

# Variation in Hungarian definiteness agreement

Peredy Márta  
ICSH9, 30<sup>th</sup> August 2009

(1a) *A nyelvészeket a nyelvi jelenségek érdeklik.*  
the linguists-ACC the linguistic phenomena are.interested.in-  
-DEF.3PL

(1b) *Bizonyos nyelvészeket a nyelvi jelenségek érdekelnek.*  
certain linguists-ACC the linguistic phenomena are.interested.in-  
-INDEF.3PL

(1c) *Bizonyos nyelvészeket a nyelvi jelenségek érdeklik.*  
certain linguists-ACC the linguistic phenomena are.interested.in-  
-DEF.3PL

Red means definiteness,  
Green means indefiniteness,  
*Italics mean objecthood.*

# 0. Goals

1. There is this phenomenon.
2. This variation cannot be accounted for in the generative syntactic framework suggested by Bartos (2000).
3. This variation can be modelled by the Maximum Entropy model of stochastic OT.
4. The variation challenges the competence/performance duality.  
*(Ok, this is performance, but what is the competence behind it?)*

## 0. Goals

1. Syntactic Object Agreement Rule

2. Test, data

3. Syntactic and semantic factors which influence  
the phenomenon

4. Stochastic OT – Maximum Entropy Model

5. Experiments

6. Results, ideas

7. More and even stranger factors

0. Goals

## 1. Syntactic Object Agreement Rule

2. Test, data

3. Syntactic and semantic factors which influence  
the phenomenon

4. Stochastic OT – Maximum Entropy Model

5. Experiments

6. Results, ideas

7. More and even stranger factors

# 1. Syntactic Object Agreement Rule

Bartos (2000):

The verbal agreement is definite iff the object is a DP.

Exceptions that can be accounted for:

a. 1<sup>st</sup>, 2<sup>nd</sup> person personal pronouns

b. dialectal: possessive construction object with a dative possessor

(2a) Jónás Tamásnak tanultunk *egy versét*.

- non-specific: in the school we learned about JT

(2b) Jónás Tamásnak tanultuk *egy versét*.

- specific: we learned that poem by heart

# 1. Syntactic Object Agreement Rule

Exceptions that can be accounted for:

c. wh-movement from embedded clause

- (3a) *Kit* szeretné<sub>l</sub>, hogy megverjék?      Bartos (2000) (164)  
(3b) ?*Kit* szeretné<sub>l</sub>, hogy megverjem?      (165a-c)  
(3c) ?*Kit* szeretné<sub>d</sub>, hogy megverjék?  
(3d) \**Kit* szeretné<sub>d</sub>, hogy megverjem?

d. Kálmán (2009): *minden* 'every'+ordinal number objects.  
definite/indefinite verb forms ~ 50-50%

- (4a) *Minden tizedik* műtétet hazánkban végeznek/végzik.  
(4b) Vigyünk/Vigyünk ki *minden tizedik* székét!

# 1. Syntactic Object Agreement Rule

Exceptions that cannot be accounted for:

“*Mindenkit*, aki épp nem nyaral és van kedve egy nyáresti koncerthez, régi és új versekkel, szeretettel várjuk!”

“Nem tudtam, *másokat* mennyire érintette mélyen *az*, hogy gyerekkorában milyen mesék, könyvek, képregények, filmek vették körül.”

“A dokk.hu fenntartja magának a jogot, hogy *az oldalra nem illő, célkitűzéseinek nem megfelelő profilt és anyagot* önkényesen, előzetes bejelentés nélkül töröljön.”



