(The) unbearable lightness of (the) English articles

Dagmar Divjak & Petar Milin
University of Birmingham
d.divjak@bham.ac.uk & p.milin@bham.ac.uk

English articles are such minion language units: frequent and hence unremarkable, they arguably carry very little communicative weight, yet are passed on from one generation of speakers to the next. Linguists struggle to formulate an account that would adequately capture their use (see, e.g., Abbott, 2006), and articles form notorious stumbling blocks for foreign learners whose native tongues do not solidify the base-of-use for articles (Thomas, 1989). We will consider the learnability of English articles from a usage-based and learning theoretical perspective. On a usage-based approach to language, abstractions would emerge from exposure to the ambient language. The wider sentence context has proven invaluable for predicting syntactic and semantic choice, and should also apply to predicting articles in context. Yet, the road to mastery is long. Young children often omit articles from their speech altogether (Gerken, 1996) or produce them as fillers (Peters, 2001). Matthews et al. (2009) found that children aged 2 to 4, when required to produce 'the', were below or at chance in all age groups, while adults remained below ceiling. What could be the underlying reason for such results?

We start from the assumption that, in usage, articles and nouns are perceptually inseparable and the choice for an article is not as free as grammar books might lead one to assume; article-noun co-occurrence is highly constrained by context, and they could be considered as forming one unit. More than half of the 1000 most frequently attested nouns in the British National Corpus appear predominantly without article (524) or with a definite article (400), while only a minority prefers the indefinite article (76); this preference is licensed in their typical contexts. The nouns' preference for co-occurrence with one of the three options (*a/an*, *the*, zero) also affects those nouns' uncertainty (entropy) and distinctiveness (relative entropy): *a/an*-dominant nouns are highly likely to allow other articles too (*the* or *zero*) and are very difficult to predict from the determiner; *the-* and zero-dominant nouns do not differ in the extent to which they allow other determiners but *the*-dominant nouns are harder to predict from their determiner than zero-dominant nouns.

Native speakers of English are exposed to the contextual properties that constrain this article choice. Native speakers of languages without articles, on the other hand, have learned to associate the exact same context with the same noun without need for articles, leading them to produce an infelicitous combination or to dispense with articles altogether. Simple computational learning simulations (using the Rescorla-Wagner rule (Rescorla & Wagner 1972) of the divergent experiences of native (L1) and non-native (L2) speakers yield two intriguing results. For L1, if the article is treated together with the noun as a unit, then the wider context becomes predictive as a function of the systematicity of its co-occurrence with the target noun; this provides support for our starting hypothesis. For L2, simulation treated the article and the noun as independent outcomes. In that learning set-up the article only becomes somewhat predictable if the language serves as "context" (L1 vs L2): L2 learners know that articles ought to appear in L2 but not in L1, which is an insight that is of little practical value. Learning theory provides us with relevant concepts to make sense of these simulation results, and using findings from a large-scale simulation study on a 1.2-billion-word Internet corpus, we will discuss how pre-exposure makes English articles so unbearably hard to pin down.

Selected References

Abbott, B. (2006). Definiteness and Indefiniteness. In L. R. Horn & G. L. Ward (Eds.), *The handbook of pragmatics*. Oxford: Blackwell.

Gerken, L. (1996). Prosodic structure in young children's language production. *Language*, 72(4), 683-712.

Matthews, D., Lieven, E. & M. Tomasello (2009) The development of referring expressions from birth to four years. *Proceedings of the Workshop "Production of Referring Expressions"* (PRE-Cogsci).

Peters, A. M. (2001). Filler syllables: what is their status in emerging grammar? *Journal of child language*, 28(1), 229-242.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: ACC.

Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied psycholinguistics*, 10 (3), 335-355.