Expectation-based processing of grammatical functions in Swedish

- presentation at CLASP 17/4 2019

Thomas Hörberg,

Department of Psychology & Department of Linguistics,

Stockholm University

Stockholms universitet

- Background and assumptions
- Expectation-based speech production: Corpus-based evidence for adaptation
- Expectation-driven model of incremental GF assignment during comprehension
- Experimental test of model predictions

Bakground and assumptions

Stockholms universitet

Expectation-based processing

- Language users aim for efficient information transfer as based on their expectations

Production

- encoding into a linguistic signal is influenced by a trade-off between
  - ensuring sufficient information transfer (”iconicity”)
  - limiting production costs (”economy”)

  (e.g., Kurumada & Jaeger 2015)

Comprehension

- decoding of the linguistic signal is done on the basis of statistical regularities in the input (e.g. Levy 2008; MacDonald 2013; Venhuize et al. 2018)

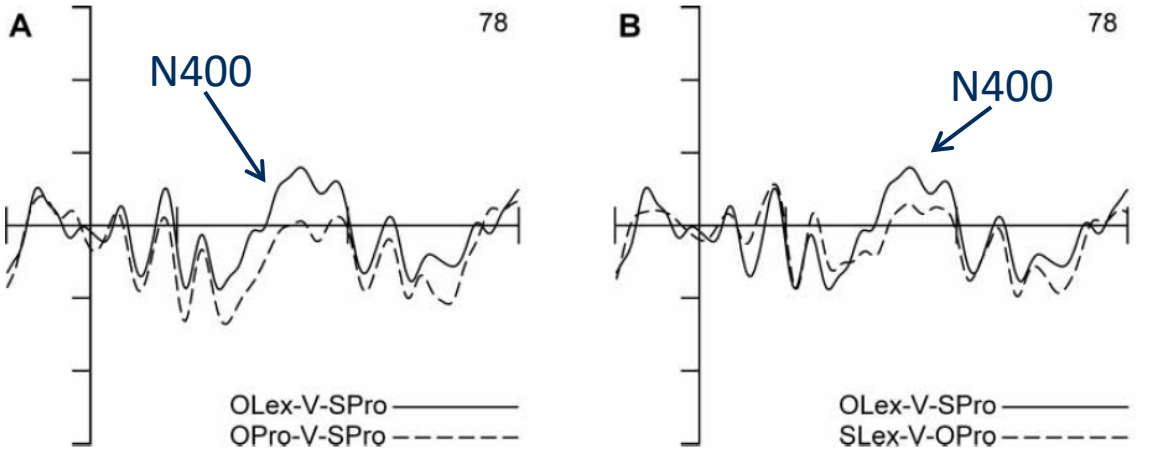
Test case: Transitive sentences in Swedish

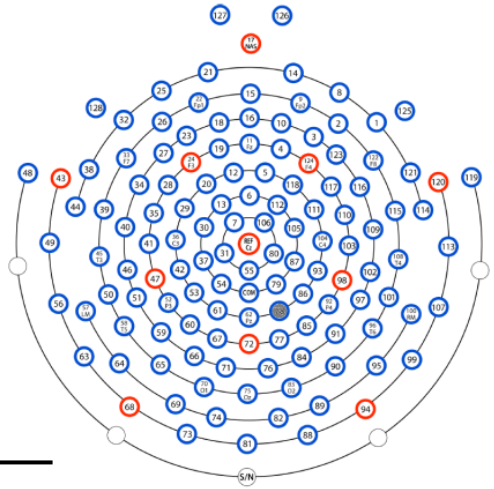
Bagarna	visade	**oss**	till	köket
baker-DEF.PL	show-PRT	2PL.OBJ	to	kitchen-DEF
Bagarna	visade	**ni**	till	köket
baker-DEF.PL	show-PRT	1PL.SBJ	to	kitchen-DEF
Oss	visade	**ni**	till	köket
1PL.OBJ	show-PRT	2PL.SBJ	to	kitchen-DEF

- potentially locally ambiguous!

Grammatical function comprehension in Swedish

(Hörberg et al. 2013)





Condition	Example sentence				
OLex-V-SPro	Bagarna	visade	**ni**	till	köket
	baker-DEF.PL	show-PRT	2PL.SBJ	to	kitchen-DEF
SLex-V-OPro	Bagarna	visade	**oss**	till	köket
	baker-DEF.PL	show-PRT	1PL.OBJ	to	kitchen-DEF
OPro-V-SPro	Oss	visade	**ni**	till	köket
	1PL.OBJ	show-PRT	2PL.SBJ	to	kitchen-DEF

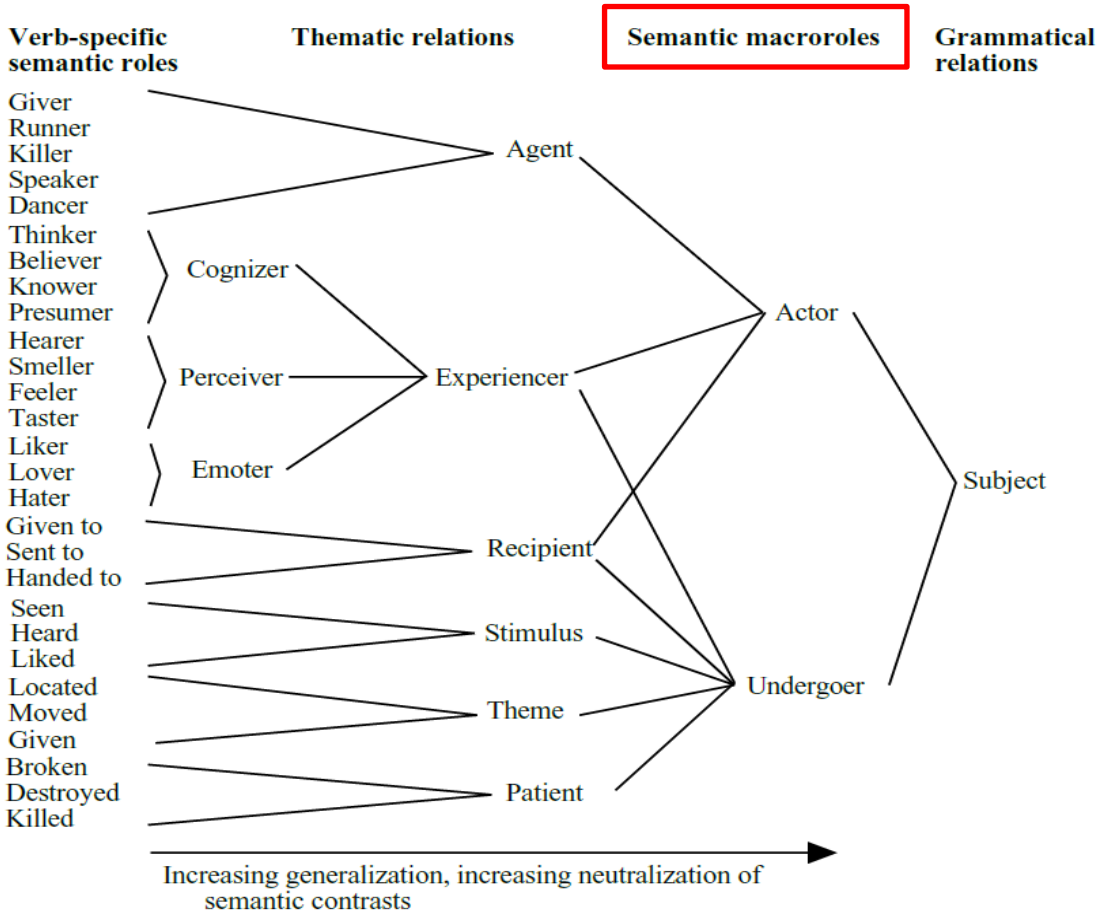
- N400 reflects re-assignment of semantic proto-roles to NP arguments

Stockholms universitet

Grammatical functions (GFs)

- Grammatical encoding:
  - morphosyntactic encoding: word order, case marking, agreement
  - syntactic behavior restrictions: e.g. co-reference: “She$_i$ kissed Charlie$_j$ and t$_{i/*j}$ went home”
- Express argument functions:
  - semantic macroroles (Actor / Undergoer)
  - information structure (Topic / Focus)

GFs and semantic macroroles



(Van Valin 2005)

GFs and semantic macroroles

Intransitive subject → Actor + Undergoer

- “Pojken faller”
- “Pojken springer”

Transitive subject → Actor

Direct object → Undergoer

- “Pojken sparkar bollen”

Subject of passive → Undergoer

- “Bollen sparkas av pojken”

GFs and information structure

- GFs strongly connected to information structure

Transitive subject: topic

Direct object: (part of) focus

Subjects (in e.g. Swedish): ‘syntacticized topics’ (Foley 2011)

