A stochastic optimality theoretic account for the variation in Hungarian object agreement

0. Introduction: In some cases, see examples (2-6) below, the grammaticality judgements about object agreement in Hungarian are very uncertain. The variation is not dialectal or register dependent but is present within one idiolect. Moreover, contrary to the usual cases of free variation, it is not both variants (the agreeing and the non-agreeing forms) that are fully grammatical, but rather neither of them. The main goal of this talk is on the one hand to explore the wide variety of effects that influence the phenomenon, on the other hand to present a stochastic optimality theoretic analysis which handles optionality and probabilistic effects as an inherent property of natural languages. This set of Hungarian data will turn out to have theoretical importance by providing examples for both *strong ganging-up cumulativity* and *counting cumulativity* effects which maximum entropy (ME) models (Goldwater and Johnson 2003, Jäger and Rosenbach 2005) can account for but neither standard OT models nor Boersma's stochastic OT model (StOT) (Boersma and Hayes 2001) can. The general idea behind the present analysis is to let morphological, syntactic, semantic and information structural constraints interact, instead of separating the evaluations to different modules.

1. Data: According to the object agreement rule in Hungarian, verbs agree in definiteness with their third person objects[1], cf. (1a) vs. (1b). Contrary to this rule, (2-6) show that the non-agreeing variants in the (b) examples are also acceptable for some speaker, moreover the non-agreeing form may perform as well as or even better than the agreeing one in (a), cf. (5) and (6) respectively. Acceptability rates given for each sentences in percentage are yielded from a small sample of 6 informants and mean the probability of choosing the definiteness value (+/-) of the agreement morpheme given in the sentence[2].

(1) a. *A macskákat egyes lakók sétáltatják.* 100%
the cats-acc some tenants walk-3pl-def
b. *A macskákat egyes lakók sétáltatnak.* 0%
the cats-acc some tenants walk-3pl-indef

(2) a. *A betegeket jó orvosok kezelik.* 92%
the patients-acc good doctors treat-3pl-def
b. *A betegeket jó orvosok kezelnek.* 8%
the patients-acc good doctors treat-3pl-indef

(3) a. *A diákokat egyes tantárgyak érdeklik.* 75%
the students-acc some subjects interested-3pl-def
b. *A diákokat egyes tantárgyak érdekelnek.* 25%
the students-acc some subjects interested-3pl-indef

(4) a. *A lányokat bizonyos fiúk verik.* 58%
the girls-acc certain boys beat-3pl-def
b. *A lányokat bizonyos fiúk vernek.* 42%
the girls-acc certain boys beat-3pl-indef

(5) a. *A nőket bizonyos ruhák öregítik.* 50%
the women-acc certain dresses make.older-3pl-def
b. *A nőket bizonyos ruhák öregítenek.* 50%
the women-acc certain dresses make.older-3pl-indef

(6) a. *A férfiakat egyes nők taszítják.* 42%
the men-acc certain women not.attract-3pl-def
b. *A férfiakat egyes nők taszítanak.* 58%
the men-acc certain women not.attract-3pl-indef

The examples are construed so that the acceptability of non-agreeing forms is high. The main factors that are most likely to contribute to the effect and will be included in the OT analysis are the following.

(i) Information structure: the topicalisation of the object, cf. (7, 4b), the focalization of the subject, cf. (1, 2).

(7) *Vernek bizonyos fiúk a lányokat.* 0%
beat-3pl-def certain boys the girls-acc
(4) b. *A lányokat bizonyos fiúk vernek.* 42%
the girls-acc certain boys beat-3pl-indef

(ii) Thematic roles. E.g. (1): agent subject, theme object; (6): experiencer object, source subject; or:

(8) *A szüleiket bizonyos gyerekek öregítenek.* 25%
the parents-acc certain children make.older-3pl-def
(5) b. *A nőket bizonyos ruhák öregítenek.* 50%
the women-acc certain dresses make.older-3pl-indef

(iii) The interchangeability of the subject and the object in the sense that the meaning of the sentence remains plausible (e.g. (4,6), but not (1-3,5)).

---

[1] It is well known that strictly speaking it is not the definiteness feature of the object that triggers the agreement, but the form (syntactic type) of the object phrase is held to determine the agreement (except one dialect where the agreement on the verb indicates the specificity of the possessive indefinite, e.g., *Olvastuk/Olvastunk néhány versedet* 'read-Past-we.Def / read-Past-weIndef some poem-Poss2Sg-Acc'), cf. Bartos (2000).

[2] If an informant accepted or rejected both variants, his judgment was evaluated as 50-50%, if he accepted only one form, it was 100-0% or 0-100%. The reason for this kind of evaluation was the idea that the speakers have to choose the one form or the other when they utter these sentences even if they "like" neither of them.

(iv) The bigger is the distance between the agreeing argument and the verb in the sentence the higher is the probability of the non-agreeing verb form:

(9) a. *Jánost mindenki érdekel.* 42% b. *Mindenki Jánost érdekel.* 25%
John-acc everybody interest everybody John-acc interest

(v) The existence of an unaccusative or an adjectival counterpart of the transitive verb (e.g. (5): *öregít* 'make look older' – *öregszik* 'grow older', *öreg* 'old') raises the probability of the non-agreeing form.

