Humboldt-Universität zu Berlin
Institut für deutsche Sprache und Linguistik
20.06.2014

Verb class effects on word order:
Evidence from a comparative corpus study

Elisabeth Verhoeven
elisabeth.verhoeven@cms.hu-berlin.de

Introduction

Experiencer-first effects:

accusative-first orders

evidence from several languages and different types of data: Lenerz 1997; Belletti & Rizzi 1998; Anagnostopoulou 1999; Scheepers 1997; Haupt et al. 2008, etc.

choice of a lower argument as subject (non-active voice)

evidence from production studies; Ferreira 1994, Lamers & De Hoop (forth.)

Phenomena of the former – but not of the latter – type

can be reflexes of the basic syntactic configuration.

Introduction

Questions

Does the **verb class** has a genuine effect on word order or is this effect fully explained by independent factors – either contextual (givenness) or inherent (animacy)?

A. Is there an experiencer-first effect and if yes, is it **independent from further semantic and pragmatic factors**?

B. Is the frequency of non-canonical orders with experiencer-object verbs part of a **general preference for the earlier realization of experiencers**?

C. There are several subclasses of experiencer object verbs depending on case (accusative, dative) or agentivity of the actor (agentive, non-agentive). **Which distinction is relevant** for the linearization properties?

Introduction

What do we expect to learn from cross-linguistic comparisons?

Languages differ with respect to

- the syntactic operations that allow for different linearizations

    OS through scrambling in German

    OS through left-dislocation in Chinese

    OS through clitic left-dislocation in Greek

- semantics of non-active voice (which restricts its use as a linearization option)

    reflexives (German)/mediopassives (Greek) and stativity

Comparing languages is informative

    for the type of syntactic operation that is sensitive to experiencer-first effects.

Contents

Word order frequencies with different verb classes in corpora

German

Does the verb class has a genuine effect on word order or is this effect fully explained by independent factors – such as animacy?

Language comparison

Scrambling in German vs. left dislocation in Greek/Chinese

German: Corpus study

corpus	W-öffentlich of COSMAS database, Written Language, 2.291.520.000 word forms
extracted	10 verbs for every verb class (four verb classes) 1000 tokens per verb (randomized): total 40000 tokens (background: verb as random factor)
valid tokens	main clauses with two realized arguments (pronouns excluded due to particular rules in German) total: 40 000; valid: 4319
decoding	order: SO vs. OS voice: active vs. non-active field: $XV_{fin}Y$ (prefield), $V_{fin}XY$ (middlefield) definiteness: definite vs. indefinite animacy: animate vs. inanimate

Verb classes

Canonical transitive verbs

beeinträchtigen ‘impair’, *behindern* ‘hinder’*, schützen* ‘protect’*,* etc.

(particular subclass of canonical verbs with include animacy configurations similar to EO verbs; the question is whether EO effects are restricted to experiencers and not to achieve generalizations about the behaviour of all canonical transitive verbs)

Dative Experiencer-Object verbs

imponieren ‘impress’*, gefallen* ‘please’*, widerstreben* ‘be reluctant’, etc.

Accusative Experiencer-Object ±agentive verbs

nerven ‘*bother‘, überraschen* ‘*surprise*’*, ärgern* ‘annoy’*,* etc.

Accusative Experiencer-Object non-agentive verbs

interessieren ‘concern’*, wundern* ‘astonish’*, freuen* ‘give pleasure’ *,* etc.