”Mannen_i skjöt offret och $\_\__i$ sprang iväg men $\_\_\__i$ blev tagen av polisen på flykten”

Argument prominence properties

- grammatical encoding of GFs conditioned on prominence (e.g. Silverstein 1976)

Animacy:	human < animate < inanimate
Person:	first, second < third
Referentiality:	pronoun < proper name < common noun
Definiteness:	definite < specific indefinite < unspecific indefinite

- Subject > Object in prominence
- Exceptions (i.e. Object > Subject) typologically marked and infrequent in discourse

Argument prominence properties

Fore (Trans-New Guinean; from Scott 1986):

Yaga: wá aegúye	Yaga:-ma wá aegúye
Pig man 3sg-hit-3sg	Pig-ERG man 3sg-hit-3sg
”The man attacks the pig”	“The pig attacks the man”

Picurís (Kiowa-Tanoan; from Zaharlick 1982):

Sənene ti-mǫn-ʔąn	ta-mǫn-mia-ʔąn	sənene-pa
Man 1SG-see-PST	1SG-see-PASS-PST	man-OBL
“I saw the man”	“The man saw me”	

GFs and prominence properties

- correlations between prominence and GFs in language use

Prominence feature	Subject	Direct object
Animate	78%	18%
1st / 2nd person	28%	3%
Pronominal	55%	18%
Definite	80%	50%
Given	71%	32%

(Hörberg 2016)

Verb semantic properties

- prominence properties interact with verb semantic properties

  (e.g. volitionality, sentience, causation, possession)

- Volitional and sentience / experiencer verbs require animate Actors
- Experiencer verbs often express private knowledge and therefore frequently occur with 1st/2nd person subjects (Dahl 2000)

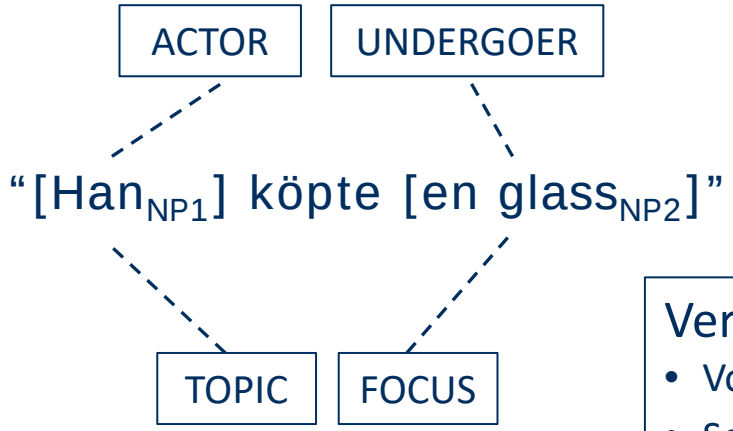
GFs and verb semantics

- Verb class can condition grammatical encoding of GFs

Acehnese (Austronesian; from Durie 1987):

Gopnyan	geu-jak	Gopnyan	geu-mat-lôn
S/he	3sg-go	S/he	3sg-hold-1sg
“She goes”		“She holds me”	
Gopnyan	rhët-geu	Gopnyan	lôn-ngieng-geu
S/he	fall-3sg	S/he	1sg-see-3sg
“She falls”		“I see her”	

Assumptions - Processing of GFs



(e.g. Hörberg 2016)

Morphosyntactic cues:

- Word order
- Case marking
- ...

Verb semantic cues:

- Volitionality
- Sentience
- Causation
- Possession

Prominence cues:

- Animacy
- Definiteness
- Givenness
- Egophoricity
- Number
- ...

Stockholms universitet

Assumptions – Processing of GFs (Hörberg 2016)

- Based on morphosyntactic, NP prominence and verb semantic information **(Argument Interpretation Cues)**
- Both GF encoding an decoding is highly incremental and expectation-based
- Production: Adapt language production in order to avoid redundancies (“economy”) but also to avoid ambiguities (“iconicity”)
- Comprehension: make tentative and probabilistic GF assignments on the basis of the information available as based upon statistical regularities

Adaptation in language production

Stockholms universitet

Adaptation in written swedish (Hörberg 2018)

- Corpus-based study of the distribution of prominence-based (animacy) and morphosyntactic cues to GFs in SVO, OVS and passives

Glossa a journal of general linguistics

Hörberg, Thomas. 2018. Functional motivations behind direct object fronting in written Swedish: A corpus-distributional account. *Glossa: a journal of general linguistics* 3(1): 81. 1–36, DOI: https://doi.org/10.5334/gjgl.502

RESEARCH

Functional motivations behind direct object fronting in written Swedish: A corpus-distributional account

Thomas Hörberg[1,2]

[1] Department of Linguistics, Stockholm University, SE-106 91 Stockholm, SE

[2] Department of Psychology, Stockholm University, SE-106 91 Stockholm, SE

thomas.hoerberg@psychology.su.se

- Do writers prefer using OVS sentences when other GF cues are available?

Sentence properties

Corpus: Svensk Trädbank (SUC + Talbanken), 1.2 million words

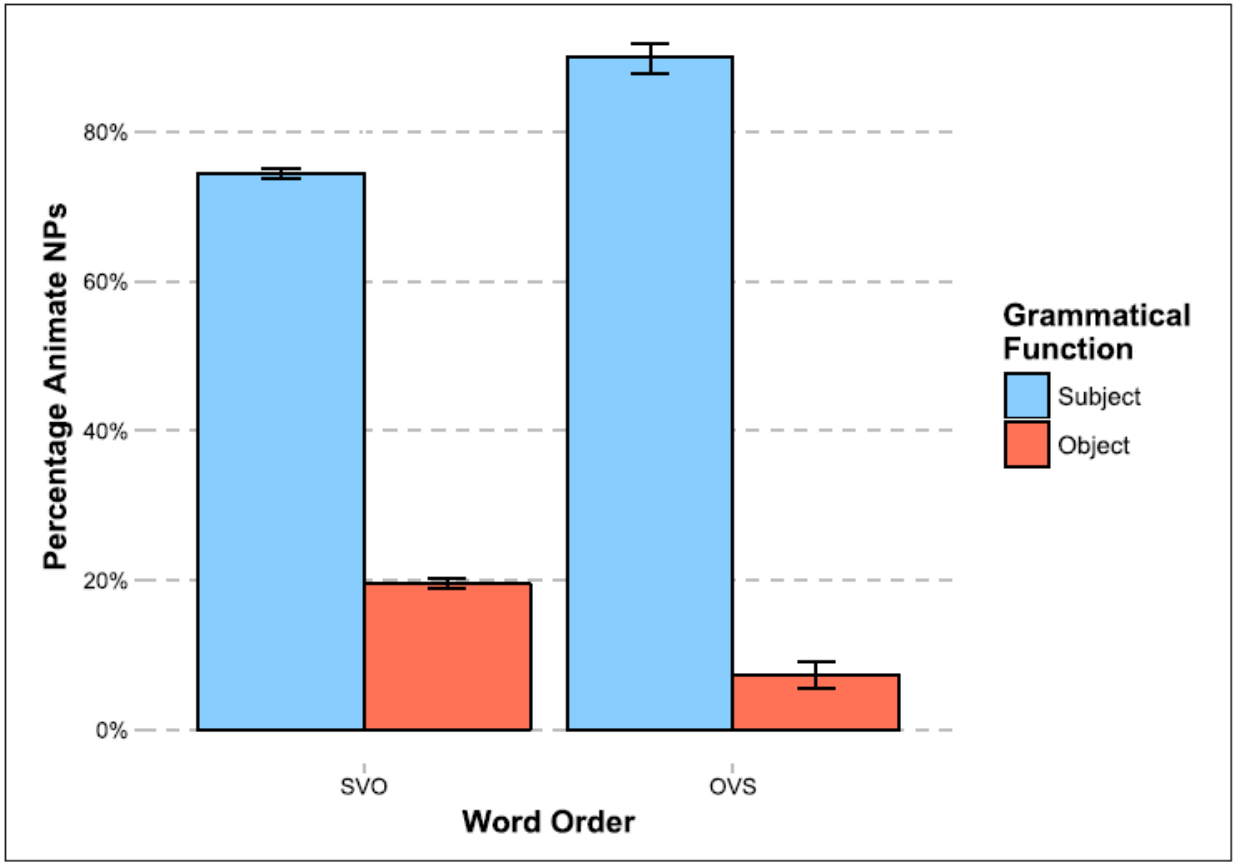
WO	Adverbial position 1	Adverbial position 2
SVO	Barnen får inte äta upp all glass innan middan	barnen inte får äta upp all glass innan middan
OVS	All glass får barnen inte äta upp innan middan	All glass får inte barnen äta upp innan middan
Passive	All glass får inte ätas upp av barnen innan middan	All glass inte får ätas upp av barnen innan middan

