An attempt at categorizing clitics in an Arabic variety

Sandra Ziagos & Dina El Zarka

Abstract

The process of building a linguistic corpus involves a variety of problems when it comes to annotation. One of the most notorious issues concerns the definition, classification and annotation of clitics (Schultze-Berndt 2006). The proposed solutions range from a general doubt about the need for a separate class of clitics to innovative categorizations (Haspelmath 2015), specifically to deal with the affix-versus-clitic challenge. The present contribution aims at highlighting the problems that arise in this context by providing an insight into the creation of a linguistic corpus for a hitherto undescribed variety of Arabic within the documentation project *The Language of the Arab Minority in Southern Iran (AiS)*, and to discuss concrete strategies for solving them.

Apart from some selective discussions in the (mostly generative) literature, there is as yet no systematic approach in Arabic Language Studies that attempts to distinguish between words, clitics and affixes. Especially regarding the cliticization of function words such as connectives or prepositions or the notoriously problematic case of (indirect) object markers and negation markers, it is noticeable that the term "clitic" is often either avoided altogether or used without being embedded in a theoretical framework. Frequently, the impression arises that "clitic" and "affix" are terminologically interchangeable choices. In the *Encyclopedia of Arabic Language and Linguistics (EALL)*, clitic elements are mentioned in the contributions on Anatolian, Andalusi, Bahraini, Bedouin, and Cairo Arabic. Crucially, the *World Atlas of Language Structures (WALS)* does not include "clitic" as a separate feature. However, it mentions only verbal affixes/clitics for Gulf Arabic and prepositional clitics for Iraqi Arabic and Syrian Arabic. Needless to say that similar facts pertain to other Arabic varieties as well; and more importantly, there are far more clitic candidates in these varieties.

Based on previous categorization approaches in the literature (e.g., Zwicky & Pullum 1983; Cardinaletti & Starke 1999; Aikhenvald 2002), we try to develop a workable annotation system for our corpus of *AiS*. For this purpose, we apply two different approaches: On the one hand, we rank the eligible criteria according to the importance ascribed to them in the literature; on the other hand, we apply a weighting method following Haspelmath (2015), and subsequently evaluate our clitic candidates against the two lists.

At present, we are finishing the process of developing a valid list of criteria, identifying all potential clitic candidates in our corpus including problematic candidates as definite articles or object affixes, and weighting them based on the number of fulfilled criteria. By the time of the conference, we will be able to present the results. The results should allow us, not only to come up with a manageable annotation system for the *AiS* corpus, but hopefully also to offer a basis for the annotation of other Arabic varieties, as the documentation of Arabic varieties according to documentary linguistics standards has recently gained ground.

References

Aikhenvald, A. Y. 2002. Typological parameters for the study of clitics, with special reference to Tariana. In Dixon, R. & Aikhenvald, A.Y. (eds.), *Word. A Cross-linguistic Typology*. CUP, pp. 42-78.

Cardinaletti, A. & Starke, M. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In van Riemsdijk, H. (ed.), *Clitics in the Languages of Europe*. Berlin: Mouton de Gruyter, pp. 145-234.

EALL: Versteegh, K et al. (eds.) 2009. *Encyclopedia of Arabic Language and Linguistics. Set Volumes 1-5*. Leiden: Brill.

Haspelmath, M. 2015. Defining vs. diagnosing linguistic categories: A case study of clitic phenomena. In Błaszczak, J. et al. (eds.), How categorical are categories? Berlin: Mouton de Gruyter, pp. 273-303.

Schultze-Berndt, E. 2004. Linguistic annotation. In Gippert, J. et al. [eds.], *Essentials of Language Documentation*. Berlin: Mouton de Gruyter, pp. 213-251.

WALS: Dryer, M.S. & Haspelmath, M. (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info

Zwicky, A.M. & Pullum, G. K. 1983. Cliticization vs. Inflection: English N'T. *Language*, 59/3, 502-513.