# 1. Syntactic Object Agreement Rule

Exceptions that cannot be accounted for:

$P_{abs}$

15 *A férfiakat* bizonyos viselkedésformák taszítják. 97%

16 *A férfiakat* bizonyos viselkedésformák taszítanak. 53%

25 *Az ás fiúkat* egyes bés fiúk verik. 92%

26 *Az ás fiúkat* egyes bés fiúk vernek. 27%

! 65 *Egyes matematikusokat* a számelméleti problémák foglalkoztatják. 88%

66 *Egyes matematikusokat* a számelméleti problémák foglalkoztatnak. 40%

73 *Egyes nőket* a sötét ruhák öregítik. 85%

74 *Egyes nőket* a sötét ruhák öregítenek. 48%

0. Goals

1. Syntactic Object Agreement Rule

**2. Test, data**

3. Syntactic and semantic factors which influence  
the phenomenon

4. Stochastic OT – Maximum Entropy Model

5. Experiments

6. Results, ideas

7. More and even stranger factors

## 2. Test

- 50 sentence pairs, which differ only in the definiteness feature of the verb.
- 37 subjects
- reading test
- sentences in random order (not in pairs, every subject got a different order)
- judgements: 0 – bad sentence, 1 – strange sentence, 2 – good sentence
- the subjects were asked to read out the sentences loudly and answer immediately.

## 2. Test

### Acceptability rates

- **Absolute acceptability rates ( $p_{abs}$ )** calculated for each sentence, the average of the judgements of that sentence.

- **Relative acceptability rates ( $p_{rel}$ )** calculated for each sentence pair.

$$p_{rel}(+def) + p_{rel}(-def) = 100\%$$

$$p_{rel}(-def) / p_{rel}(+def) = p_{abs}(-def) / p_{abs}(+def)$$

## 2. Test

- **Information structure** of the sentences:

Usually the verb (or the verbal modifier, if any) is the information focus. Exception: sentences with the universal quantifier.

- **The naturalness of the sentences** obviously influences the judgements. → relative acceptability rates compensate this effect at least partly.

Data

0. Goals

1. Syntactic Object Agreement Rule

2. Test, data

**3. Syntactic and semantic factors which influence  
the phenomenon**

4. Stochastic OT – Maximum Entropy Model

5. Experiments

6. Results, ideas

7. More and even stranger factors

### 3. Syntactic and semantic factors which influence the phenomenon

#### 1. definiteness of the grammatical object and subject

	$p_{rel}$
41 <i>A nőt</i> bizonyos ruhák öregítik.	69%
42 <i>A nőt</i> bizonyos ruhák öregítenek.	31%
73 <i>Egyes nőt</i> a sötét ruhák öregítik.	61%
74 <i>Egyes nőt</i> a sötét ruhák öregítenek.	39%

#### 2. the distance between the “(non)agreeing” NPs and the verb

	dist(O, V)	
75 <i>Jánost</i> mindenki érdekli.	2	76%
76 <i>Jánost</i> mindenki érdekeli.		24%
79 Mindenki <i>Jánost</i> érdekli.	1	82%
80 Mindenki <i>Jánost</i> érdekeli.		18%

### 3. Syntactic and semantic factors which influence the phenomenon

#### 3. word order

33 <i>A lányokat</i> egyes fiúk dühítik.	65%
34 <i>A lányokat</i> egyes fiúk dühítenek.	35%
99 Dühítik egyes fiúk <i>a lányokat</i> .	93%
100 Dühítenek egyes fiúk <i>a lányokat</i> .	7%

#### Semantic factors (Group I)

#### 4. thematic roles

{AG, PAT} or {THEME, EXP}

correlation	PAT/EXP
$p_{abs}$ („agree”)	0,01
$p_{abs}$ („non-agree”)	-0,66
$p_{rel}$	0,64



### 3. Syntactic and semantic factors which influence the phenomenon

5. intransitive equivalent  
(unaccusative verb or adjective)  
*öregít* 'make look older' –  
*öregszik* 'grow older';  
*megnevettet* 'make laugh' –  
*nevet* 'laugh'

correlation	intr
$p_{abs}$ („agree”)	-0,32
$p_{abs}$ („non-agree”)	0,35
$p_{rel}$	-0,4

6. the interchangeability of the subject and the object in the sense that the meaning of the sentence remains plausible.

correlation	ich
$p_{abs}$ („agree”)	-0,03
$p_{abs}$ („non-agree”)	0,35
$p_{rel}$	-0,37

- 5 *A gyerekeiket egyes szülők verik.* **93%**
- 6 *A gyerekeiket egyes szülők vernek.* **7%**
- 25 *Az ás fiúkat egyes bés fiúk verik.* **75%**
- 26 *Az ás fiúkat egyes bés fiúk vernek.* **25%**

### 3. Syntactic and semantic factors which influence the phenomenon

6. type of the determiners

7. phonological realization

The uncertainty is higher if the two verb forms differ in an overt morpheme (*bánt**anak** / bánt**ják***) than if they only differ in the presence or absence of a morpheme (*bánt $\emptyset$  / bánt**ja***).