- NP:s of any length
- Up to 4 verbs
- Adverbials + verb particles optional
- Annotation: NP animacy, case, auxiliary, verb particle

Number of sentences

Corpus	Genre	Texts	Sentences	Hits						
				SVO		OVS		Passive		Total N
				N	%	N	%	N	%	
SUC	Press: reportage	44	7278	1149	86.3	68	5.1	47	3.5	1264
	Press: Editorial	17	2385	385	81.7	32	6.8	22	4.7	439
	Press: Reviews	27	3961	536	79.8	51	7.6	34	5.1	621
	Skills, Trades and Hobbies	58	8933	1343	82.2	118	7.2	55	3.4	1516
	Popular Lore	48	6525	1160	85.4	55	4.0	72	5.3	1287
	Biographies and Memoirs	26	3598	627	83.8	51	6.8	35	4.7	713
	Miscellaneous	70	10847	1239	78.5	50	3.2	145	9.2	1434
	Learned and Scientific Writing	83	9633	1398	83.1	63	3.7	159	9.4	1620
	General fiction	82	13028	2527	86.2	185	6.3	35	1.2	2747
	Mysteries and Science fiction	19	4070	665	84.2	58	7.3	9	1.1	732
	Light reading	20	2908	611	84.6	53	7.3	5	0.7	669
	Humor	6	1071	183	76.6	27	11.3	2	0.8	212
TB	Brochure texts	25	1733	298	83.2	20	5.6	20	5.6	338
	Newspaper texts	28	1669	277	85.8	16	5.0	14	4.3	307
	Educational texts	14	1624	292	88.0	10	3.0	20	6.0	322
	Debate articles	18	1134	259	89.0	12	4.1	8	2.7	279
Total		585	80397	12949	83.7	869	5.6	682	4.4	14500

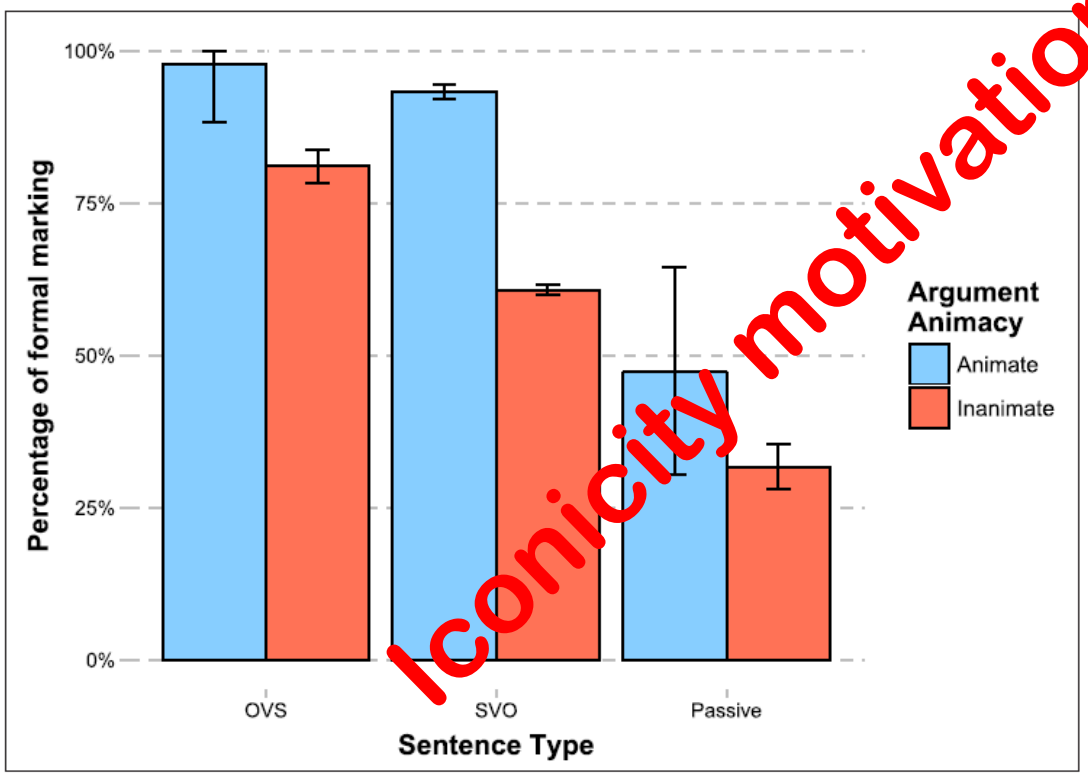
Adaptation in written Swedish



- Use of ambiguous OVS word order preferred when a prominence based GF cue (animacy) is availible

Stockholms universitet

Adaptation in written Swedish

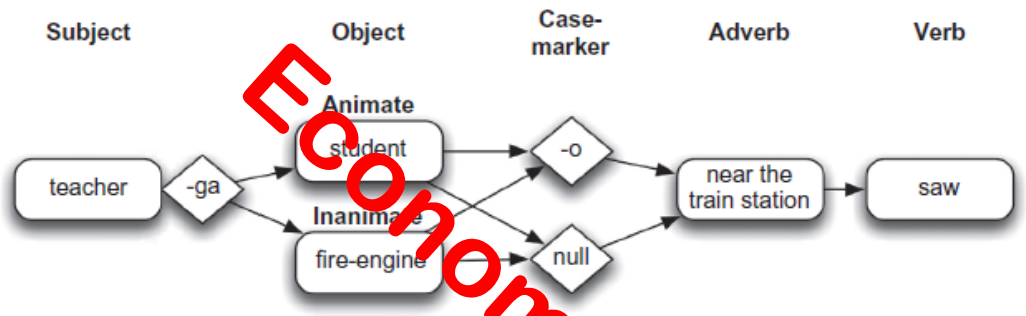


Property	OVS		SVO		Passive		*p*
	N	%	N	%	N	%	
Case marking	588	67.7%	6375	49.2%	59	8.7%	<.0001
Auxiliary verb	323	37.2%	3993	30.8%	167	24.5%	<.0001
Verb particle	121	13.9%	1335	10.3%	22	3.2%	<.0001
Any formal marker	713	82.0%	8429	65.1%	221	32.4%	<.0001

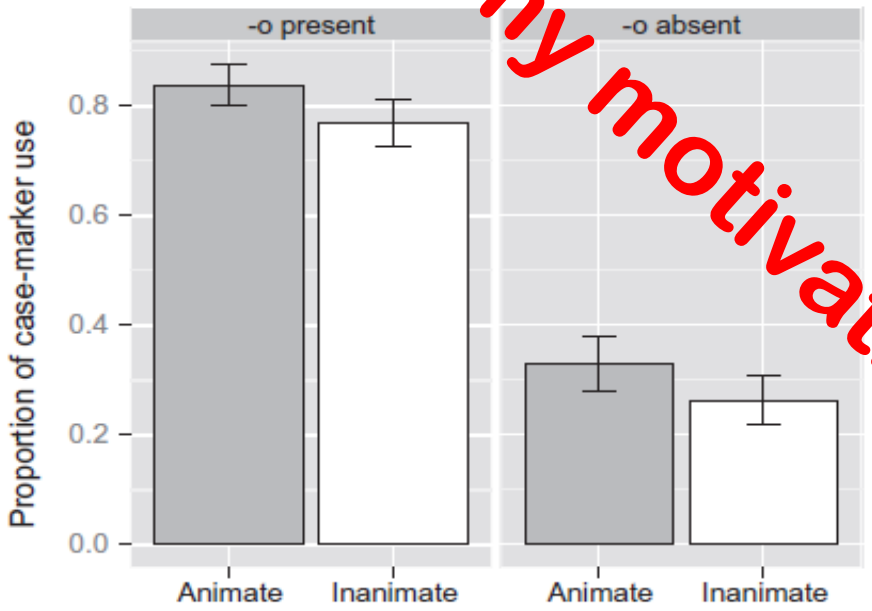
- Use of ambiguous OVS word order preferred when any other GF cue is availible

Stockholms universitet

Adaptation in spoken Japanese



Economy motivation



- Statistically, speakers avoid overt case marking when prominenced-based (animacy) cue is availible

(Kurumada & Jaeger 2015)

Stockholms universitet

Adaptation in speech production

- Writers and speakers adapt their productions towards
  - avoiding ambiguities
  - avoiding redundancies
- This is likely to reflect a trade-off between
  - ensuring sufficient information transfer (”iconicity”)
  - limiting production costs (”economy”)

