (or degradation) in the global context of the experiment.

The same set of stimuli and experimental lists were used in the *predictable noise context*. Each participant was presented with 120 unique sentences blocked by channel conditions, i.e., blocked by noise vocoding channels. There were four blocks of stimuli. Thirty sentences were presented in each of the four blocks. In the first block, all sentences were 8 channels noise vocoded, followed by blocks

of 6 channels, 4 channels, and 1 channel noise vocoded speech consecutively [5]. Within each block, 10 HP, 10 MP and 10 LP sentences were presented. All the sentences were pseudo-randomized so that not more than three sentences of the same predictability condition appeared consecutively in each block. This ascertained there was a certainty of next-trial speech quality (within each block) and an uncertianty of next-trial sentence predictability across all four blocks.

1.2.3 Procedure

Participants were asked to use headphones or earphones. A prompt to adjust loudness was displayed at the beginning of the experiment: A noise-vocoded sound not used in the main experiment was presented, and participants were asked to adjust the loudness at their level of comfort. One spoken sentence was presented in each trial. Eight practice trials were presented before presenting 120 experimental trials. They were asked to enter what they had heard (i.e., to type in the entire sentence) *via* keyboard. Guessing was encouraged. At the end of each trial, they were asked to judge their confidence in their response on a scale of 1 to 4, 1 being 'just guessing' to 4 being 'highly confident'. The response was not timed. The experiment was about 40 minutes long.

1.3 Analyses

In the sentences used in our experiment, verbs evoke predictability of the sentence-final noun. Therefore, the effect of predictability (evoked by the verb) on language comprehension can be rightfully measured if we consider only those trials in which participants identify the verbs correctly. Verb-correct trials were considered as the sentence in which participants realized the context independent of whether they correctly understood the sentence final target noun. Morphological inflections and typos were considered as correct. We first filtered out those trials in which verbs were not identified correctly, i.e., trials with incorrect verbs. Therefore, we included 2469 out of 5760 trials in *unpredictable channel context* and 2374 out

of 6000 trials in *predictable channel context* from the analyses. The 1 channel noise vocoding condition was dropped from the analyses as there were no correct responses in any of the trials in this condition.

We preprocessed and analysed data in R-Studio (Version 3.6.1; R Core Team, 2019) following the procedure described in Chapter 4.4.4. Response accuracy was analyzed with Generalized Linear Mixed Models (GLMMs) with lme4 [11] and lmerTest [12] packages. Binary responses (correct/incorrect) for all participants in both groups (predictable channel context and unpredictable channel context) were fit with a binomial logistic mixed-effects model. Noise condition (categorical; 4-, 6-, and 8-bands), target word predictability (categorical; HP, MP, LP), global channel context (categorical; predictable channel context and unpredictable channel context), and the interaction of noise condition and target word predictability were included in the fixed effects.

We first fitted a model with maximal random effects structure that included random intercepts for each participant and item [13]. Both, by-participant and by-item random slopes were included for noise condition, target word predictability and their interaction. To find the optimal model for the data, non-significant higher-order interactions were excluded from the fixed-effects structure (and from the random-effects structure) in a stepwise manner. Model selection was based on Akaike Information Criterion (AIC) [14, 15] unless otherwise stated. Random effects not supported by the data that explained zero variance according to singular value decomposition were excluded to prevent overparameterization. This gave a more parsimonious model [16]. Such a model was then extended separately with: i) item-related correlation parameters, ii) participant-related correlation parameter, and iii) both item- and participant-related correlation parameters. The best fitting model among the parsimonious and extended models, based on AIC, was then selected as the optimal model for our data.

We applied treatment contrast for noise condition (8-bands as a baseline; factor levels: 8-bands, 4-bands, 6-bands) and sliding difference contrast for target word predictability (factor levels: MP, LP, HP) and channel context (factor levels:

unpredictable, predictable). The results from the optimal model are shown in Table X.X.X, and are reported below in the Results section.

1.4 Results

1.5 Conclusion

References

[1] Mante S Nieuwland et al. "Large-scale replication study reveals a limit on probabilistic prediction in language comprehension". In: *eLife* 7 (2018), pp. 1–24.