29 *Ezeket a kislányokat* egyes idősebb fiúk bántják. 85%

57 *Ezt a kislányt* minden idősebb fiú bántja. 99%

8. information structure

9. relative frequencies

10. genericity

11. etc.

0. Goals

1. Syntactic Object Agreement Rule

2. Test, data

3. Syntactic and semantic factors which influence  
the phenomenon

**4. Stochastic OT – Maximum Entropy Model**

5. Experiments

6. Results, ideas

7. More and even stranger factors

## 4. Stochastic OT – Maximum Entropy Model

A model with the following properties is needed:

- **Violable constraints:** Object agreement in Hungarian is not a strict, but rather a violable constraint. → OT
- **Weighted constraints:** The relevance of the different factors has to be determined quantitatively. → ranks
- **Variation:** There isn't one optimal candidate (winner), but the speakers choose, with some probability, the one candidate or the other. → stochastic model

OT model is good method to estimate the weights of the different factors.

## 4. Stochastic OT – Maximum Entropy Model

### Maximum Entropy Model (Goldwater and Johnson 2003)

- characterizes the constraints with ranks
- and the candidates with their probabilities
- all candidates have a positive probability

$$p(cand_k) = \frac{1}{Z} \cdot \exp\left(-\sum_j r_j c_j(cand_k, input)\right)$$

- parallelism with the Boltzmann-factor

$$E = \sum_j r_j c_j(cand_k, input)$$

## 4. Stochastic OT – Maximum Entropy Model

### Simulation

- The constraints represent the factors mentioned above.
- In every competition there are two candidates which differ in the definiteness morpheme of the verb.
- On the bases of probabilities (relative acceptability rates) derived from the data, the model calculates the constraint ranks (the importance of the different factors).

### The learning algorithm

**SGA:** Stochastic Gradient Ascent for MaxEnt (Jäger 2003)

$$r_i^{n+1} = r_i^n + \eta \cdot (c_i(h^n, input^n) - c_i(datum^n, input^n))$$

## 4. Stochastic OT – Maximum Entropy Model

### The set of constraints

*Harmonic Alignment* method (Prince and Smolensky 2004): Constraints are defined by the possible combinations of two parameter sets.

Set 1: +/-D – the definiteness of the verbal morpheme

Set 2: the factors which may influence it

+/-do – the definiteness of the object,

+/-ds – the definiteness of the theme subject,

+/- darg2 – the definiteness of the second NP in the sentence

+/-intr – the existence of an intransitive equivalent of the verb,

+/-ich – the interchangeability of the role of the subject and the object,

+/-pat – the thematic grid is {AG, PAT} or {THEME, EXP}.

Constraints: e.g.:

+D/+do – requires a definite verb in the case of a definite object,

-D/-intr – requires a indefinite verb if the verb does not have an intransitive counterpart.

Distance: is expressed by the number of violations.