Expectation-based model of incremental GF assignment

Stockholms universitet

Expectation-based model of incremental GF assignment

- Expectation-driven model of incremental GF assignment on basis of distributional patterns of AICs in Swedish transitive sentences

1. Use corpora to estimate distribution of AICs over subjects and objects in Swedish transitive sentences

2. Calculate probabilities for a given word order / GF assignment based on AICs provided by constituents over time (i.e., NP1, verb, NP2)

3. Estimate incremental change in the expectation for a given WO on basis of these probabilites

Sentence properties

Corpus: Svensk Trädbank (SUC + Talbanken), 1.2 million words

WO	Adverbial position 1	Adverbial position 2
SVO	Barnen får inte äta upp all glass innan middan	barnen inte får äta upp all glass innan middan
VSO	Innan middan får barnen inte äta upp all glass	Innan middan får inte barnen äta upp all glass
OVS	All glass får barnen inte äta upp innan middan	All glass får inte barnen äta upp innan middan

- NP:s of any length
- Up to 4 verbs
- Adverbials + verb particles optional

Corpus properties

Svensk Trädbank: balanced written Swedish texts

Corpus	Genre	N texts	N sentences	N hits	SVO	VSO	OVS
SUC	Press: reportage	44	7278	1495	1495	1149	68
	Press: Editorial	17	2385	473	473	385	32
	Press: Reviews	27	3961	712	712	536	52
	Skills, Trades and Hobbies	58	8933	1840	1840	1343	118
	Popular Lore	48	6525	1503	1503	1160	55
	Belles Letters, Biography, Memoirs	26	3598	805	805	627	51
	Miscellaneous	70	10847	1540	1540	1239	50
	Learned and Scientific Writing	83	9633	1809	1809	1398	64
	General fiction	82	13028	3110	3110	2527	186
	Mysteries and Science fiction	19	4070	826	826	665	58
	Light reading	20	2908	749	749	611	53
	Humor	6	1071	248	248	183	27
TB	Brochure texts	25	1733	390	390	298	20
	Newspaper texts	28	1669	361	361	277	16
	Educational texts	14	1624	374	374	292	10
	Debate articles	18	1134	316	316	259	12
TOTAL		585	80397	**16551**	**12949**	**2730**	**872**

Estimation of AIC distributions over GFs

- Penalized logistic regression modeling
- Estimates individual AIC strengths in terms of predicting WO / GF assignment
- Predicts probability for OS order at NP1, verb, and NP2 conditional on AICs

NP properties
- Animacy
- Givenness
- Definiteness
- Number
- Egophoricity (1st / 2nd vs. 3rd person)
- Pronominality
- Case

Verb semantic properties
- Volitionality
- Causation
- Sentience
- Possession

Additional properties
- NP length
- Adverbial-initial
- Embedded
- Auxiliary verbs

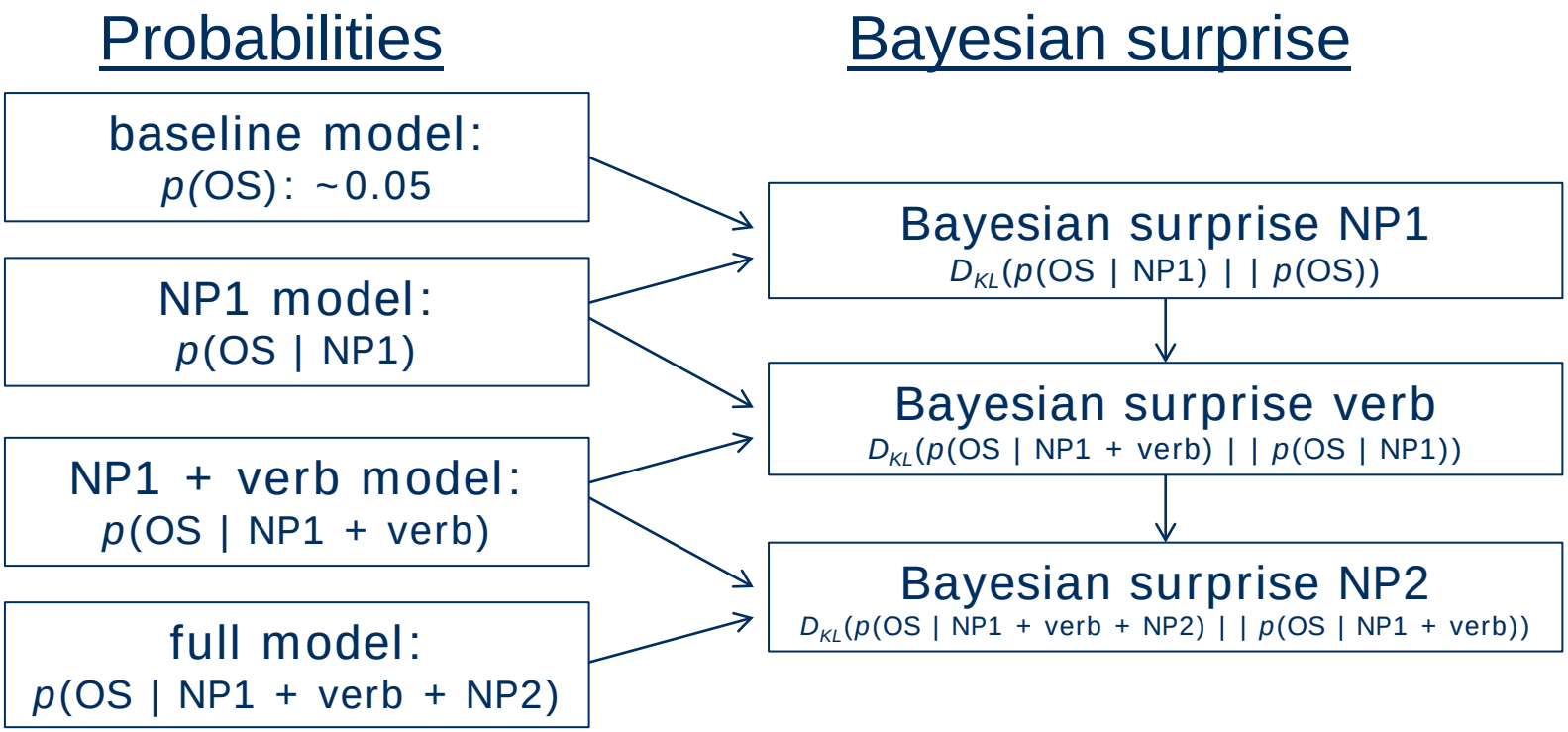
Expectation-based modeling of incremental GF assignment

- Model the incremental change in the expectation of OS word order over time
- Expectations based upon AICs provided by constituents over time (i.e., NP1, verb(s) and NP2)

Expectation-based modeling of incremental GF assignment

- expectation change for OS modelled in terms of Bayesian surprise / relative entropy / the Kullback–Leibler divergence:
  $D_{KL}(P||Q) = \Sigma_i \log(P_i/Q_i)P_i$
- Similar to word surprisal ($-\log(p(W_i \mid W_{i-1}...W_1))$ but applies to constituents with respect to predicting WO / GF assignment
  (e.g. Levy 2008)

