Cumulative usage effects in inflectional paradigm: Probability of being affixed

Daiki Hashimoto[1]

[1]*University of Tokyo (Japan)*
daiki@phiz.c.u-tokyo.ac.jp

Past usage patterns affect production of a linguistic unit in future speech. This effect is called a "cumulative usage effect." For example, words with lower informativity, which have often been affected by probabilistic reduction, are produced with shorter duration even when they occur in a less predictable context (Seyfarth [1]; Sóskuthy & Hay [2]). This effect can be neatly captured by Exemplar Theory (Pierrehumbert [3]), which consists of the following key hypotheses:

H1: Sounds encountered in speech are stored with detailed phonetic information as "exemplars" in the mental space.

H2: Exemplars with similar properties (e.g., phonetic, semantic, and social information) cluster together, and "categories" are formed.

H3: A speaker begins speech production by activating an intended category. Then, the speaker chooses a couple of exemplars belonging to the category, and average their phonetic values to form a production target.

The informativity-oriented reduction can be captured, by positing that words that are likely to undergo probabilistic reduction may be represented by a larger number of exemplars with reduced phonetic properties.

This raises a research question of whether cumulative usage effects are also observed in relation to inflectional paradigms of lexemes. The aim of this talk is to inform our understanding of how lexemes are represented in our mind by addressing the following concrete RQ:

RQ: "Is the duration of lexeme-final /s/ (e.g., *bu̱s* and *bu̱ses*) affected by the likelihood of being followed by bound morphemes (i.e., *-es*, *-ed*, and *-ing*)?"

It is known that a coronal fricative phoneme /s/ is produced with shorter duration in word-medial position than in word-final position in American English (Umeda [4]): therefore, it can be expected that lexeme-final /s/ is produced with shorter duration in New Zealand English when they are attached by bound morphemes (e.g., *bu̱ses* and *i̱ced*) in comparison with when they are not (e.g., *bu̱s* and *i̱ce*). We crucially hypothesize that exemplars encoding affixed forms and unaffixed forms of a lexeme are closely represented, because they share a variety of properties (e.g., phonological information, syntactic information, semantic information, and social information). The following prediction can be put forward on the bases of the hypotheses:

Prediction: Lexeme-final /s/ may be produced with shorter duration in lexemes that are in general more likely to be followed by bound morphemes (e.g., *-s*, *-ed*, and *-ing*).

This prediction is illustrated in Figure 1. According to our data, which is explained below, *chase* is a word that is likely to be affixed, while *base* is a word that is unlikely to be affixed. Namely, *chase* is likely to be produced with shorter lexeme-final /s/ because of the following bound morphemes, and consequently this lexical category may be represented by a larger number of exemplars with shorter lexeme-final /s/. Because of this representation, it can be predicted that the lexeme-final /s/ in *chase* may be produced with shorter duration in general even when it is not followed by bound morphemes. This is because an exemplar with shorter /s/ is more likely to be chosen in production of this word. The reverse is true for the lexical category *base*.
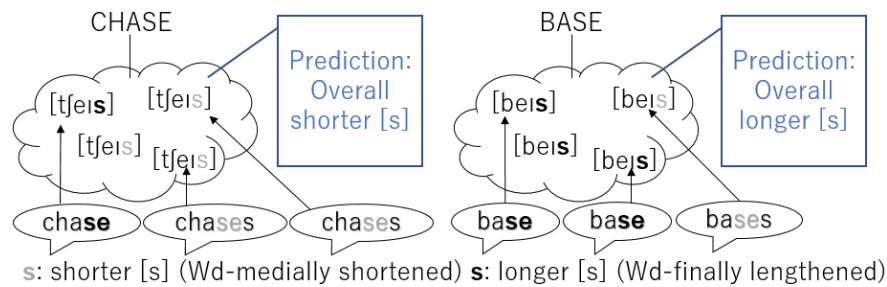
Fig. 1 Prediction about cumulative usage effects on lexeme-final /s/

Lexemes ending with non-morphemic /s/, which can be countable nouns or verbs, were collected from CELEX (Baayen et al. [5]). Then, the recorded tokens of the unaffixed and affixed forms (i.e., followed by bound morphemes) were collected from ONZE Corpus (Gordon et al. [6]). Some tokens were removed for acoustic or morphological reasons. Finally, 16,059 tokens with lexeme-final /s/ were collected: 9,120 unaffixed tokens and 6,939 affixed tokens.

The 16,059 tokens of lexeme-final /s/-sounds were hand-fitted into a mixed-effects linear regression model using the *lmer* function in the *lme4* library implemented in R. The model includes some control variables such as word predictability and prosodic factors. As expected, lexeme-final /s/ is shorter when they are affixed ($p<0.05$). As for the probability of being affixed, the effect is significant only in affixed forms ($p<0.05$), but it is non-significant in unaffixed forms ($p=0.46$), see Figure 2. This result is in line with our Prediction, and it suggests that cumulative usage effects are also observed in relation to inflectional paradigms of lexemes. This can be captured by positing that our exemplar space is developed in accordance with speech experience, that is, a category of a lexeme that is likely to be affixed is represented by a larger number of exemplars that are NOT affected by word-final lengthening.
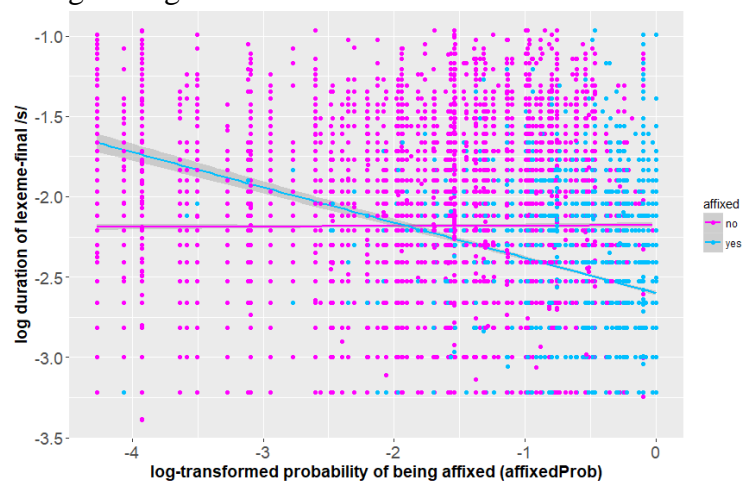
Fig.2 Relationship between log-duration of /s#/ and log-probability of being affixed

References

[1] Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133, 140-155.
[2] Sóskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition*, 166, 298-313.
[3] Pierrehumbert, J. B. (2001). Exemplar dynamics, word frequency, lenition, and contrast, In J. Bybee, & P. Hopper (Eds) *Frequency effects and the emergence of linguistic structure* (pp. 135-157). Amsterdam: John Benjamins.
[4] Umeda, N. (1977). Consonant duration in American English. *Journal of Acoustic Society of America*, 61, 846-858.
[5] Baayen, H., Richard P., Léon G. (1995). *The CELEX Lexical Database*. Philadelphia: Linguistic Data Consortium (University Pennsylvania).
[6] Gordon, E., Maclagan, M., Hay, J. (2007). The ONZE Corpus. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (Eds), *Models and methodsin the handling of uncontroversial digital corpora. Diachronic corpora* (Vol. 2, pp.82-104). Basingstoke, Hampshire: Palgrave Macmillan.