Choice of order

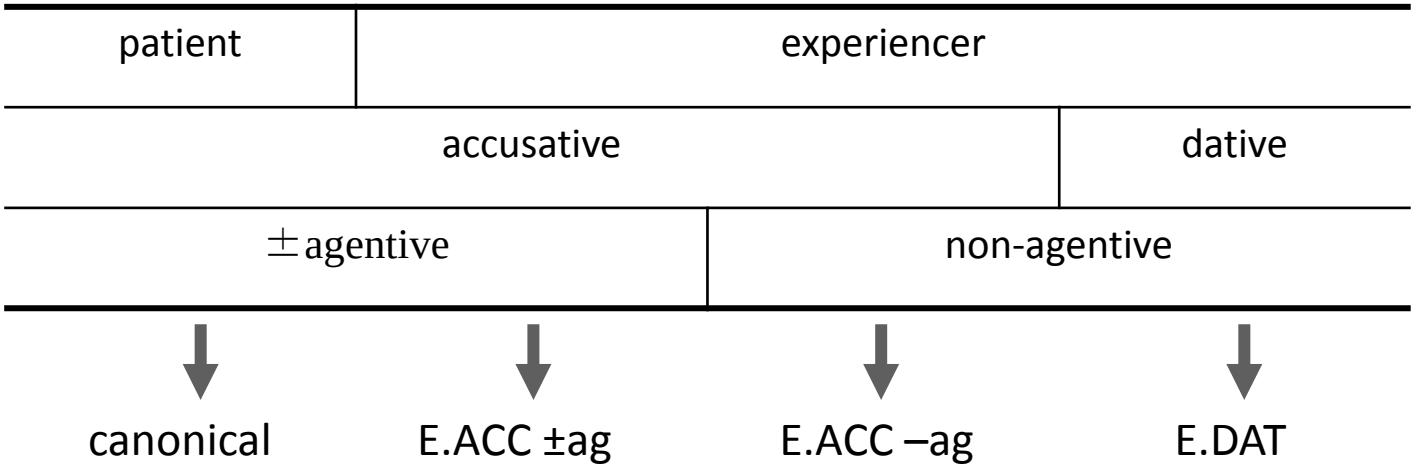
Questions

A. Do **verb classes** have an **influence on word order?**

A1. (If yes) is this influence independent from other factors (animacy/definiteness)?

A2. (If yes) where is the locus of the influence?

- thematic relation of the **undergoer**: patient vs. experiencer
- thematic properties of the **actor**: agentive/non-agentive
- **case**: dative vs. accusative



Examined factors

verb class	canonical
	E.ACC ±agentive
	E.ACC non-agentive
	E.DAT
clausal domain	middlefield (basic configuration)
	prefield (derived configuration)
animacy	disharmonic (animate O & inanimate S)
	other permutations
definiteness	disharmonic (definite O & indefinite S)
	other permutations
random factor	verb

Cases of interest

SX order, prefield

Die Nachricht überraschte den Stürmer gestern.

‘The news surprised the striker yesterday.’

SX order, middlefield

Gestern überraschte die Nachricht den Stürmer.