0. Goals

1. Syntactic Object Agreement Rule

2. Test, data

3. Syntactic and semantic factors which influence  
the phenomenon

4. Stochastic OT – Maximum Entropy Model

**5. Experiments**

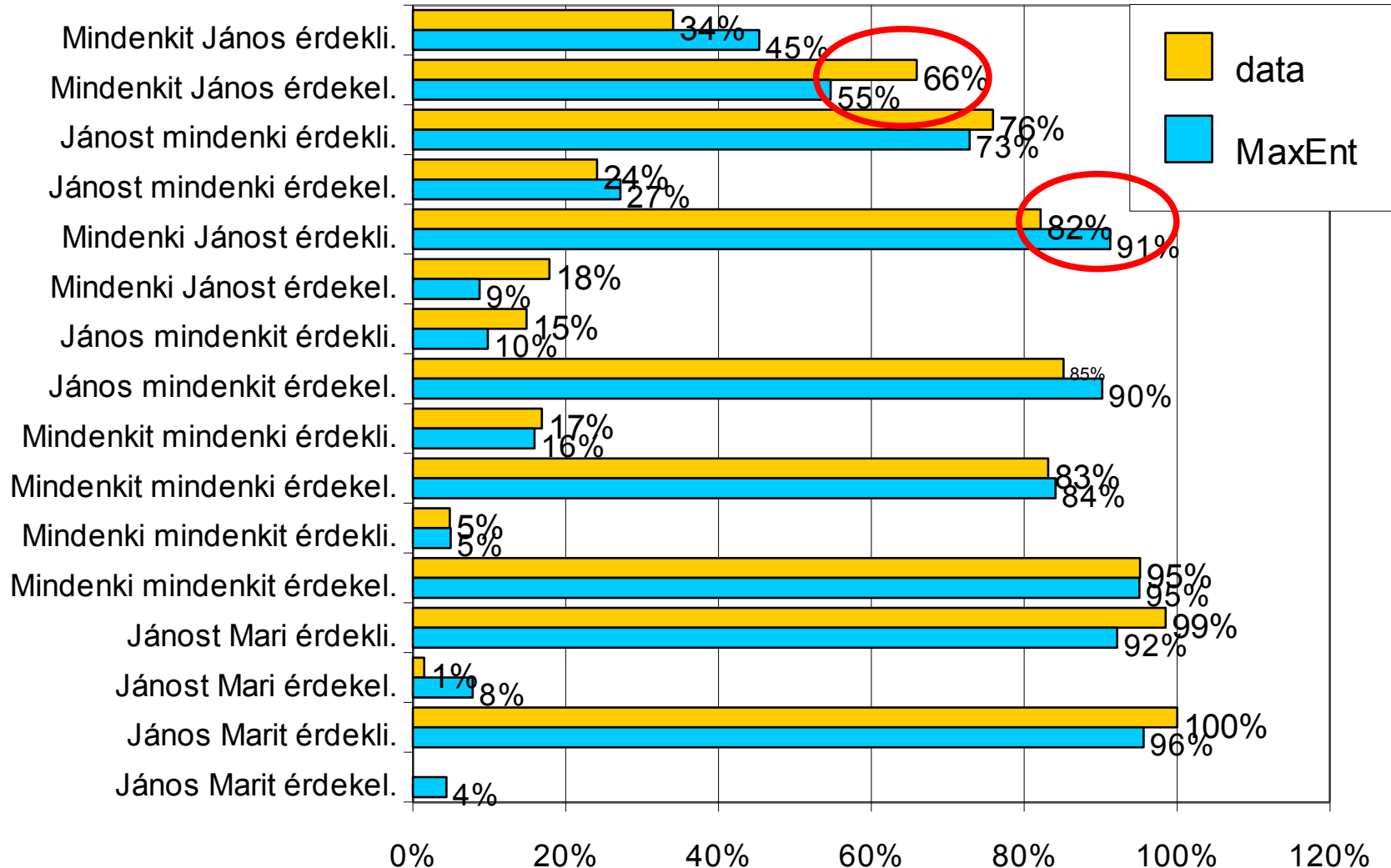
6. Results, ideas

7. More and even stranger factors