(vi) The uncertainty is higher if the definite and indefinite verb forms differ in an overt morpheme, as in the case of third person plural subject (e.g. *ver**nek*** 'beat-Pres-3PlIndef'/ *ver**ik*** 'beat-Pres-3PlDef'; *vert**ek*** 'beat-Past-3PlIndef'/ *vert**ék*** 'beat-Past-3PlDef'), than if they only differ in the presence or the absence of a morpheme, as in the case of third person singular subject (e.g. *ver**Ø*** 'beat-Pres-3SgIndef'/ *ver**i*** 'beat-Pres-3SgDef'; *vert**Ø*** 'beat-Past-3SgIndef'/ *vert**e*** 'beat-Past-3SgDef').

(vii) The type of the determiner of the object and the subject.

2. Problem: As far as I know there is no Hungarian grammar claiming that the object agreement is optional or driven by other factors than the form of the object. A model with the following properties is needed. 1. Violable constraints: Object agreement in Hungarian is not a strict, but rather a violable constraint. 2. Optionality: There isn't one optimal output, but the speakers choose, with some probability, the one candidate or the other. 3. Weighted constraints: Although the factors summarized in the previous section seem to be plausible their relative importance, i.e., their weights, have to be determined quantitatively.

3. Model: Standard Optimality Theory (Prince and Smolensky 1993) works with violable constraints in strict domination, and selects one optimal candidate. In contrast, the maximum entropy model of stochastic OT characterizes the constraints with positive weights and the candidates with their probabilities (instead of selecting one winner). Optionality corresponds to cases when more then one candidate has relatively high probability. I will use a constraint set generated by the combination of the examined parameter (the agreement morpheme on the verb, ±D) and the binary factors mentioned above (e.g.: ±DOBJ: definiteness of the object, ±DTH: definiteness of the theme, ±UNACC: the existence of an unaccusative pair of the verb, ±TOPOBJ: the topichood of the object, ±MORPH: the morphological relation of the agreement morpheme and its alternative, etc.). E.g., -D/+DOBJ requires the cooccurrence of the indefinite agreement morpheme on the verb and a definite object. I will use Jäger (2003)'s *Stochastic Graduate Algorithm* with the aid of which the weights can be determined on the basis of a set of training data. The training data will be obtained from 10 informants, every parameter combination examined will be represented by at least two sentences. Given the weights, the probabilities of candidates predicted by the model can be determined and compared with the test values.

4. Results: **a.** Topicality (i) and thematic roles (ii) are well known to determine the unmarked mapping of semantic arguments to grammatical functions (Aissen 1999), and it is not surprising that the acceptability of non-agreeing objects correlates with the marked mappings. The interchangeability of the two arguments (iii) is also related to the argument structure, i.e., the selectional criteria of the verb.

b. According to my analysis, in Hungarian, where there is no passivisation, the presence of an unaccusative or adjectival form (v) also turns out to be relevant. This is, however, an analogical phenomenon, the weight of which, compared to "pure" syntactic and semantic factors, can be measured in this model. The fact that to omit the definiteness marker (when it alternates with the zero morpheme) is more difficult than just to replace it with another morpheme marking indefiniteness (vi) also can be considered as a loose formal analogy.

c. The studied agreement phenomenon provides example for cumulativity effects the importance of which lies in the fact that standard OT cannot account for them. On the one hand, two or more weaker constraints (the factors above) can jointly "beat" a stronger one (the object agreement constraint). This is called ganging up cumulativity. On the other hand, a weaker constraint assigning more violation to a candidate (e.g., the constraint which is sensitive to the distance of the verb and the argument triggering the agreement (iv)) can "beat" a stronger constraint assigning just one violation to another candidate.

References:

Aissen, J. 1999 *Markedness and subject choice in optimality theory*. Natural Language and Linguistic Theory 17(4): 673-711.

Bartos, H. 2000 *Az inflexiós jelenségek szintaktikai háttere*. Strukturális Magyar Nyelvtan 3. Morfológia. Akadémiai Kiadó

Boersma, P. and Hayes, B. 2001 *Empirical tests of the Gradual Learning Algorithm*. Linguistic Inquiry 32(1): 45-86.

Goldwater, S. and Johnson, M. 2003 *Learning OT constraint rankings using a maximum entropy model*. In: Jennifer Spenader, Anders Eriksson, and Östen Dahl (eds.) Proceedings of the Stockholm Workshop on Variation within Optimality Theory, 111-120. Stockholm University.

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

Jäger, G. and Rosenbach, A. 2005 *The winner takes it all – almost. Cumulativity in grammatical variation*. In: Uli Sauerland and Penka Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*, Palgrave McMillan

Prince, A. and P. Smolensky. 1993 *Optimality Theory*. Ms., Rutgers University., 136 o.