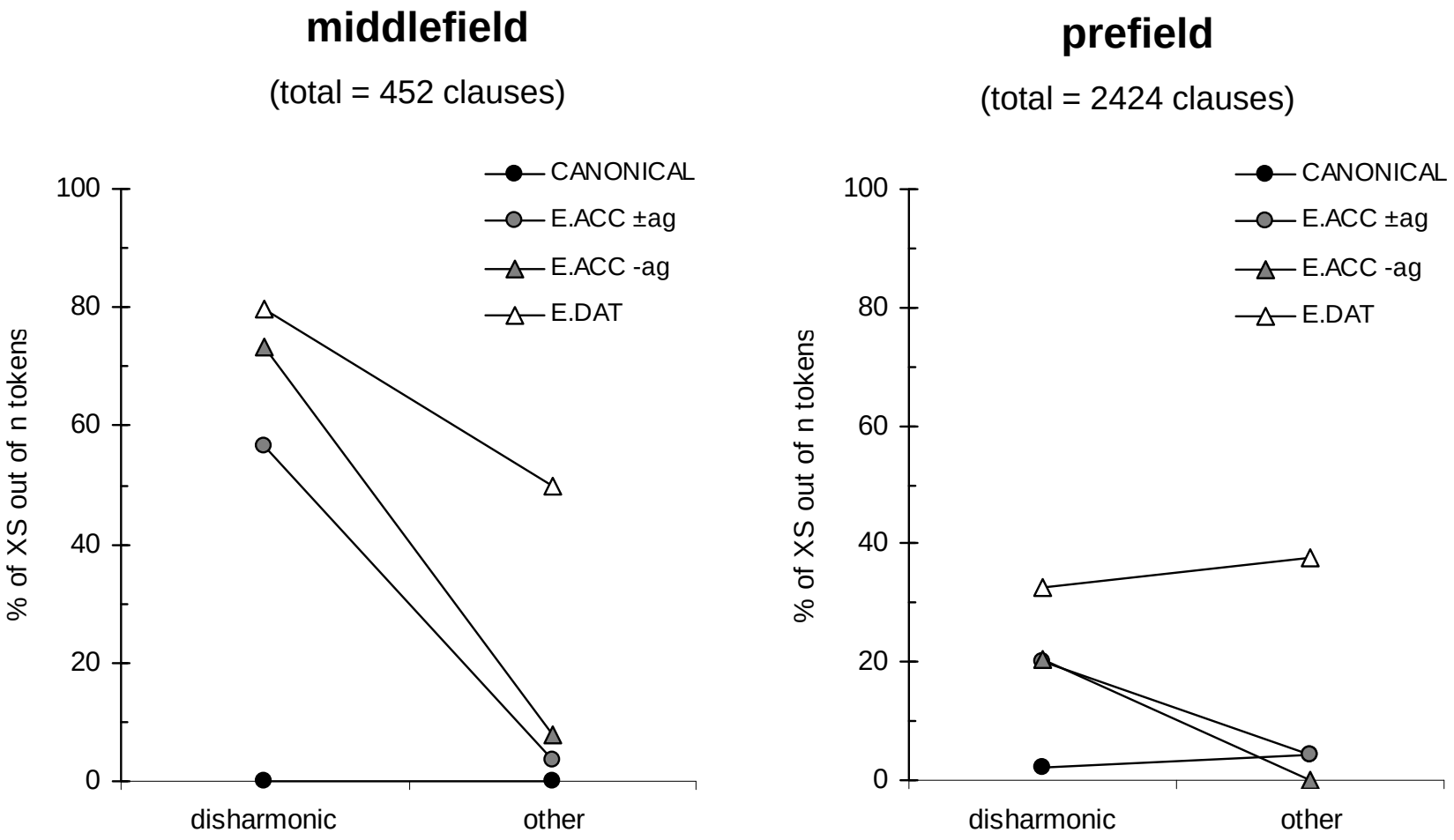
XS order, prefield

Den Stürmer überraschte die Nachricht gestern.

XS order, middlefield

Gestern überraschte den Stürmer die Nachricht.

Animacy and clausal domain



Results I: verb-class contrasts

AIC	patient	experiencer		
	accusative			dative
	±agentive		non-agentive	
2470.1	***α***	***β***		***γ***
2475.7	***α***	***β***	***γ***	***δ***

Comparison of all models (fourfold, threefold, twofold contrasts) reveals that the maximal goodness of fit is reached by "**Canonical vs. E.ACC vs. E.DAT**"

(Log-likelihood test comparing this model with the full model: not significant).

Calculations with the function *glmer* of the *R*-package *lme4* (Bates et al. 2011)

Results II: confirmed effects

definiteness

main effect and interaction effects not statistically confirmed (LLT : not significant).

verb class : animacy

log-likelihood test: $\chi^2(2) = 13.4$, $p < .01$

canonical verbs:	no animacy effect
EO.ACC:	experiencer fronting with disharmonic animacy
EO.DAT:	experiencer fronting independent of animacy

source of the case effect: blocking effect with ACC&harmonic

verb class : field

log-likelihood test: $\chi^2(2) = 13.9$, $p < .001$

larger effects of class in the middlefield than in the prefield

relevance: assumptions about German syntax; previous findings (Bader & Häussler 2010)

animacy : field

log-likelihood test: $\chi^2\ (1) = 9.1$, $p < .01$

larger effect of disharmonic animacy in the middlefield.

Choice of subject

Questions

B. Do **verb classes** have an **influence on the choice of subject?**

B1. Independence from other factors

B2. Locus of influence

Relevance of this question

If the linearization effect relates to a general preference for realizing the experiencer early in the utterance, it should be equally manifested in alternative structures:

- fronting the EO.

- choosing an ES construction.