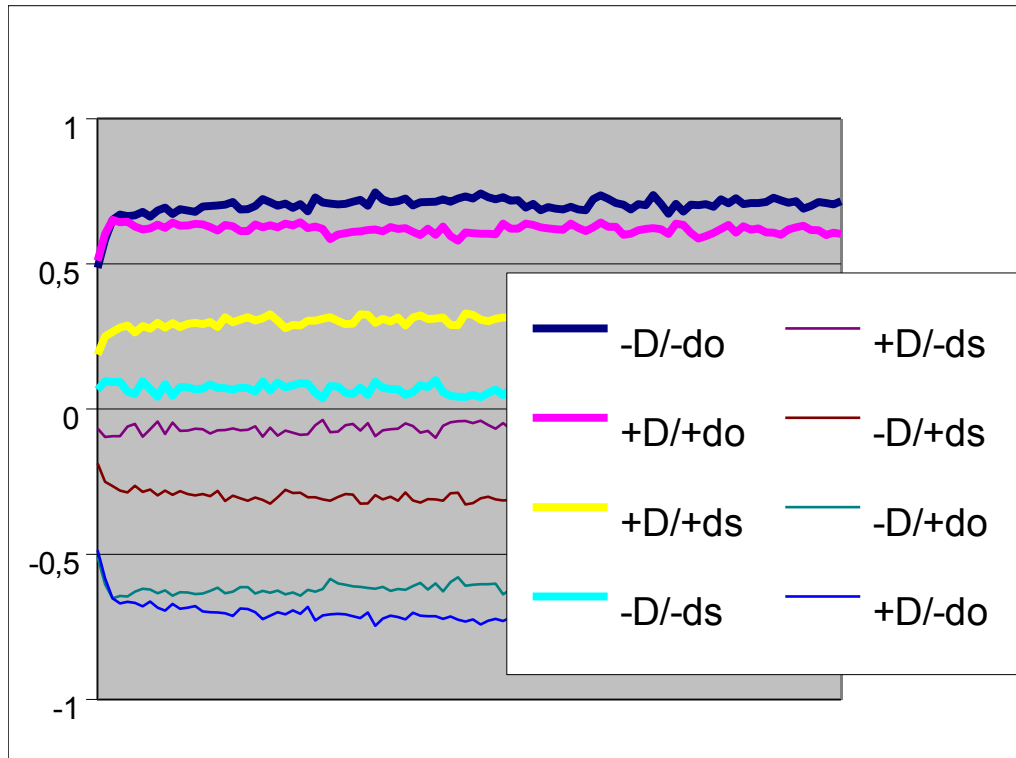
# 5. Experiments

## Experiment 1: Group V. - acceptability rates



## 5. Experiments

### Experiment 1: Constraints and their ranks



$r(-D/-do)=0,72$   
 $r(+D/+do)=0,6$   
 **$r(+D/+ds)=0,31$**   
 $r(-D/-ds)=0,06$