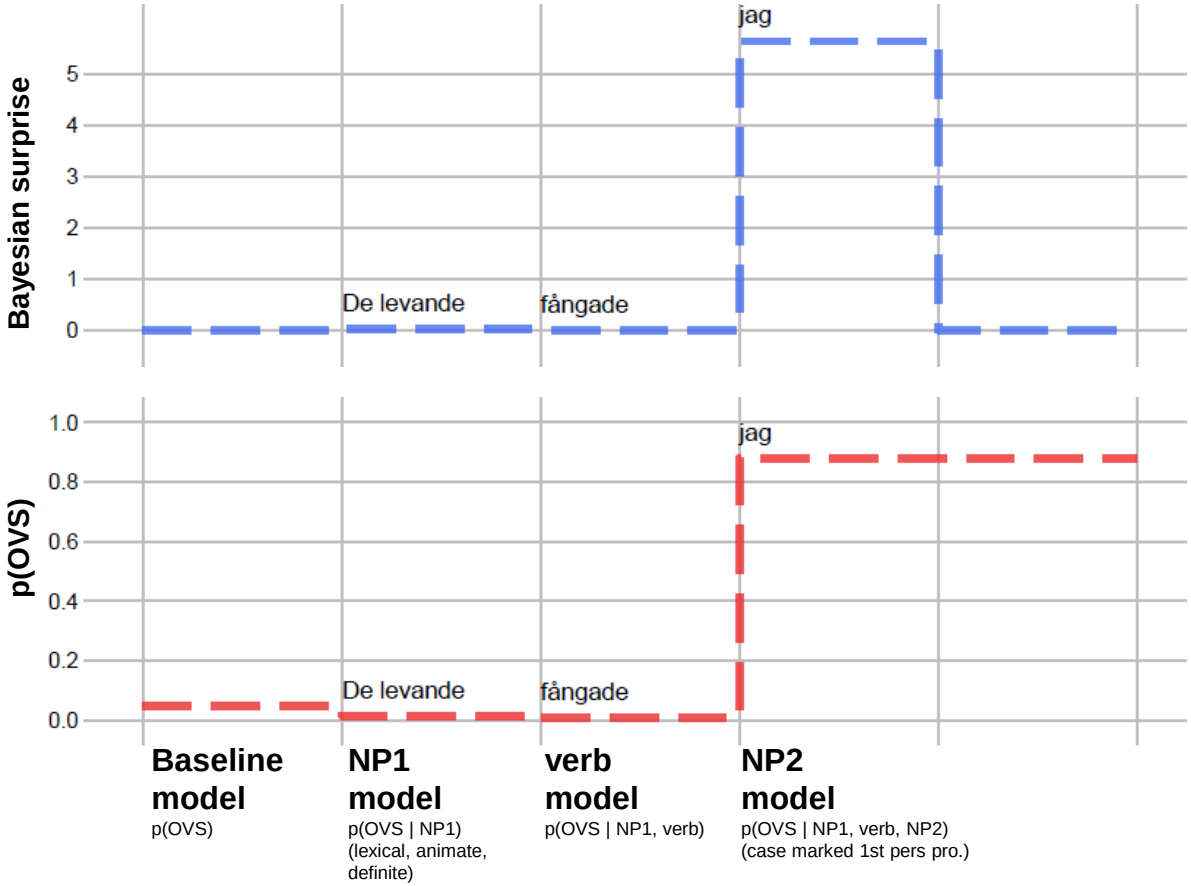
Expectation-based modeling of incremental GF assignment



Stockholms universitet

Bayesian surprise in original data

- OVS sentences with initial lexical NP and final case marked NP



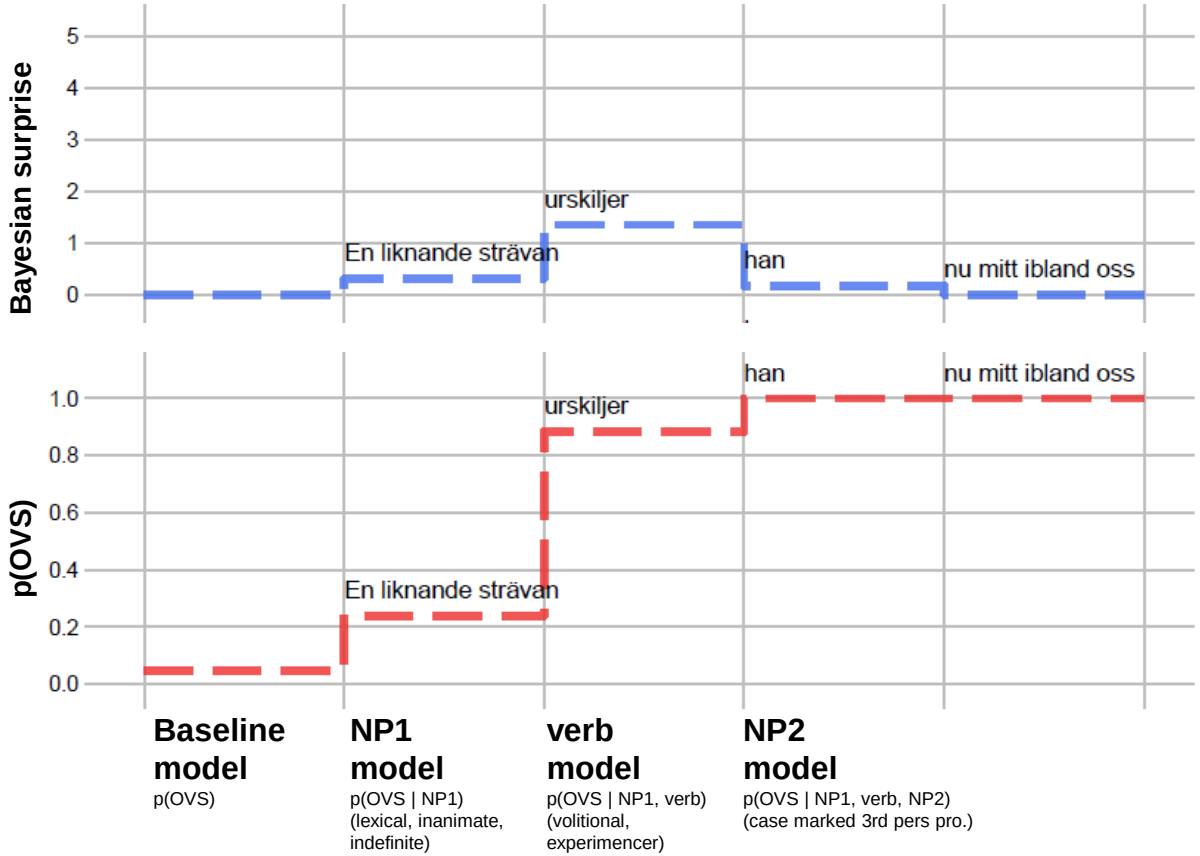
[De levande D_{KL} = 0.02] [fångade D_{KL} = 0.00] [jag D_{KL} = 5.64]
The living caught I

”The living, I caught them”

Stockholms universitet

Bayesian surprise in original data

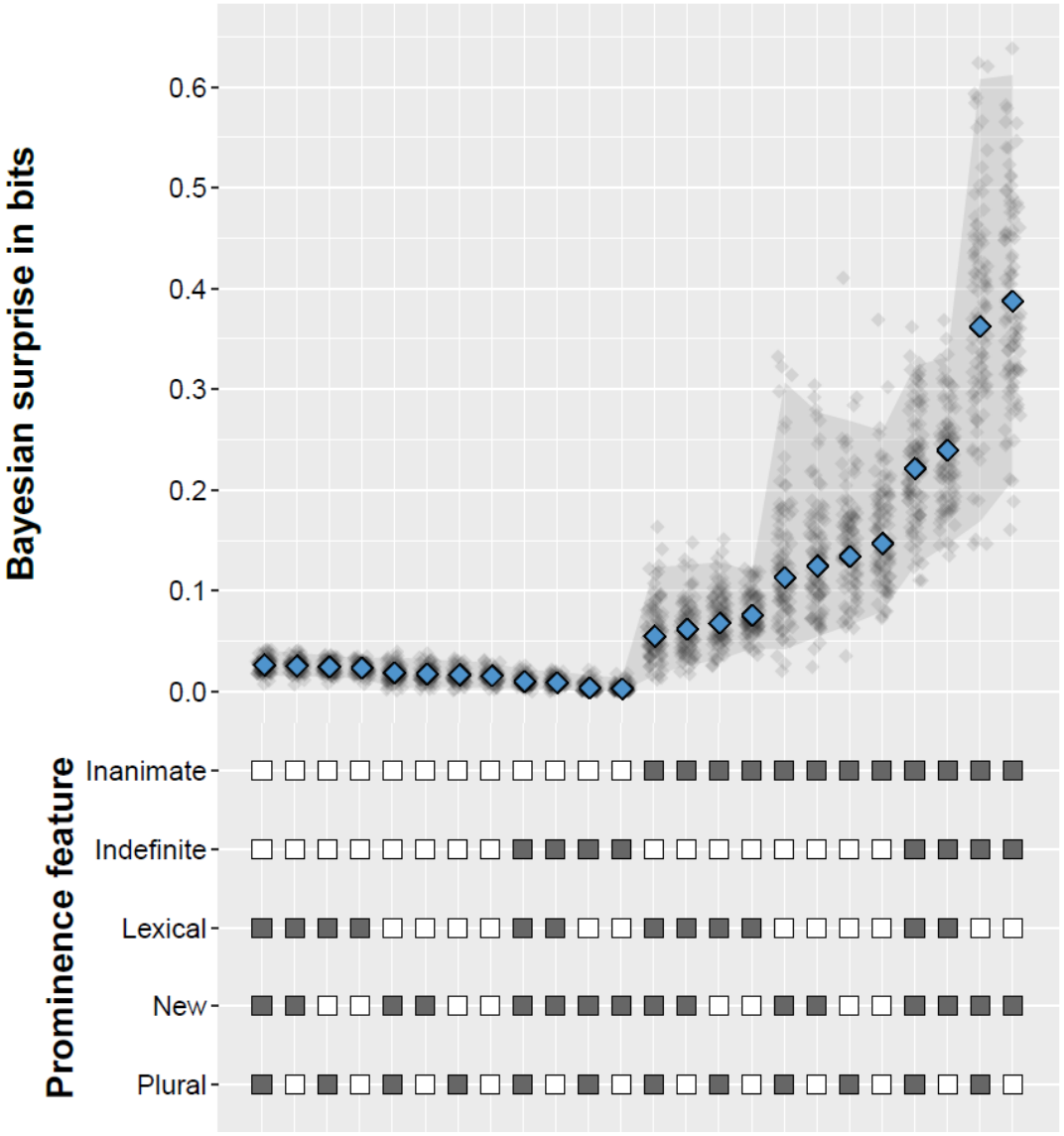
- OVS sentences with initial lexical NP and final case marked NP