Cases of interest

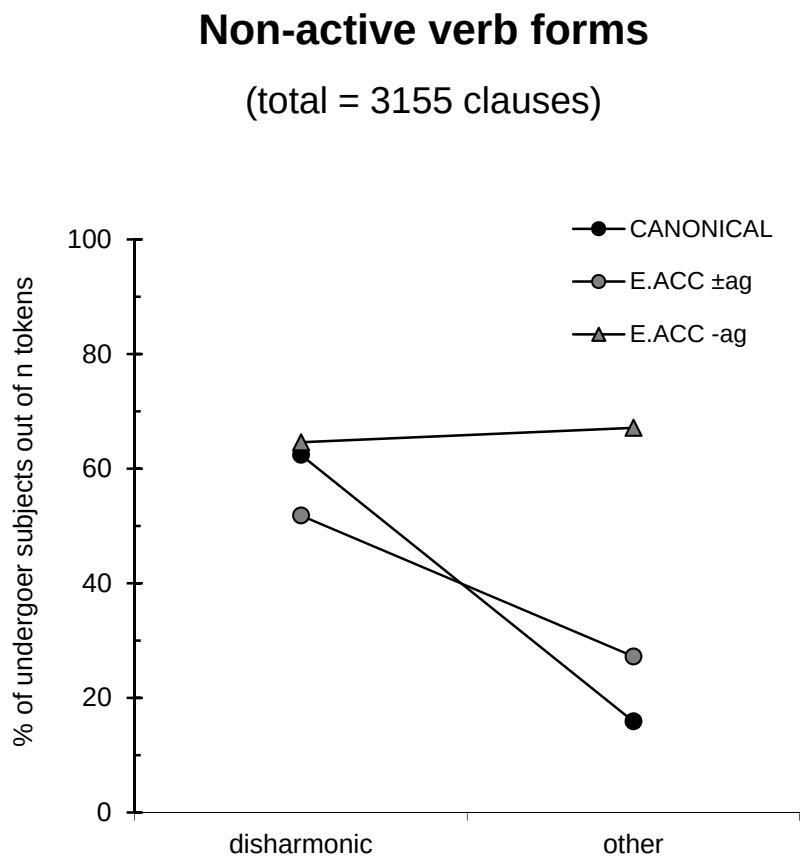
actor subject

Die Profis enttäuschten die Fans.
'The professional players disappointed the fans.'

undergoer subject

Die Fans wurden / waren von den Profis enttäuscht.
'The fans were disappointed by the professional players.'

Animacy and choice of subject



Results: verb-class contrasts

AIC	patient	experiencer	
	±agentive		non-agentive
3150.7	***α***		***β***
3152.2	***α***	***β***	***γ***

Comparison of all models (threefold, twofold contrasts) reveals that the maximal goodness of fit is reached by "±**agentive vs. non-agentive**"

(Log-likelihood test comparing this model with the full model: not significant).

verb class : animacy

log-likelihood test: $\chi^2(1) = 7.9$, $p < .01$

source of this effect: non-agentive verbs display high proportions of non-active voice across animacy levels

Comparative corpus study

Greek and Chinese

corpus	Chinese: *CCL* Corpus; 264 444 436 Modern Chinese characters, 84 127 123 Old Chinese characters; Greek: Hellenic National Corpus, ILSP, 47.000.000 words
extracted	10 verbs for every verb class (two verb classes) 250 tokens per verb (randomized): total 5000 tokens (background: verb as random factor)
valid tokens	main clauses with two lexical arguments Chinese; valid: 640 Greek; valid: 701
decoding	order: SO vs. OS voice: active vs. non-active definiteness: definite vs. indefinite animacy: animate vs. inanimate clitic doubling (Greek)

Object-first orders: Greek

Clitic left-dislocation

tin	*maría*	*tin*
DEF:ACC.SG.F	Maria:ACC.SG.F	3.SG.ACC.F
enoxlí	*o*	*pétros*.
bother:3.SG	DEF:NOM.SG.M	Peter:NOM.SG.M

'Petros bothers Maria.'

Discourse triggers for clitic left-dislocation:

-with canonical verbs: contrastive topicalization or hanging topics (see Alexopoulou and Kolliakou 2002, Anagnostopoulou 1997, Grohmann 2003)

-with EO verbs: obligatory according to intuitions (Anagnostopoulou 1999), not confirmed in corpus (Verhoeven 2009) or acceptability experiments (Temme & Verhoeven 2014)

Object-first orders: Chinese

Left dislocation

Zhāngsān	*wǒ*	*yǐjīng*	*jiàn*	*guo*	*le*
Zhangsan	1.SG	already	see	EXP	CRS

‘Zhangsan, I have already seen him.’ (see Li and Thompson, 1981:15)

Discourse triggers for left-dislocation in Chinese:

-Left-dislocated objects occur as hanging topics (see Huang et al. 2009, Li and Thompson, 1981)

-Rare in speech production (1 OSV out of 397 clauses with two lexical arguments in Verhoeven 2014)

Object-first in corpus

	Greek		Chinese	
	canonical	EO	canonical	EO
SVO	255	233	316	206
SOV	–	7	5	46
OVS	9	9	–	–
OSV	–	1	–	–
VOS	2	9	–	–
VSO	1	1	–	–
Total	267	260	321	252

Object-fronting in languages with left-dislocation is more rare than object-fronting in scrambling languages, vgl. German, 222 Accusative-first out of 1712 clauses, 13% (across verb classes and animacy levels).