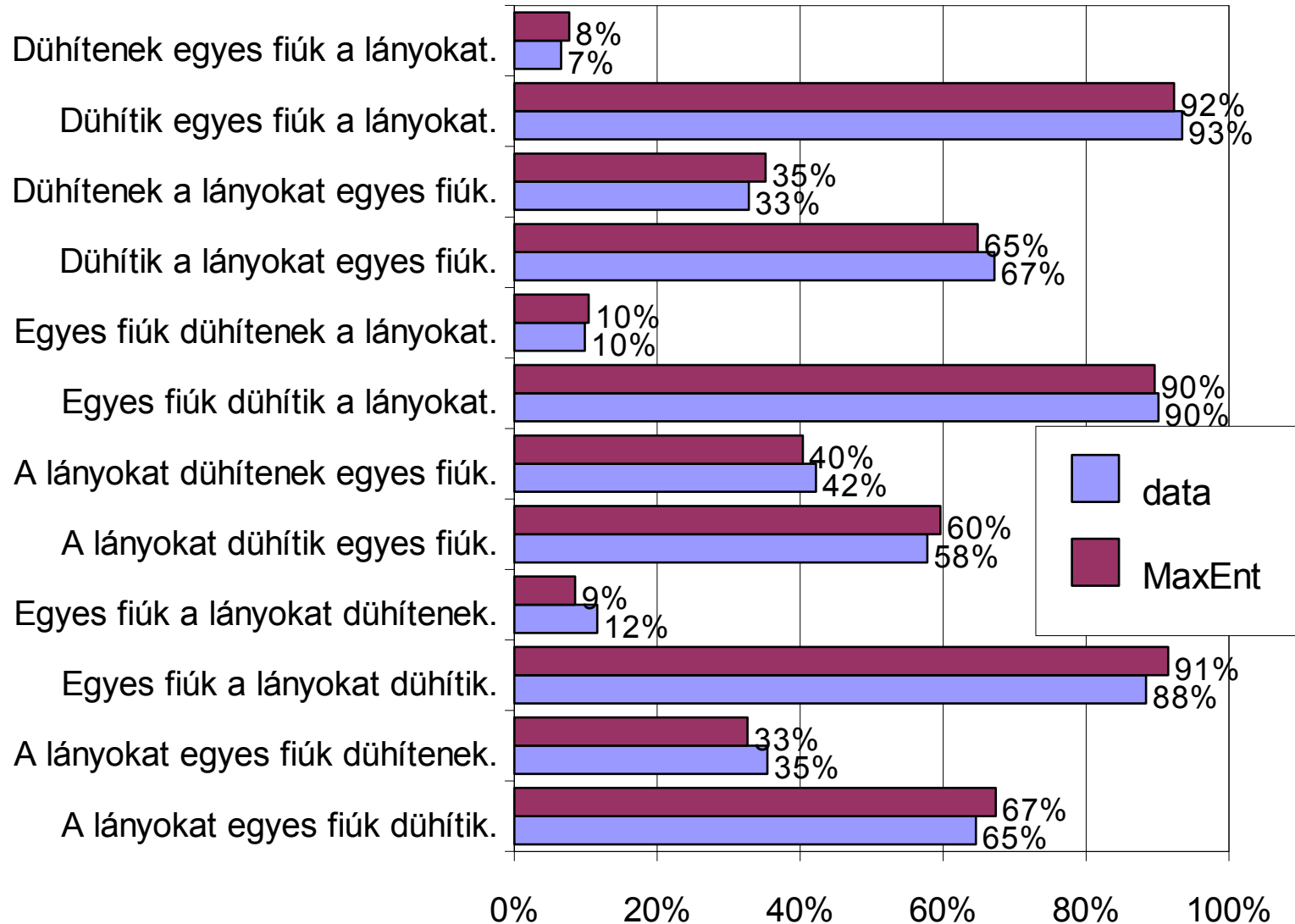
## 5. Experiments

### Experiment 1 - Predictions

	data	MaxEnt
1.		
75 <i>Jánost</i> mindenki érdekli.	76%	73%
79 Mindenki <i>Jánost</i> érdekli.	82%	91%
33 <i>A lányokat</i> egyes fiúk dühítik.	65%	
91 Egyes fiúk <i>a lányokat</i> dühítik.	88%	
2.		
75 <i>Jánost</i> mindenki érdekli.	76%	73%
82 <i>Mindenkit</i> János érdekeli.	66%	55%
Group I +defO -defS +defV	75%	
Group IV -defO +defS -defV	43%	