[En liknande strävan $D_{KL} = 0.31$] [urskiljer $D_{KL} = 1.35$] [han $D_{KL} = 0.17$] nu...
A similar endavour discerns he now...
”A similar endavour he now discerns”

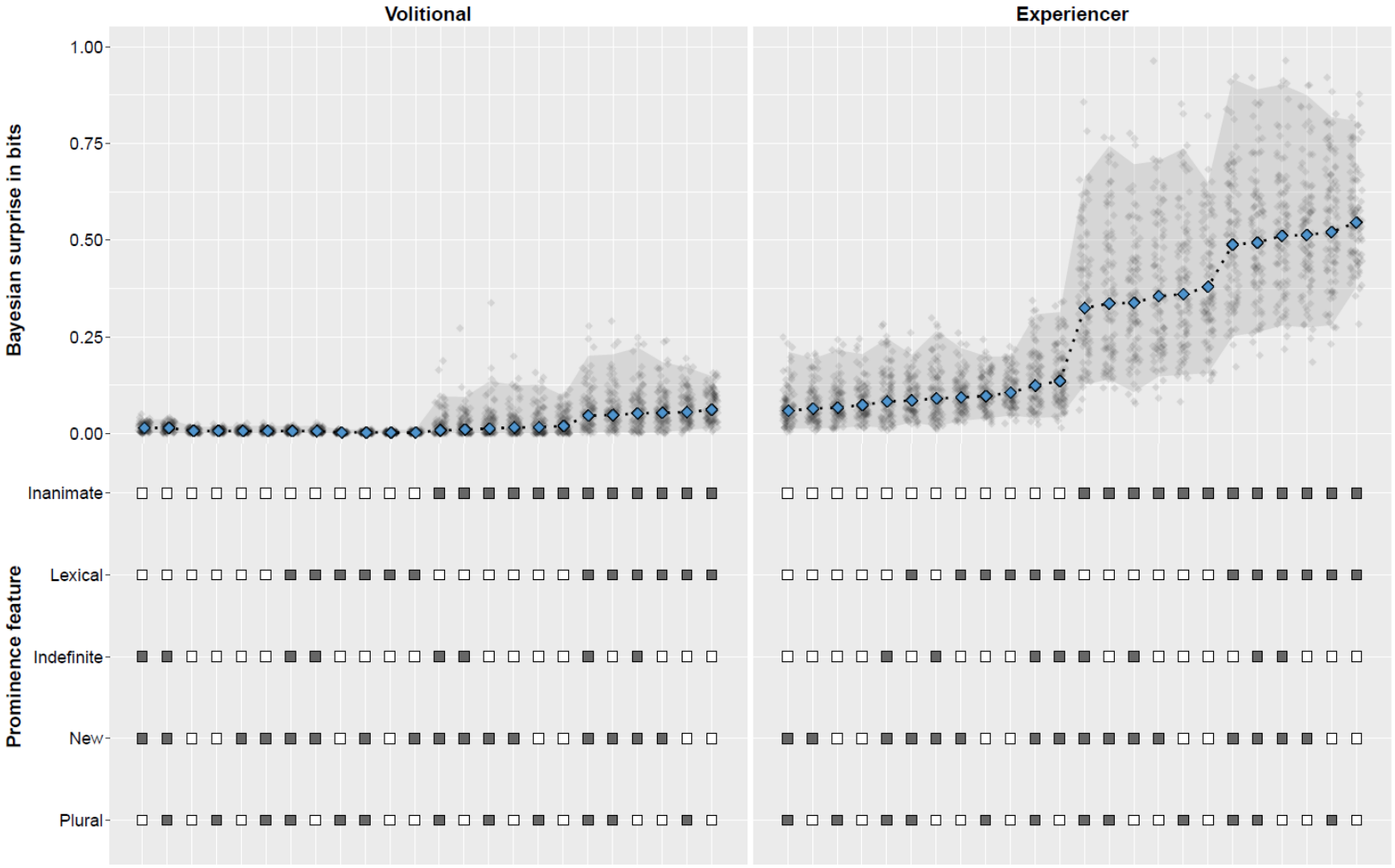
Stockholms universitet

Bayesian surprise at NP1 (3rd pers.)



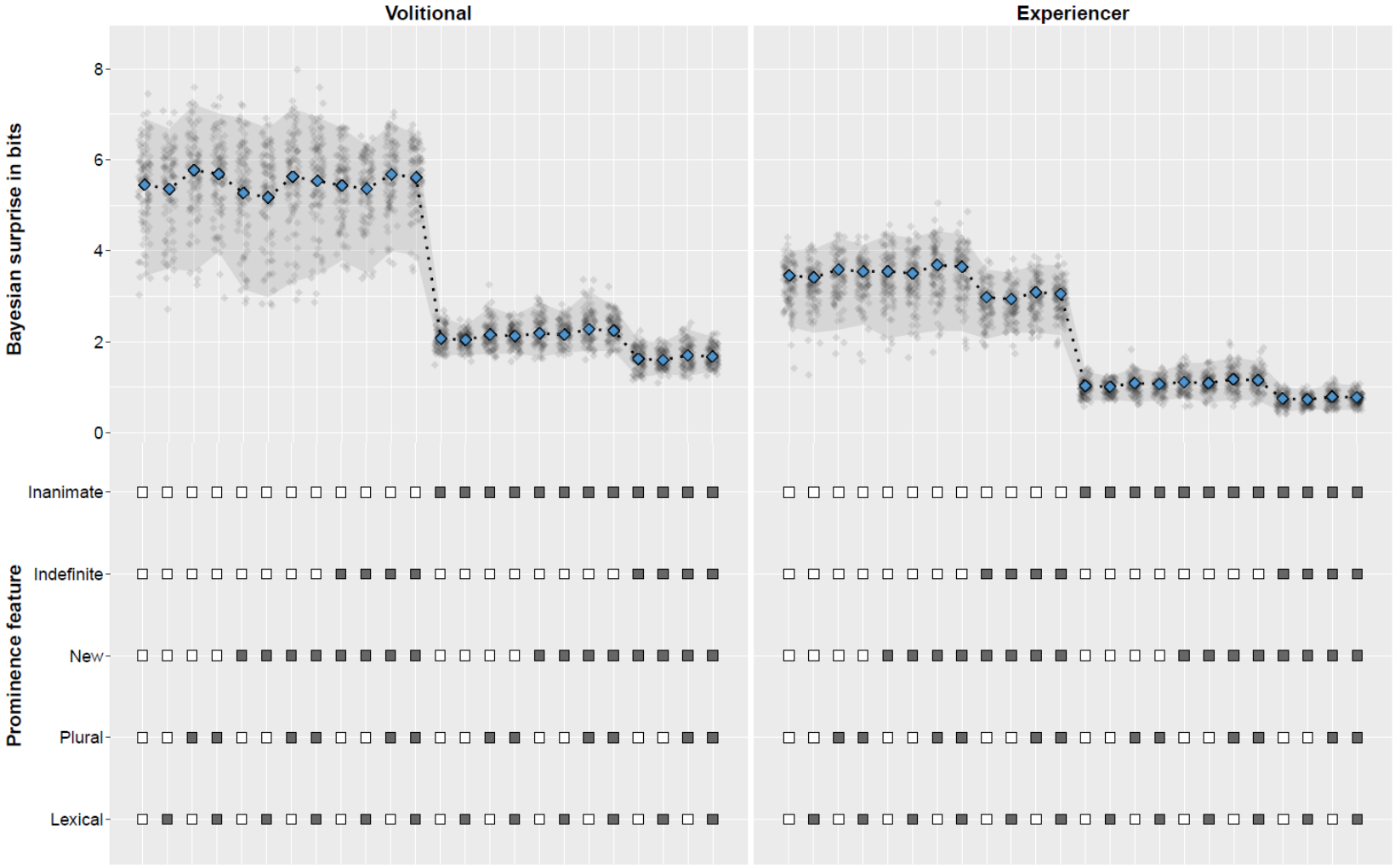
Stockholms universitet

Bayesian surprise at verb (3rd pers. int. NP)



