INPUT COMPLEXITY AND RULE INDUCTION. AN ENTROPY MODEL

Silvia Rădulescu (Utrecht University), Frank Wijnen (Utrecht University) & Sergey Avrutin (Utrecht University)
S.Radulescu@uu.nl

In language acquisition, children manage impressively fast to infer generalized rules from a limited set of linguistic items, and apply those rules to novel strings. This study investigates what triggers and what limits the inductive leap from memorizing specific items to extracting abstract rules that apply productively beyond those items. Our new entropy model predicts that *generalization is a cognitive mechanism that results from the interaction of input complexity (entropy) and brain's limited processing and memory capacity (i.e. limited channel capacity)*.

It was argued that children detect patterns in auditory input, like phonotactic information (Chambers, Onishi & Fisher, 2003), and word boundaries (Saffran, Aslin & Newport, 1996) by statistical learning. Statistical learning deals with computing probabilities that specific items co-occur in the input, and it cannot account for abstractions beyond those items. Previous studies (Gómez & Gerken, 2000) drew a distinction between abstractions based on specific items (e.g. *ba* follows *ba*) and category-based abstractions (generalizing over specific elements, e.g. Noun-Verb constructions). An algebraic system was proposed (Marcus, Vijayan, Rao & Vishton, 1999) to account for extracting rules that apply to categories, such as "the first item is the same as the third item" (*li_na_li*). This system addresses abstractions to novel items, but it does not explain how humans tune into such algebraic rules, and what the factors (if any) in the input are that facilitate or impede this process. Our entropy model addresses these questions and bridges the gap between previous findings, thus unifying them under one consistent account. According to our model, less complexity in the input facilitates memorization of specific items, which allows for abstractions based on those items, while a higher input complexity that overloads the channel capacity drives the tendency to make category-based generalizations (i.e. reduce the number of features that items can be coded for, and group them in abstract categories and acquire relations between these categories).

In our first experiment we exposed adults to 3-syllable AAB strings that implemented a miniature artificial grammar to probe the effect of *input complexity* on rule induction. We manipulated two factors (number of syllables and their frequency) and we used *entropy* (a function of the two factors) as a measure of complexity (calculated in *bits*), to design three experimental conditions: low entropy - 3.5 bits (4×6 As/4×6 Bs), medium entropy – 4 bits (2×12 As/2×12 Bs), and high entropy – 4.58 bits (1×24 As/1×24 Bs). Participants gave grammaticality judgments on 4 types of test strings: grammatical trained AAB strings, grammatical AAB strings with new syllables, ungrammatical new A_1A_2B strings (three different syllables), and ungrammatical A_1A_2B strings with trained syllables. In a second experiment we exposed adults to a similar AAB grammar, but the three conditions had other degrees of *entropy*: 2.8 bits (4×7 As/4×7 Bs), 4.25 bits (2×14 As/2×14 Bs), and 4.8 bits (1×28 As/1×28 Bs). Participants were tested on the same types of test strings as in the first experiment. As predicted, the results of the first experiment showed that the higher the input complexity, the higher the tendency to abstract away from specific items and make a category-based generalization (i.e. accept new AAB strings). The same *effect of input complexity* on rule induction was replicated in the second experiment. When put together, the results from both experiments are in line with the predictions of our model: they show a progressively increasing tendency of generalizing beyond specific items, as the entropy increases (Fig.1). Unlike previous findings, this model also gives a quantitative measure for the likelihood of making generalizations in different ranges of input complexity. To further test our model and its domain generality, similar studies will be run with infants, and also using visual input.
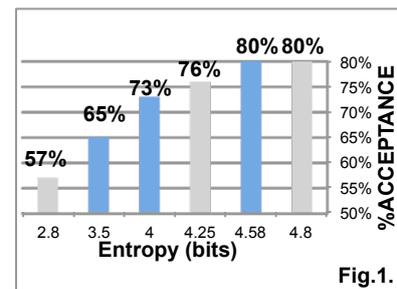

Fig.1.