Competition in the aspect-mood domain: The standardization of a diachronic data set of New Persian

Diachronic studies of the verbal categories in New Persian (10th to 20th centuries) have been conducted either with traditional methods of extraction and analysis of the examples (e.g., Lazard 1963; Natel-Khanlari 1986; Ahmadi-Givi 2001) or through limited sampled data (e.g., Lenepveu-Hotz 2012). This contribution aims at filling the gap of corpus-based studies in this field by means of a discussion of a newly designed data set, concentrated on the aspect-mood markers of New Persian. The ultimate goal is to follow the diachronic changes in the functions that are expressed by these markers. The data set includes 77,000 verb tokens, sampled from 55 texts across eleven centuries (5 per century). In each text, 1400 verbs were extracted from two different batches (700 consecutive verbs from each one), and all of these verbs were labelled for TAM categories as well as their morphological structure, clause type, presence of negation, and event type (stative/dynamic). In total, four inflectional markers as well as four major periphrastic constructions coding aspect-mood categories are attested (see Figure 1).
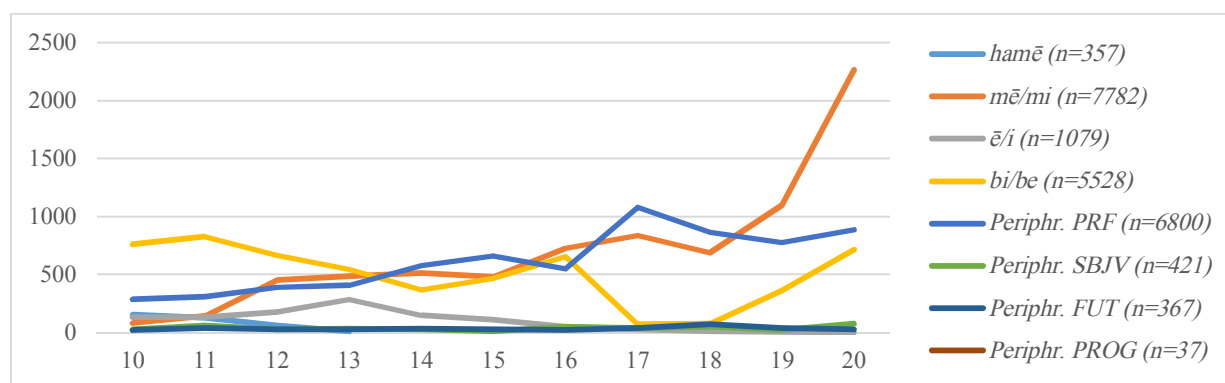

Figure 1. The token frequency of aspect-mood markers in the data set (by century)

While the token frequency is generally genre-based, the proportions of each function expressed by the markers and their changing trends throughout are significant, revealing extensive competition between the markers as well as zero-marking strategies.

In the course of labelling, a number of standardization issues were considered. Firstly, the progressive category was divided into two subcategories (durative and focalized), following Bertinetto et al. (2000), in order to trace the imperfective marker's development more efficiently. Secondly, the future and irrealis were considered as extended interpretations of the general imperfective marker. Thirdly, three stative verbs expressing BE, HAVE, and SHOULD were observed to generally resist aspect-mood inflectional marking, and their exclusion from the analyses showed a more straightforward picture of the generalization of the developing markers. The data set is still being expanded and revised as required by the research topics which are being addressed, but the preliminary results are promising.

References:
Ahmadi-Givi, Hassan. 2001. *Dastur-e tārixi-e fe'l* [The historical grammar of verb]. Tehran: Ghatreh.
Bertinetto, Pier Marco, Karen H. Ebert & Casper de Groot. 2000. The progressive in Europe. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 517–558. Berlin & New York: Mouton de Gruyter.
Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose persan*. Paris: Librairie C. Klincksieck.
Lenepveu-Hotz, Agnès. 2012. *Etude diachronique du système verbal persan (X^e-XVIe siècles): d'un équilibre à l'autre?*. Paris: Ecole Pratique des Hautes Études PhD dissertation.
Natel-Khanlari, Parviz. 1986. *Tārix-e zabān-e Fārsi* [The history of Persian]. Tehran: Ferdows.