Probabilistic phonotactics in visual word recognition within and across morphological boundaries

Chiara Celata*, Giulia Bracco°
**Scuola Normale Superiore di Pisa, °Università di Salerno*

Transition probabilities (TPs) refer to the conditional probability with which a segment σ occurs in a corpus, given *N* preceding segment(s). TPs are intuitively similar to a frequency measure but produce an opposite effect on word recognition. High TPs imply high neighborhood density and therefore inhibitory effects on word or syllable recognition. TPs have been shown to influence infants' discrimination of native language sounds and words, children's and adults' spoken word recognition, wordlikeness ratings, speech segmentation and speech production (e.g. [1], [2], [3], [4], [5], [6]).

This study investigates the role of TPs in the processing of consonant clusters (as opposed to consonant+vowel sequences) and of complex words (as opposed to monomorphemic words) in two experiments of visual word recognition. Consonant clusters are challenging phonological objects not only for phonetic complexity but also for cognitive processing [7] and are known to convey information on the morphological structure of words ([8], [9] [10]).

Two fragment priming lexical decision tasks, in which an Italian target word (e.g. *borsa* 'bag') is primed by its initial fragment (*bor*), are performed by 124 native speakers. In both experiments, based on a reference corpus of written Italian [11], we calculate TP and token frequency values for each biphone corresponding to the last segment of the fragment prime and the following segment in the target (i.e. ***rs*** in *bo**rs**a*). TP and frequency values are equally calculated for the sequence corresponding the fragment prime and the following segment (i.e. ***bors*** in ***bors****a*). In the first experiment (E1), biphones are CV or CC sequences (e.g. *bir-bi**ri**llo* 'skittle' vs. *bor-bo**rs**a* 'bag') and the aim is to verify whether biphone's and sequence's TPs play a different role in consonant clusters as opposed to consonant+vowel sequences. The second experiment (E2) focuses on consonant clusters and compares monomorphemic words (e.g. *bis-bi**st**ecca* 'beef-steak') with prefixed words in which the biphone crosses a morphemic boundary (e.g. *bis-bi**sn**onna* 'great-grandmother', 'bis' being an Italian prefix). Linear mixed models [12] are run with log-transformed reaction times as the dependent variable, Target Frequency, Prime Frequency, Biphone Frequency, Biphone TP, Sequence Frequency and Sequence TP as independent factors with fixed effects, Word and Subject as factors with random effects.

The results of E1 (Figure 1) show that latencies in the recognition of CC and CV targets are predicted by different factors. Target Frequency has a strong role in both CC and CV models; however, the models differ for the role of sublexical factors. In particular, there is an inhibitory effect of Sequence TP on target recognition in the CC model, suggesting that probabilistic information is processed in the recognition of targets containing consonant clusters and not of targets containing CV sequences. The results of E2 additionally show that probabilistic and frequency-related information influence the processing of prefixed and monomorphemic words differently. Sequence TP has an inhibitory effect on the recognition of monomorphemic targets, consistently with the CC results in E1, but an opposite facilitatory effect is found in the recognition of prefixed words. Moreover, both Target Frequency and Prime Frequency have a facilitatory effect on prefixed words, whereas Prime Frequency has an inhibitory effect on monomorphemic words, suggesting a competition effect for the latter but not for the former.

The results of this study enlarge the scope of current investigation on probabilistic phonotactics, showing, for an understudied language, the role of TPs in the processing of consonant clusters and of morphologically complex words. Implications for the phonotactics-morphology interface and for theories of visual word recognition are also discussed.

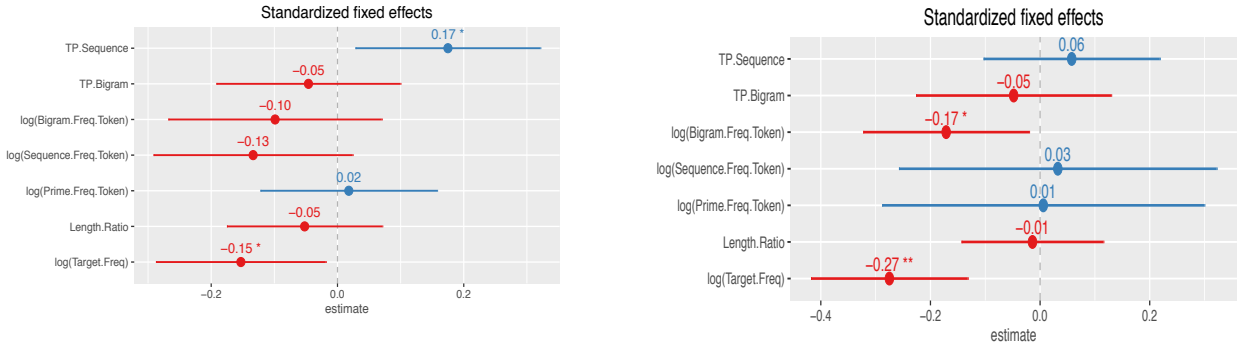

Figure 1. *Standardized fixed effects for the CC (left) and CV (right) models in E1. Stars (*) represent the significance levels : '***' = 0.001, '**' = 0.01 , '*' = 0.05.*

[1] Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645.

[2] MacRoy-Higgins, M., Shafer, V.L., Schwartz, R.G. & Marton, K. (2014) The influence of phonotactic probability on word recognition in toddlers. *Child Language Teaching and Therapy* 30(1), 117-130 .

[3] Vitevitch, M.S., Luce, P.A., Charles-Luce, J. & Kemmerer, D. (1997) Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech* 40, 47-62.

[4] Mattys, S.L. & Jusczyk, P.W. 2001. Do infants segment words or recurring contiguous Patterns? *Journal of Experimental Psyhcology: Human Perception and Performance* 27(3), 644-655.

[5] Gustafson, E. & Goldrick, M. 2017. The role of linguistic experience in the processing of probabilistic information in production. *Language, Cognition and Neuroscience*, 33(2), 211-226.

[6] Vitevitch, M. S., & Luce, P. A. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374-408.

[7] Hoole, P., Bombien, L., Pouplier, M., Mooshammer, C. & Kühnert, B. 2012. *Consonant clusters and structural complexity*. Mouton de Gruyter.

[8] Dressler, W.U. & Dziubalska-Kolaczyk K. 2006. Proposing morphonotactics. *Italian Journal of Linguistics* 18, 249-266.

[9] Calderone, B., Celata, C., Korecky-Kröll, K. & Dressler, W.U. 2014. A computational approach to morphonotactics: evidence from German. *Language Sciences* 46 (A), 59-70.

[10] Celata, C., Korecky-Kröll, K., Ricci, I. & Dressler, W.U. 2015. Phonotactic processing and morpheme boundaries: word- final /Cst/ clusters in German. *Italian Journal of Linguistics* 27(1), 85-110.

[11] Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C. & Thornton, A.M. 2005. *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS)*. http://linguistica.sns.it/CoLFIS/Home.htm

[12] Baayen, R.H., Davidson, D.J. & Bates, D.S. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.