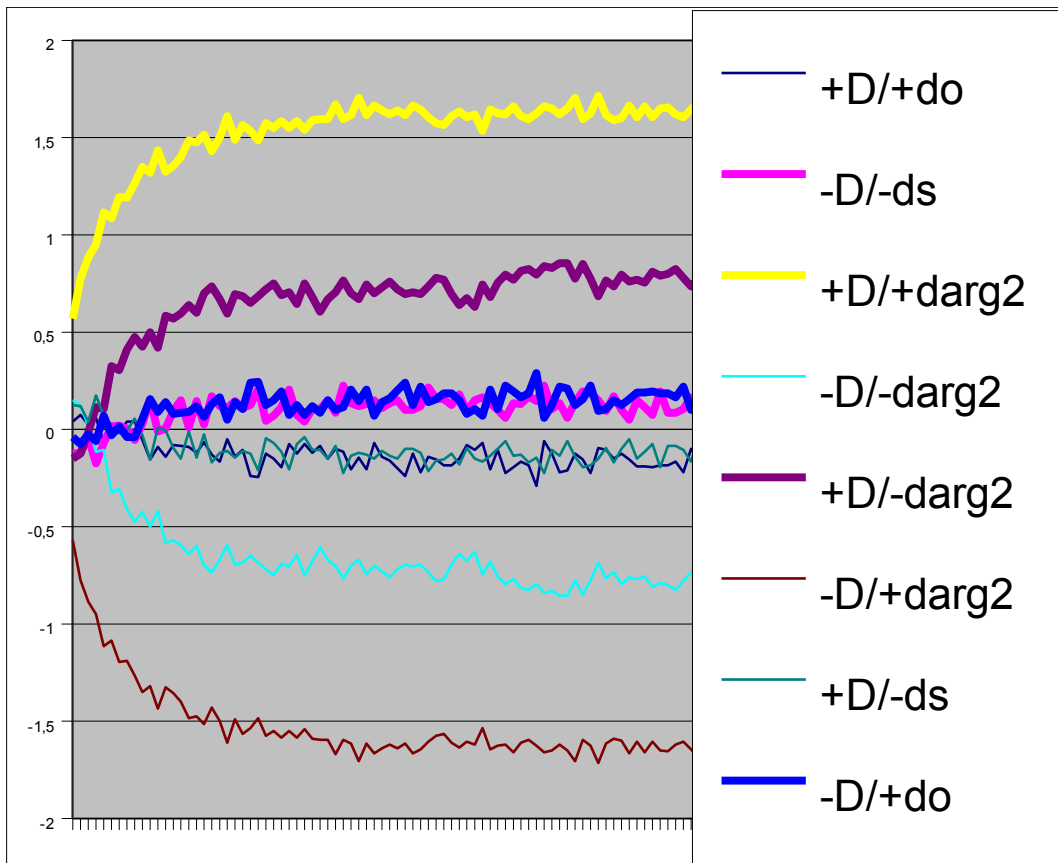
## 5. Experiments

### Experiment 2: Group VI. - acceptability rates



## 5. Experiments

### Experiment 2: Constraints and their ranks



$r(+D/+darg2)=1,6$   
 $r(+D/-darg2)=0,79$   
 $r(-D/+do)=0,16$   
 $r(-D/-ds)=0,13$

## 0. Goals

1. Syntactic Object Agreement Rule

2. Test, data

3. Syntactic and semantic factors which influence  
the phenomenon

4. Stochastic OT – Maximum Entropy Model

5. Experiments

**6. Results, ideas**

7. More and even stranger factors

## 6. Results, ideas

1. Exp. 1.: The definiteness of the subject plays an important role.
2. Exp. 2.: The SO order is preferred even postverbally.
3. The definiteness marking of the verb can be accounted for without the notion of agreement.
4. Contradictory factors → unsable points → variation  
(Rebrus 2009)
5. Instead of specificity difference, difference in theta grid?  
  
(2a) Jónás Tamásnak tanultunk egy versét.  
- in the school we learned about JT → {EXP, TH}  
  
(2b) Jónás Tamásnak tanultuk egy versét.  
- we learned that poem by heart → {AG, PAT}

## 0. Goals

1. Syntactic Object Agreement Rule

2. Test, data

3. Syntactic and semantic factors which influence  
the phenomenon

4. Stochastic OT – Maximum Entropy Model

5. Experiments

6. Results, ideas

**7. More and even stranger factors**



## 7. More and even stranger factors

### Rhythm and rhyme

**75** Jánost mindenki érdekl. **79%**

**76** Jánost mindenki érdekel. **21%**

**45** Jánost minden érdekl. **63%**

**46** Jánost minden érdekel. **38%**

**87 Mindenkit mindenki érdekl. 14%**

**88 Mindenkit mindenki érdekel. 86%**

**89 Mindenki mindenkit érdekl. 6%**

**90 Mindenki mindenkit érdekel. 94%**

I thank Katalin É. Kiss, Tamás Bíró and Dániel Vásárhelyi for comments and discussion.

## References

- Bartos, H. 2000 *Az inflexiók jelenségei szintaktikai háttere*. In: Strukturális Magyar Nyelvtan 3. Morfológia. Akadémiai Kiadó
- 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.
- Kálmán, L. 2009. április 7., *Az analógiás nyelvelmélet (és számítógépes modellezésének első lépései)*, RIL HAS
- Prince, A. and P. Smolensky. 1993 *Optimality Theory*. Ms., Rutgers University., 136 o.
- Rebrus, P. 2009. március 10., *Paradigmák és alakváltozatok a magyarban*, RIL HAS