Stockholms universitet

Bayesian surprise at NP2 (1st pers. subj. pronoun)



Stockholms universitet

Expectation-based modeling of incremental GF assignment

- Expectation-based incremental GF assignment drawing upon the distribution of AICs can be modeled in terms of Bayesian surprise:
  - predicts baseline assumption of the initial NP-as-subject (in line with e.g. Hörberg et al. 2013)
  - predicts prominence-by-verb class interaction effects on processing costs during incremental GF assignment

Experimental test of model predictions

Stockholms universitet

Experimental test of model predictions

- Testing the strongest model predictions:
  - baseline assumption of the initial NP-as-subject
  - animacy × verb class interaction
- Self-paced reading
- Reading times assumed to reflect processing

Self paced reading

####### ######## ### ##### ### # ########

Self paced reading

Bollen ######## ### ##### ### # ########

ball.the

Self paced reading

####### **sparkar** ### ##### ### # ########

kick

Self paced reading

```
####### ######## jag #### ### # ########
                |
```


Stockholms universitet

Self paced reading

####### ######## ### **mitt** ### # ########

middle

Self paced reading

####### ######## ### #### **upp** # ########

up

Self paced reading

####### ######## ### ##### ### i ########

in

Self paced reading

####### ######## ### ##### ### # **krysset**

top.corner.the

Stockholms universitet

Self paced reading

####### ######## ### ##### ### # ########

”The ball, I kick it right up into the top corner”

- Dependent variable: time latency between button presses
- Analyses done on region RTs rather than word RTs
- Task: Comprehension question following each sentence

Sparkar jag bollen mitt upp i krysset?

”Do I kick the ball right up into the top corner?”

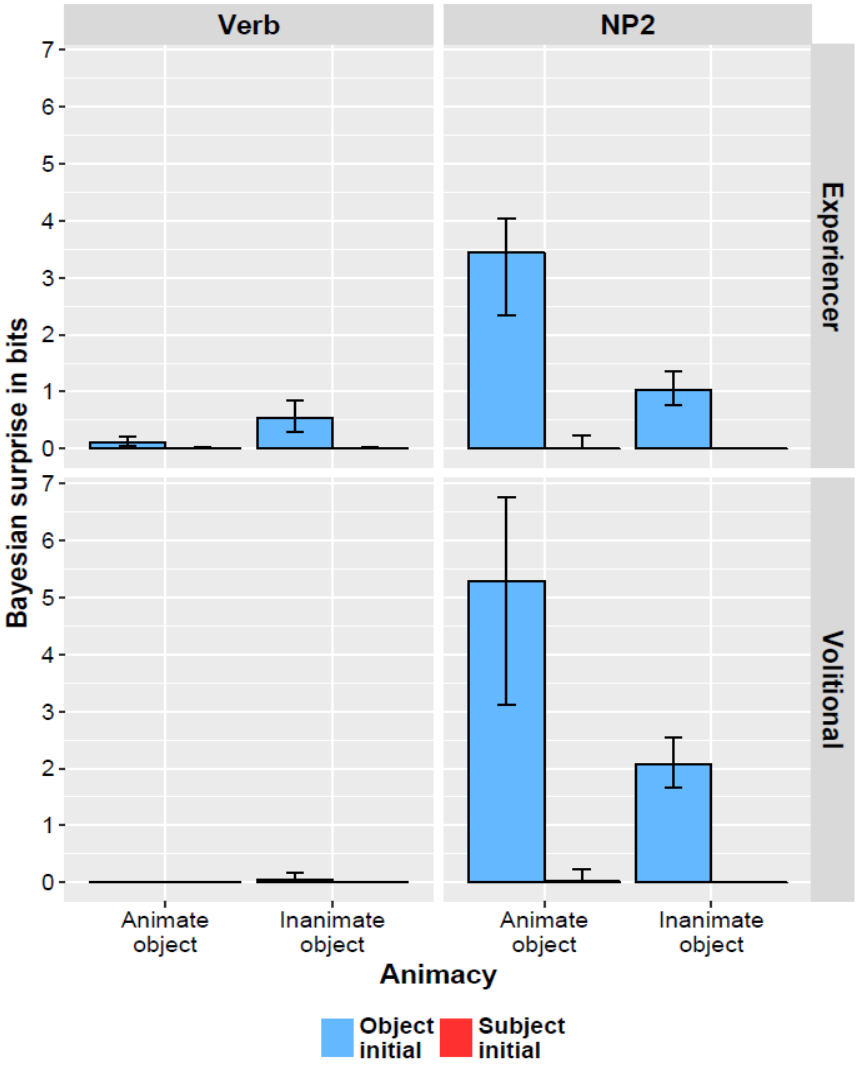
Experimental sentences

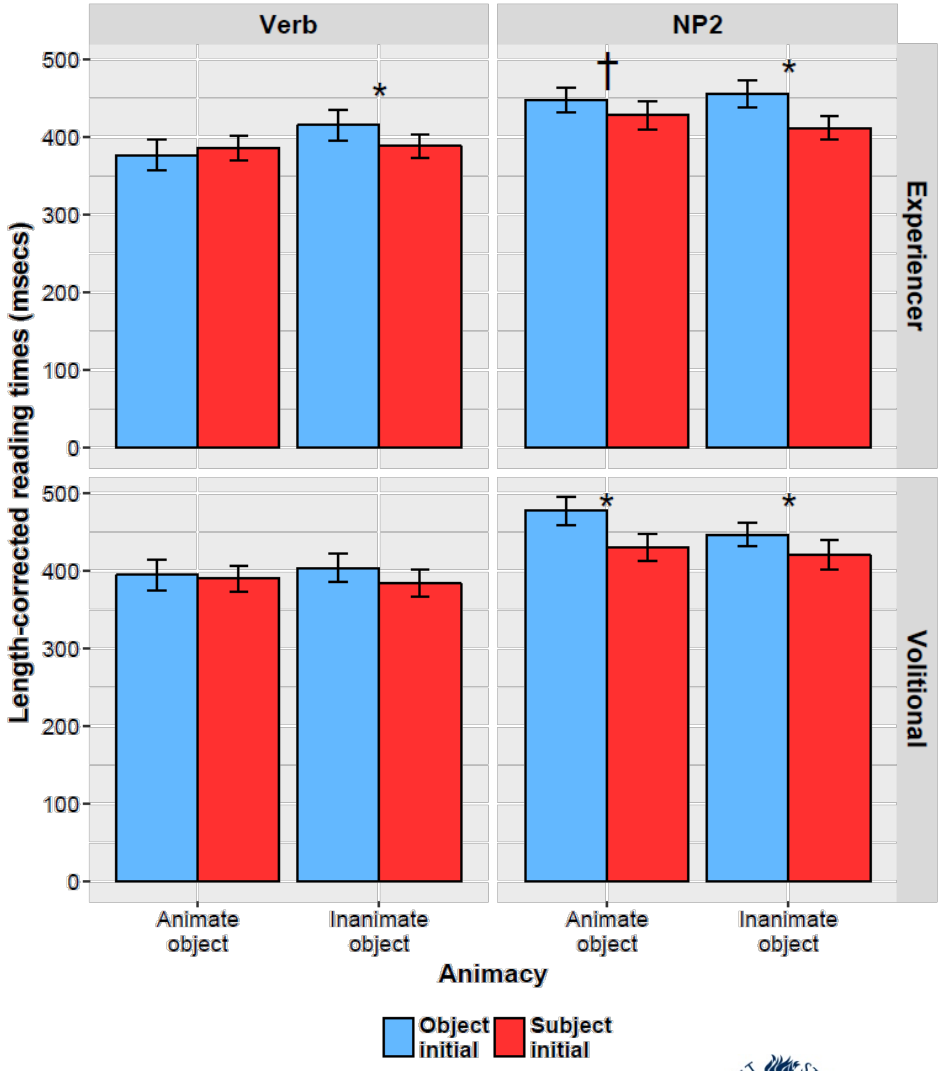
WO	Verb class	Obj. animacy	Example
OVS	Volitional	Inanimate	Bollen sparkar jag mitt upp i krysset ”The ball, I kick it right up into the top corner”
		Animate	Killen sparkar jag mitt på smallbenet ”The guy, I kick him in the middle of the shin”
	Experiencer	Inanimate	Bollen glömmer jag mitt på fotbollsplanen ”The ball, I forget it in the milddle of the football field”
		Animate	Killen glömmer jag sent på kvällen ”The guy, I forget him late at night”
SVO	Volitional	Inanimate	Jag sparkar bollen mitt upp i krysset ”I kick the ball right up into the top corner”
		Animate	Jag sparkar killen mitt på smallbenet ”I kick the guy in the middle of the shin”
	Experiencer	Inanimate	Jag glömmer bollen mitt på fotbollsplanen ”I forget the ball in the milddle of the football field”
		Animate	Jag glömmer killen sent på kvällen ”I forget the guy late at night”

