Paradigmatic Variation in Hungarian in contexts where syntactic and semantic functions are unstable

Márta Peredy
Research Institute for Linguistics, Hungarian Academy of Sciences

The goal of the talk: Rebrus and Törkenczy (2009) argue that paradigmatic variation occurs at the *unstable points* of the paradigm. Unstable points are those elements of the paradigm which are influenced by more contradictory but equally strong effects. In addition to the morphophonological constraints, they mention two main types of the relevant effects: the one requires the uniformity of forms belonging to the same semantic/syntactic function, the other requires the contrast between forms belonging to different functions. In this framework the semantic and syntactic functions realised by a word form are assumed to be unambiguous. However, in the present work I am going to argue that in a larger syntactic context even the syntactic and semantic functions of a word form may be unstable and therefore may lead to variation. It means that in the context of larger constructions we predict an other kind of pragmatic variation where it is not different allomorphs that vary but different elements from the paradigm. One could say that the variation of semantic and syntactic functions is not a morphological issue anymore but it would only be true in a modular view while the presented data will show that in this kind of variation semantic and syntactic effects strongly interact with morphophonological and frequency effects. In fact, rhythm and rhyme may have some influence, too.

Phenomenon: I will demonstrate my claim with data of variation in Hungarian object agreement. Verbs agree with their objects in definiteness. However, contrary to the traditional generative agreement rule, under special circumstances we find extremely high variation, sometimes the acceptability of an "ungrammatical" sentence exceeds the acceptability of its "grammatical" counterpart. The probability of the lack of object agreement is the highest if the agreeing argument is either indefinite but specific or definite but interpreted as generic, i.e., when syntactic and semantic definiteness do not coincide. In a reading test, where 37 subjects judged 100 simple sentences, the following factors turned out to increase the instability of the syntactic/semantic definiteness function of the verb: (i) the definiteness of the subject and the object is different, (ii) the thematic role of the subject and the object is near to each other, (iii) the role of the two arguments is interchangeable, (iv) both arguments precede the verb, (v) the object precedes the subject.

Preliminary results: On the basis of the reading test, the following purely formal factors strongly influence the variation in unstable contexts characterized above.

1. The similarity of the competing morphemes: The variation, i.e., the acceptance of the non-agreeing verb forms is higher if two overt morphemes vary with each other (this is the case in 3PL, e.g., *bántanak* 'annoy-DEF.3PL' vs. *bántják* 'annoy-INDEF.3PL') than if one of the agreement morphemes is zero (this is the case in 3SG, e.g., *bántja* 'annoy-DEF.3SG' vs. *bántØ* 'annoy-INDEF.3SG'). Compare the acceptance of (1b) and (2b)! (Percentages show the acceptability rates of each sentences. All differences are significant at p=0.05.) Moreover what counts in reality is the difference of the surface forms. Even if the indefinite 3SG agreement morpheme is a zero morpheme (speaking in terms of underlying representations) but there is still a varying ending in the surface forms, which is the case with epenthetic stems like *érdekel* 'interest', then we get the same effect, see (3). This shows that it is only the surface forms that influence this phenomenon and not the underlying ones.

2. Frequencies of the two competing verb forms: The acceptability rate of (3b) is significantly higher than that of (1b). One reason for this (in addition to the semantic factors not discussed here) is the relative frequency differences between the two verbs *érdekel* 'interest' and *bánt* 'annoy'. The relative frequency of the indefinite *érdekel* to the definite

érdekli form is 4.5 while this ratio is only 0.8 for *bánt*. (Frequencies come from the Hungarian National Corpus.)

3. Rhythm and rhyme: Some data suggests that rhythm and rhyme could play a role, too. In (4) and (5), both the subject and the object are indefinite, but still the acceptance of the definite verb form in (4) is still 20%, significantly higher than the 5% in (5). My assumption is that the reason for this kind of difference lies in the fact that the preverbal argument in (4) rhymes better with the ungrammatical ending of the verb than in (5). The comparison of (3b) and (6) is a further example which suggest that the reason why the acceptance of (6) is much higher can be that the preverbal argument rhymes with the ungrammatical verb ending while in (3b) it does not.

Further results: In the present work I will complete the results of this test with another one where on the one hand, the three above mentioned factors will be systematically controlled and on the other hand, complex sentences containing object relative clauses will be tested as well, since these are very good candidates for syntactically and semantically unstable contexts.

Model: I analyse the data in a non-modular stochastic framework. I use the Maximum Entropy Model (Goldwater and Johnson 2003) and for the simulations the Stochastic Gradient Ascent online learning algorithm (suggested by Jäger 2003), with the aid of which the relevance of different factors of different modules can be compared and their weights can be calculated on the basis of the measured acceptability rates.

Conclusion: I attempted to show that in contexts where the syntactic and semantic function of a word is unstable, the role of purely formal effects suddenly increase and may give rise to variation.

Examples

(1a) Ezeket a kislányokat egyes idősebb fiúk bánt**ják**. 89%
    these-ACC the girls-ACC certain older boys annoy-**DEF.3PL**

(1b) Ezeket a kislányokat egyes idősebb fiúk bánt**anak**. 17%
    these-ACC the girls-ACC certain older boys annoy-**INDEF.3PL**

(2a) Ezt a kislányt minden idősebb fiú bánt**ja**. 100%
    this-ACC the girl-ACC every older boy annoy-**DEF.3SG**

(2b) Ezt a kislányt minden idősebb fiú bánt. 2%
    this-ACC the girl-ACC every older boy annoy-**INDEF.3SG**

(3a) Jánost mindenki érdek**li**. 95%
    John-ACC everybody interest-DEF.3SG

(3b) Jánost mindenki érdek**el**. 30%
    John-ACC everybody interest-INDEF.3SG

(4) Mindenkit minden**ki** érdek**li**. 20%
    everybody-ACC everybody interest-DEF.3SG

(5) Mindenki minden**kit** érdek**li**. 5%
    everybody-ACC everybody interest-DEF.3SG

(6) Jánost min**den** érde**kel**. 50%
    John-ACC everything interest-INDEF.3SG

Reference

Goldwater, S. and Johnson, M. 2003 *Learning OT constraint rankings using a maximum entropy model*, in J. Spenader, A. Eriksson, and Ö. Dahl (eds.) Proceedings of the Stockholm Workshop on Variation within Optimality Theory. Stockholm University, 111–120.

Jäger, G. 2003 *Maximum Entropy Models and Stochastic Optimality Theory.* Manuscript, University of Potsdam. ROA 625-1003.

Rebrus, P. and M.Törkenczy. 2009. Paradigmatic Variation in Hungarian. 9th International Conference on the Structure of Hungarian (ICSH9), Debrecen, Hungary