Non-active forms

Chinese

Kèrén *bèi* *tāng* *mízhù*.
guest BEI soup attract
'The guest is enticed by the soup.'

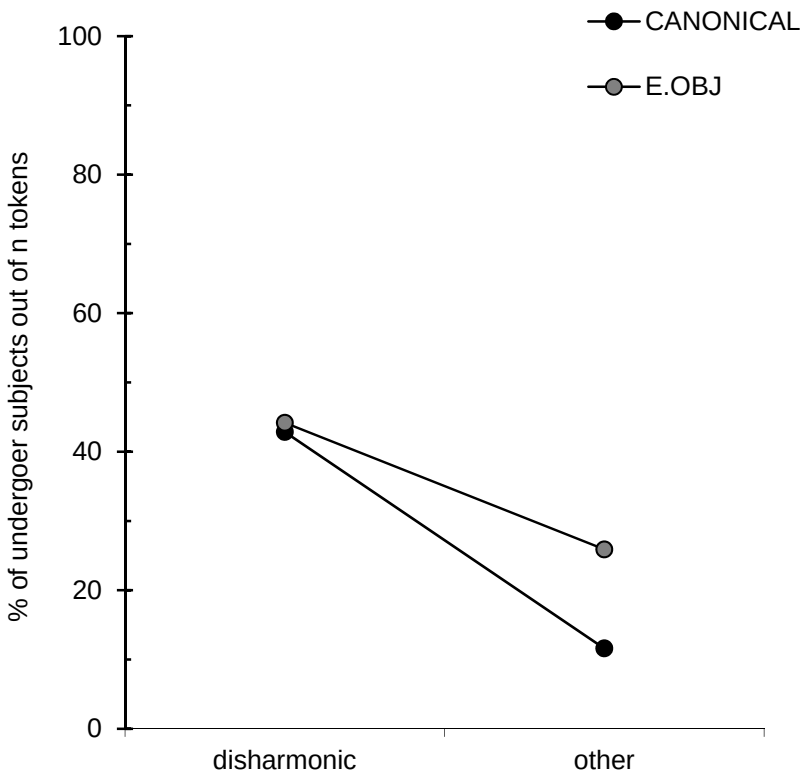
Greek

i *maría* *enoxlíte*
DEF:NOM.SG.F Maria:NOM.SG.F bother:MEDIOPASS.3.SG
apó *ton* *pétro*.
by DEF:ACC.SG.M Peter:ACC.SG.M
'Petros bothers Maria.'

Choice of subject

Greek

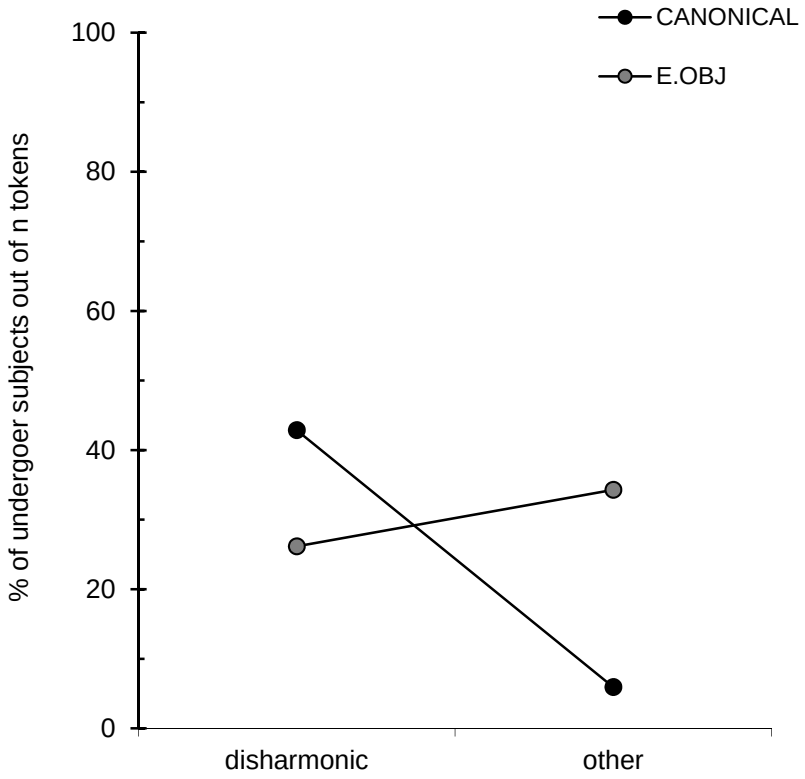
(valid = 701)