[2] Katherine A DeLong, Thomas P Urbach, and Marta Kutas. "Probabilistic word pre-activation during language comprehension inferred from electrical brain activity". In: *Nature Neuroscience* 8.8 (July 10, 2005), pp. 1117–1121. URL: http://dx.doi.org/10.1038/nn1504.

[3] Micha Heilbron et al. "A hierarchy of linguistic predictions during natural language comprehension". In: *bioRxiv* (2020).

[4] Antje Strauß, Sonja A. Kotz, and Jonas Obleser. "Narrowed Expectancies under Degraded Speech: Revisiting the N400". In: *Journal of Cognitive Neuroscience* 25.8 (Aug. 2013), pp. 1383–1395. URL: http://dx.doi.org/10.1162/jocn_a_00389.

[5] Signy Sheldon, M. Kathleen Pichora-Fuller, and Bruce A. Schneider. "Priming and sentence context support listening to noise-vocoded speech by younger and older adults". In: *The Journal of the Acoustical Society of America* 123.1 (Feb. 2008), pp. 489–499. URL: http://dx.doi.org/10.1121/1.2783762.

[6] Ruth E. Corps and Hugh Rabagliati. "How top-down processing enhances comprehension of noise-vocoded speech: Predictions about meaning are more important than predictions about form". In: *Journal of Memory and Language* 113 (Aug. 2020), p. 104114. URL: http://dx.doi.org/10.1016/j.jml.2020.104114.

[7] Jonas Obleser et al. "Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions". In: *Journal of Neuroscience* 27.9 (Feb. 28, 2007), pp. 2283–2289. URL: http://dx.doi.org/10.1523/JNEUROSCI.4663-06.2007.

[8] Jonas Obleser and S. A. Kotz. "Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network". In: *Cerebral Cortex* 20.3 (June 26, 2009), pp. 633–640. URL: http://dx.doi.org/10.1093/cercor/bhp128.

[9] Jonas Obleser and Sonja A. Kotz. "Multiple brain signatures of integration in the comprehension of degraded speech". In: *NeuroImage* 55.2 (2011), pp. 713–723. URL: https://www.sciencedirect.com/science/article/pii/S1053811910016034.

[10] Jonas Obleser. "Putting the Listening Brain in Context". In: *Language and Linguistics Compass* 8.12 (Dec. 2014), pp. 646–658. URL: http://dx.doi.org/10.1111/lnc3.12098.

[11] Douglas Bates et al. "Fitting Linear Mixed-Effects Models Usinglme4". In: *Journal of Statistical Software* 67.1 (2015). URL: http://dx.doi.org/10.18637/jss.v067.i01.

[12] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. "lmerTest Package: Tests in Linear Mixed Effects Models". In: *Journal of Statistical Software* 82.13 (2017). URL: http://dx.doi.org/10.18637/jss.v082.i13.

[13] Dale J. Barr et al. "Random effects structure for confirmatory hypothesis testing: Keep it maximal". In: *Journal of Memory and Language* 68.3 (Apr. 2013), pp. 255–278. URL: http://dx.doi.org/10.1016/j.jml.2012.11.001.

[14] C. E. Grueber et al. "Multimodel inference in ecology and evolution: challenges and solutions". In: *Journal of Evolutionary Biology* 24.4 (Jan. 27, 2011), pp. 699–711. URL: http://dx.doi.org/10.1111/j.1420-9101.2010.02210.x.

[15] Shane A. Richards, Mark J. Whittingham, and Philip A. Stephens. "Model selection and model averaging in behavioural ecology: the utility of the IT-AIC framework". In: *Behavioral Ecology and Sociobiology* 65.1 (Aug. 2011), pp. 77–89. URL: http://dx.doi.org/10.1007/s00265-010-1035-8.

[16] Douglas Bates et al. "Parsimonious Mixed Models". In: *arXiv* (2015). arXiv: 1506.04967. URL: http://arxiv.org/abs/1506.04967.