Probabilistic Approach to Stress Assignment in Arabic

Although previous studies (e.g., Broselow 1976, McCarthy 1990, Watson 2002) have been successful in analyzing the interplay between metrical or syllabic constrains on stress assignment in Arabic, these analyses are predominately rule/constraint-based. This predisposition toward rule/constraint-based analyses does not adequately address one important variable in language – FREQUENCY, as frequency has been found to have a significant effect on various facets of language acquisition and processing (Ellis 2002). To address this inadequacy of discussion on the role of frequency in the literature, this study argues that while the stress system in Arabic is rule/constraint-governed, it could be probabilistic, bottom-up and frequency-informed as well. The purpose of the study is two-fold: a) to provide a statistical description of frequency cues that are relevant to stress assignment, and b) to examine the viability of using these cues either separately or collectively to predict stress in Arabic.

The dataset is constructed from a respected frequency dictionary that is compiled from a 30-million-word corpus (Buckwalter & Parkinson 2010). From the 5000 words included in the dictionary, 4283 multisyllabic words are extracted, corresponding to 105 stress patterns. From these stress patterns, the frequency distribution is calculated for a) the stress pattern, b) the stress position, c) the syllabic structure of the stressed syllable, d) the conditional probability of the stressed syllable and e) the interaction between the number of syllables and aforementioned frequency ques. A brief sample of this analysis is presented in (1) and (2), where "high" indicates that a syllable more is frequently stressed given the corresponding frequency cue, and vice versa.

(1)

	Ante	Penu	Final
Bi-syllabic	n.a	high	low
Tri-syllabic	high	mid	low
Quadri-syllabic	high	mid	low

(2)

	CVC	CVVC	CV	CVV
Bi-syllabic	high	high	High	mid
Tri-syllabic	high	mid	Mid	high
Quadri-syllabic	mid	low	Low	high

Although the predicting power of these frequency cues is rather limited when considered separately, a significantly stronger predicting power emerges when the frequency cues are considered collectively. By applying the Hidden Markov Model, we are able to use the frequency cues to calculate and compare the probability of an attested stress pattern with other hypothetically possible options that are not attested in Arabic, e.g., [ˈCV.CV.CV] v.s *[CV.ˈCV.CV] v.s. *[CV.CV.ˈCV]. The result shows that nearly 97% (101 out of 105) of the attested stress patterns have higher probability than the hypothetically possible ones.

This result shows that the probabilistic model is successful at predicting the stress position and rejecting ill-formed stress patterns in nearly all instances tested. This strong predicting power confirms the viability of a probabilistic approach to stress assignment that is not grounded upon any pre-set rules or constraints but it relies on the interplay of various frequency cues from which rule-like patterns emerge. This result suggests an alternative approach to understanding stress in Arabic, which in turn encourages scholars to reconsider the necessity and validity for rules/constraints-based analyses for stress assignment in particular and for language acquisition in general.

Subfields: Phonology, Corpus Linguistics

Reference

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.

Buckwalter, T., Parkinson, D. B.. (2011). *A frequency dictionary of Arabic: core vocabulary for learners.* London: Routledge.