Interaction, LLT, $\chi^2(1) = 4.2$, $p < .05$

Chinese

(valid = 640)



Interaction, LLT, $\chi^2(1) = 7.2$, $p < .01$

Conclusions

Conclusions

Word order

Experiencer-first triggers scrambling in German, but not left-dislocation in Greek/Chinese.

In particular for *German*: difference between canonical and EO verbs; additional effect of case (with ACC blocking fronting under harmonic animacy)

Choice of subject

Animacy and verb class interact in all languages such that

the contrast between animacy levels is smaller for EO verbs.

Differences in magnitude may be due to differences in the semantics of non-active forms.

high frequency in German : non-compensatory effects with word order

Supplementary material

Verb classes

Canonical transitive verbs

beeinträchtigen, behindern, blenden, infizieren, *heilen*, *retten*, *schützen, vergiften*, *wecken, zerquetschen*

Dative Experiencer-Object verbs

imponieren, *auffallen, behagen*, *einfallen, entfallen, gefallen*, *leidtun*, *missfallen, entgehen, widerstreben*

Accusative Experiencer-Object ±agentive verbs

nerven, überraschen, ärgern, erschrecken, reizen, langweilen, amüsieren, irritieren, aufregen, enttäuschen

Accusative Experiencer-Object non-agentive verbs

interessieren, wundern, freuen, befremden, bedrücken, entsetzen, ekeln, faszinieren, empören, anwidern

Corpus

	canonical	EO ACC ±ag	EO ACC –ag	EO DAT	total
total extract	10 000	10 000	10 000	10 000	40 000
main decl. clauses, two args.	1 248	2 047	2 873	3 593	9 761
two lexical arguments	990	974	1 191	1 164	4 319
active voice	767	527	418	1 164	2 876

Table 1. Categories in the corpus

Verb classes and agentivity

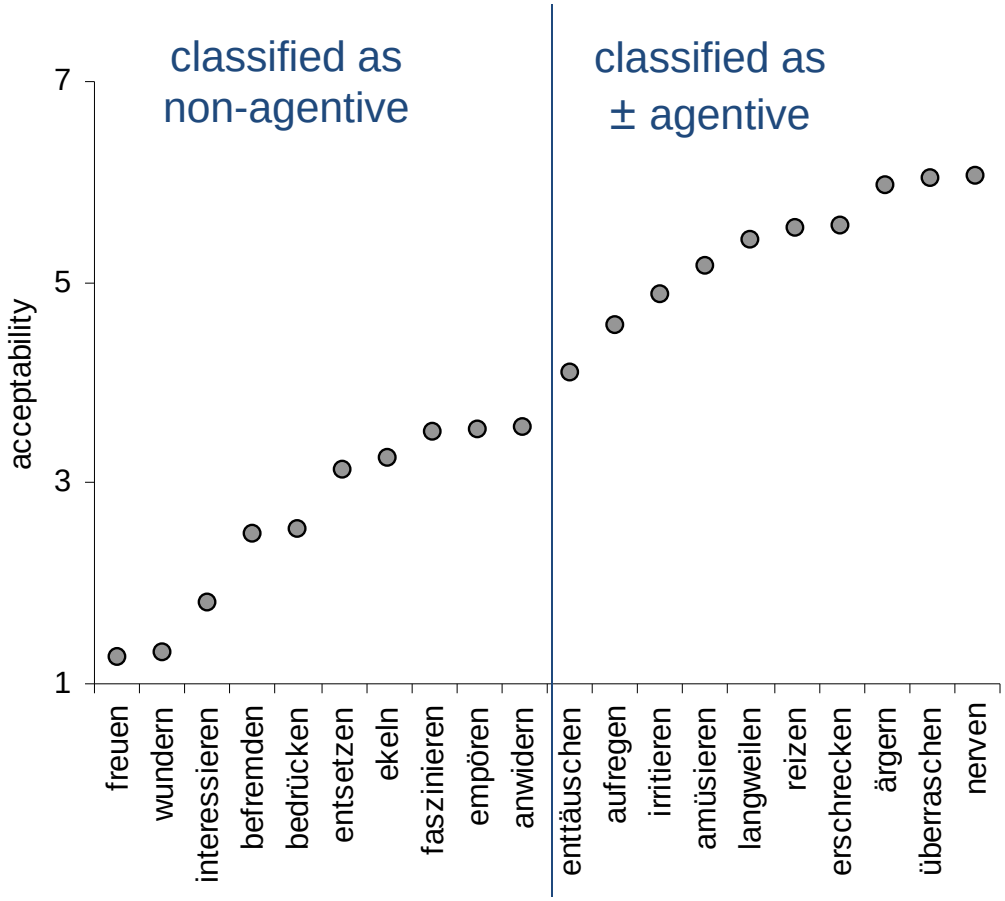
Agentivity test frame with control verbs:

X entschied, Y zu V ‚X decided to V Y'

scalar acceptability judgments

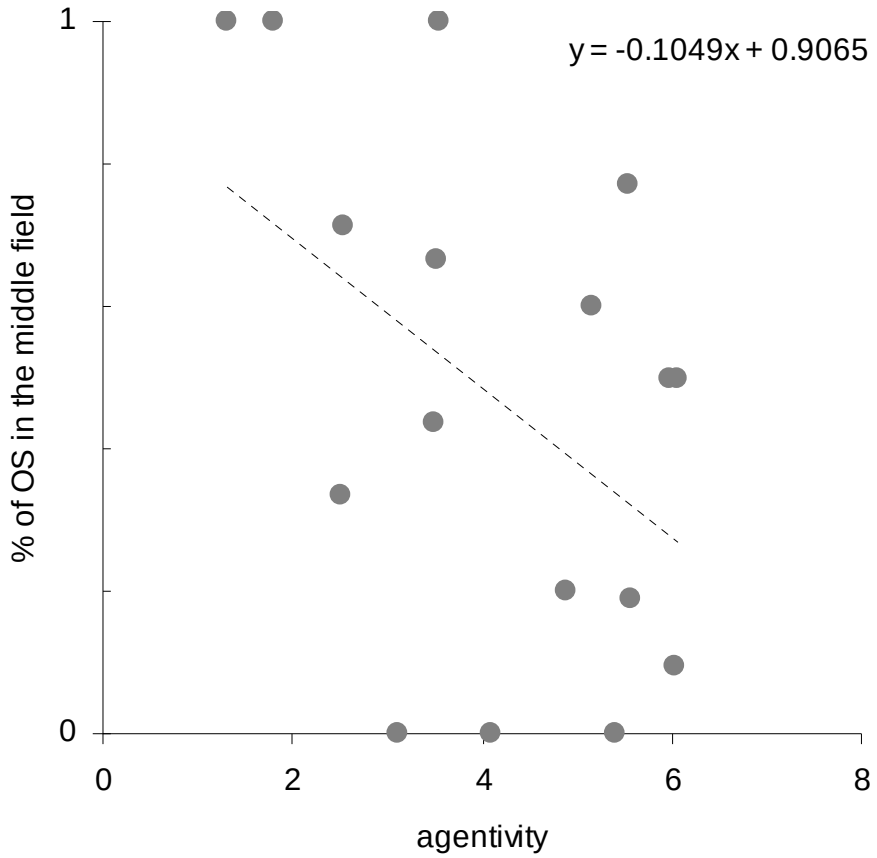
(*1*: non-acceptable; *7*: acceptable)

n of speakers: 32



Corpus study: EO verbs

Is agentivity a good predictor for the choice of OS in the middle field?



Pearson *R* = –0.45

Corpus study: EO verbs

Two *generalized mixed logit models* on the data with ACC EO verbs

argument order ~ verb class (fixed factor); verb (random factor)

(a) verb class = factor with two levels

(b) verb class = interval-scaled factor

(a) two-levels model (+-agentive vs. non-agentive):

Akaike Information Criterion: 173.02

(b) scalar model:

Akaike Information Criterion: 172.16

Results (plotted for definiteness levels)