Morphological Cues vs. Number of Nominals in Learning Verb Types from Child-Directed Speech

Syntactic bootstrapping approaches to verb learning propose that children keep track of the number of nominals surrounding a verb to derive transitive or intransitive verb types (e.g. Naigles & Swensen, 2007). Criticisms of the syntactic bootstrapping theory question the generalizability of the mechanism to languages with nominal ellipsis and varying word order (e.g. Bowerman & Brown, 2007). In responding to these criticisms, Lee and Naigles (2005) analyzed child-directed speech in Mandarin, a language with argument-drop and flexible word order, concluding that the number of noun phrases surrounding a verb provides information for learners to conjecture about verb types. However, though diverging from English regarding the extent of argument ellipsis, Mandarin also mostly marks grammatical relations through word order.

Our study is a machine-learning of verb types using two Turkish corpora containing 12,276 and 20,687 child-directed utterances respectively (1). Turkish, unlike English and Mandarin, (a) is a language that relies on morphology, not so much on word order, to assign grammatical relations, and unlike English, but similarly to Mandarin (b) allows extensive nominal ellipsis. We recorded child-caregiver interactions of two female learners in different socioeconomic backgrounds, one hour every two weeks between 0;9-1;10.

In determining our learning target, we manually tagged each verb that occurred 10 or more times in each corpus according to their argument structure (2). The WEKA machine-learning toolkit was used to systematically evaluate different subsets of linguistic features in utterances and rank them according to their performance in discriminating transitive from intransitive verbs, and within intransitives discriminating unaccusative from unergative verbs. We took the distance between two verbs to be the Euclidean distance between their feature vectors, using the k-means clustering algorithm and compared the resulting cluster assignments with the manually tagged actual categories of the verbs to see how well a particular feature set performs.

Categorizing every verb as transitive (the most frequent—majority class—category) would be right for approximately 60% of the verbs. We found that using the number of nominals as input can predict the transitive-intransitive distinction correctly for around 70% of the verbs, i.e., 10% better than the majority class prediction. Using the frequency of accusatives, we can distinguish transitives from intransitives correctly for about 90% of the verbs, significantly better than the majority class baseline. We, then, explored whether verbal morphology would be useful in distinguishing between different classes of verbs. None of the verbal morphemes served as crucial cues for classifying transitive vs. intransitive verbs; lack of verbal morphology contributed criterial value in subcategorizing intransitives. Using the frequency with which verbs are used in their bare form as in commands, we were able to predict the unaccusative-unergative distinction with 74% accuracy in one corpus, and 83% in the other, significantly above the majority class baseline (60%) (see Figure 1 for clusters of verbs, on the basis of features of accusative and bare verb, i.e., imperative).

We conclude that in classifying verb types, morphological cues provide more reliable cues than the number of nominals types in Turkish, a morphologically rich language.

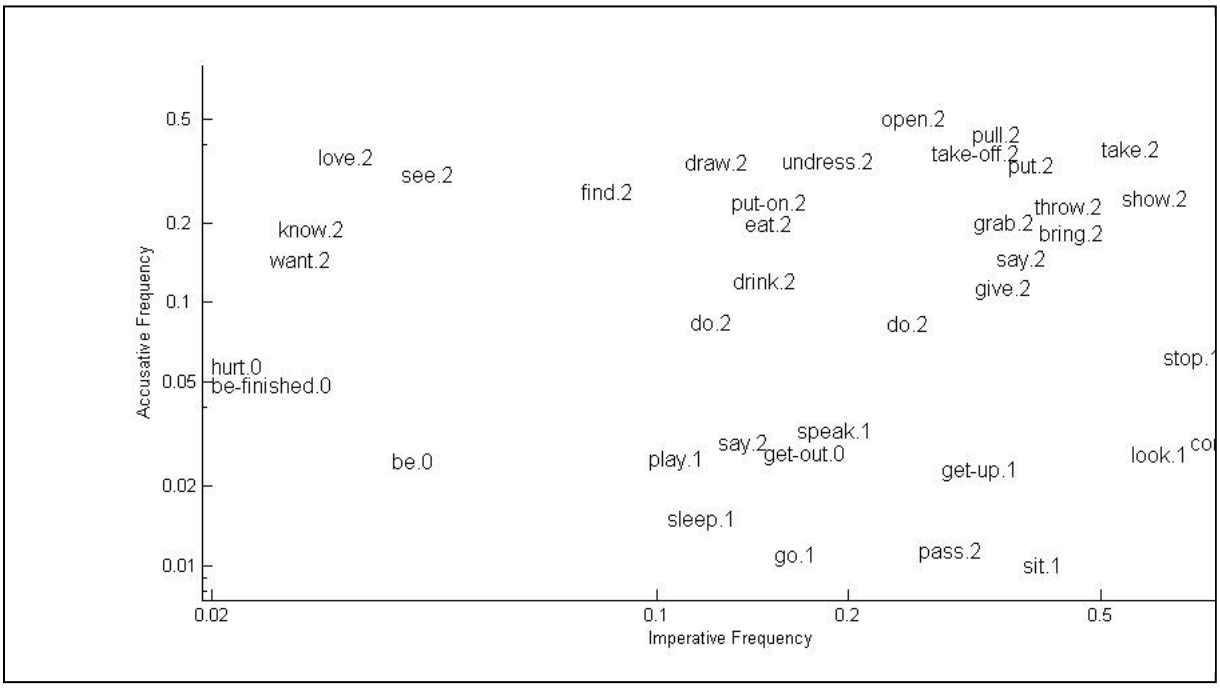
500 words

Examples and Figures

(1) Example from the corpora. Nanny is talking to the child.
NAN-CHI: kuş-lar-ı-mı arı-yo(r)-sun?
%mor: N|kuş-PL-ACC-QUE V|ara-IPVF-2S
'Are you looking for the birds?

(2) Verbs are tagged according to their argument structures as transitive and intransitive and within the intransitive category as unergative and unaccusative. Here are some examples:
a. Transitive (including ditransitives): e.g. *ye-* 'eat' *kır-* 'break' *koy-* 'put'
b. Intransitives: (i) Unergatives. e.g., *koş-* 'run' *ağla-* 'cry'
(ii) Unaccusatives. E.g., *düş-* 'fall' *eri-* 'melt'

Figure 1. English translations of the most frequent verbs shown with the frequency with which they take accusative-marked nominals (Y-axis), and the frequency with which they are used in the bare imperative form (X-axis). A number indicating the argument structure category has been appended to each verb: 0=unaccusative, 1=unergative, 2=transitive.



References

Bowerman, M. and Brown, P.(2007) *Crosslinguistic perspectives on argument structure: Implications for learnability*. New York: Lawrence Erlbaum Assoc.

Lee, J. & Naigles, L.R. (2005) Input to verb Learning in mandarin Chinese: A role for syntactic bootstrapping. *Developmental Psychology 41*, 529-540.

Naigles, L., & Swensen, L. D. (2006). Syntactic supports for word learning. In E. Hoff, & M. Shatz (Eds.), Blackwell Handbook of Language Development. London: Blackwell Publishing.