Regions

- e.g. /Bollen sparkar $_{\text{reg.1}}$ / jag mitt $_{\text{reg.2}}$ / upp i $_{\text{reg.3}}$ / krysset $_{\text{FW}}$ /
- Region 1 RTs correspond to verb surprisal, region 2 RTs to NP2 surprisal

Results – RT differences vs predictions



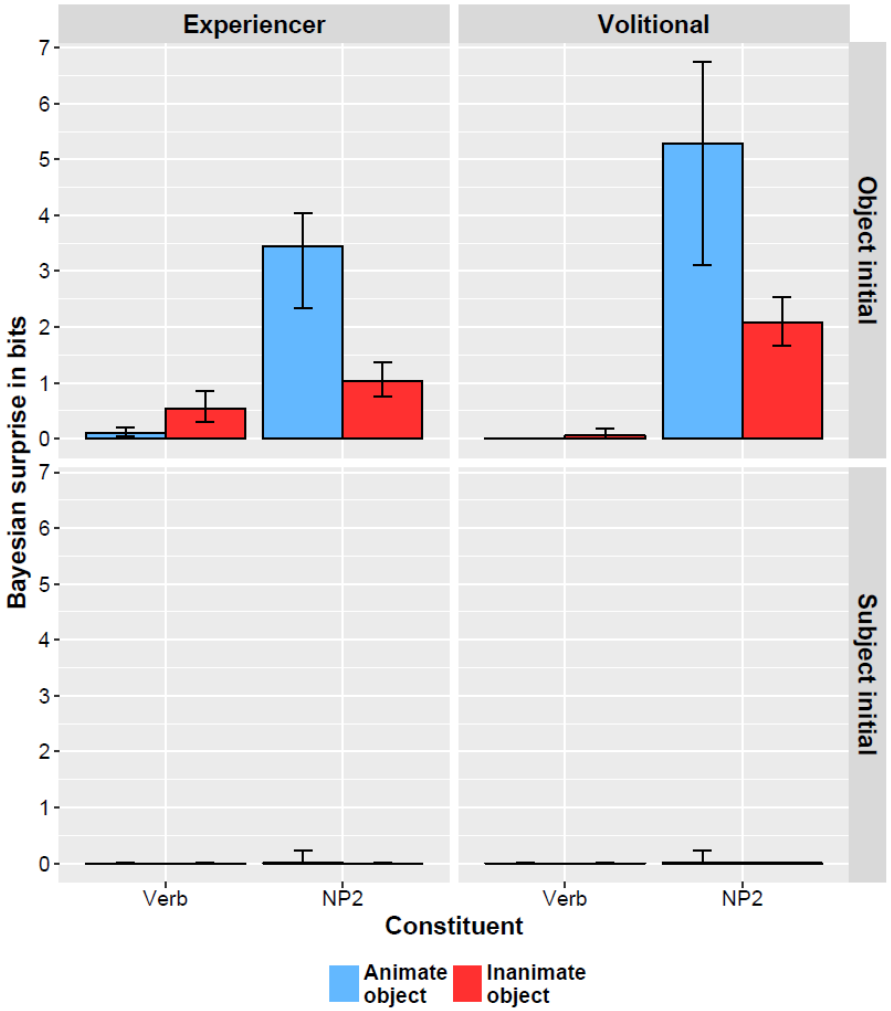


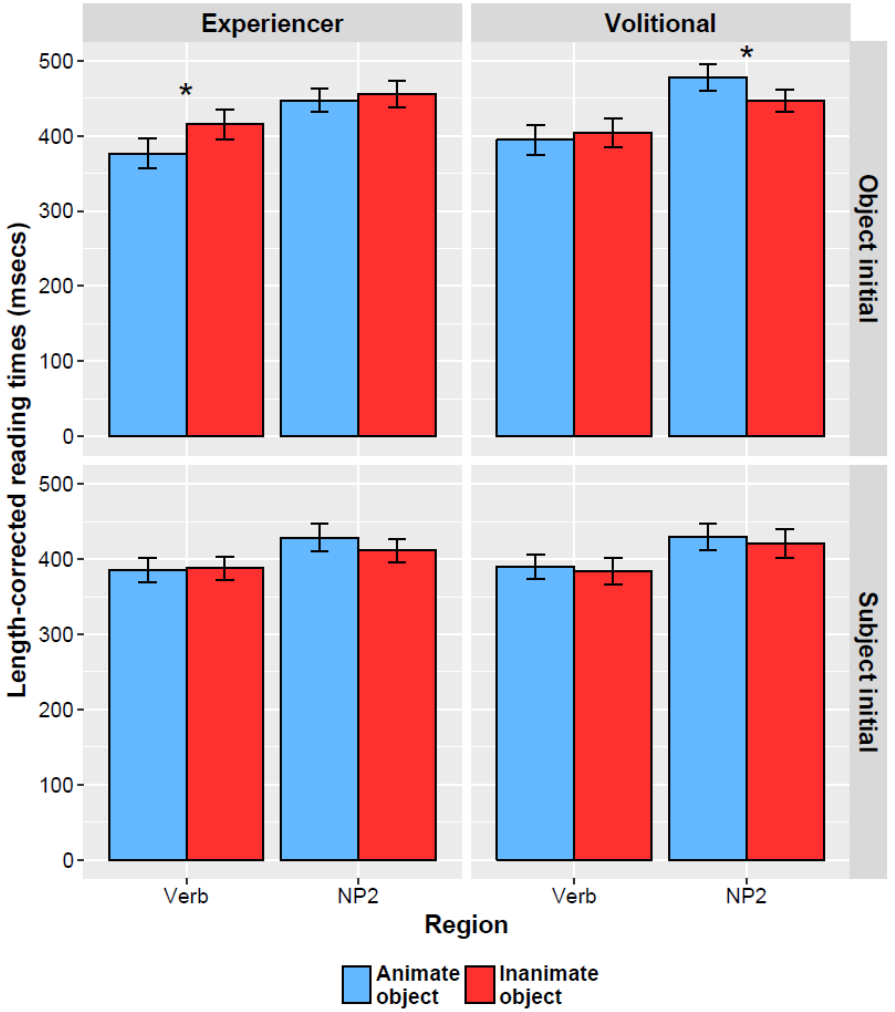
Subject-inital sentences preferred:

- faster RTs in SVO sentences vs. OVS sentences

Stockholms universitet

Results – RT differences vs predictions





- Faster verb RTs for animate vs. inanimate objects in experiencer verb sentences
- Faster NP2 RTs for inanimate vs. animate objects in volitional verb sentences

Stockholms universitet

Summary of results

- Confirms model predictions of OS vs. SO-differences - a baseline assumption of the initial NP-as-subject
- *At large* confirms animacy × verb class interaction effects on incremental GF assignment
  - no animacy effect on NP2 RTs in experiencer verb sentences
  - model might underestimate experiencer × egophoricity interaction

Overall conclusions

- Writers adapt their productions of transitive events in order to avoid ambiguities
- Processing costs during the comprehension of transitive sentences overall in line with expectation-based model of incremental GF assignment
- Converging evidence for the expectation-based perspective on language processing

Thanks for your attention!

Stockholms universitet

References

Dahl, Ö. (2000). Egophoricity in discourse and syntax. *Functions of Language*, 7(1),37-77.

Durie, M. (1987). Grammatical Relations in Acehnese. *Studies in Language*, 11(2), 365–399.

Foley, W. A. (2011). A typology of information packaging in the clause. In T. Shopen (Ed.), *Language Typology and Syntactic Description: Vol. 1: Clause Structure* (pp. 362–446).

Hörberg, T., Kallioinen, P. & Koptjevskaja-Tamm, M. (2013). The neurophysiological correlate to grammatical function reanalysis in Swedish. *Language and Cognitive Processes*, 28 (3), 388–416.

Hörberg, T. (2016). *Probabilistic and Prominence-driven Incremental Argument Interpretation in Swedish*. Doctoral Thesis, Department of Linguistics, Stockholm University.

Hörberg, T. (2018). Functional motivations behind direct object fronting in written Swedish: A corpus-distributional account. *Glossa: a journal of general linguistics* 3(1): 81. 1–36.

Kurumada, C. & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language,* 83, 152–178.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4.

Scott, G. (1986). On Ergativity in Fore and Other Papuan Languages. Pacific Linguistics A, 70, 167–175.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical Categories in Australian Languages* (pp. 112-171). New Jersey: Humanities Press.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019). Expectation-based Comprehension: Modeling the Interaction of World Knowledge and Linguistic Experience. *Discourse Processes*, 56(3), 229–255.

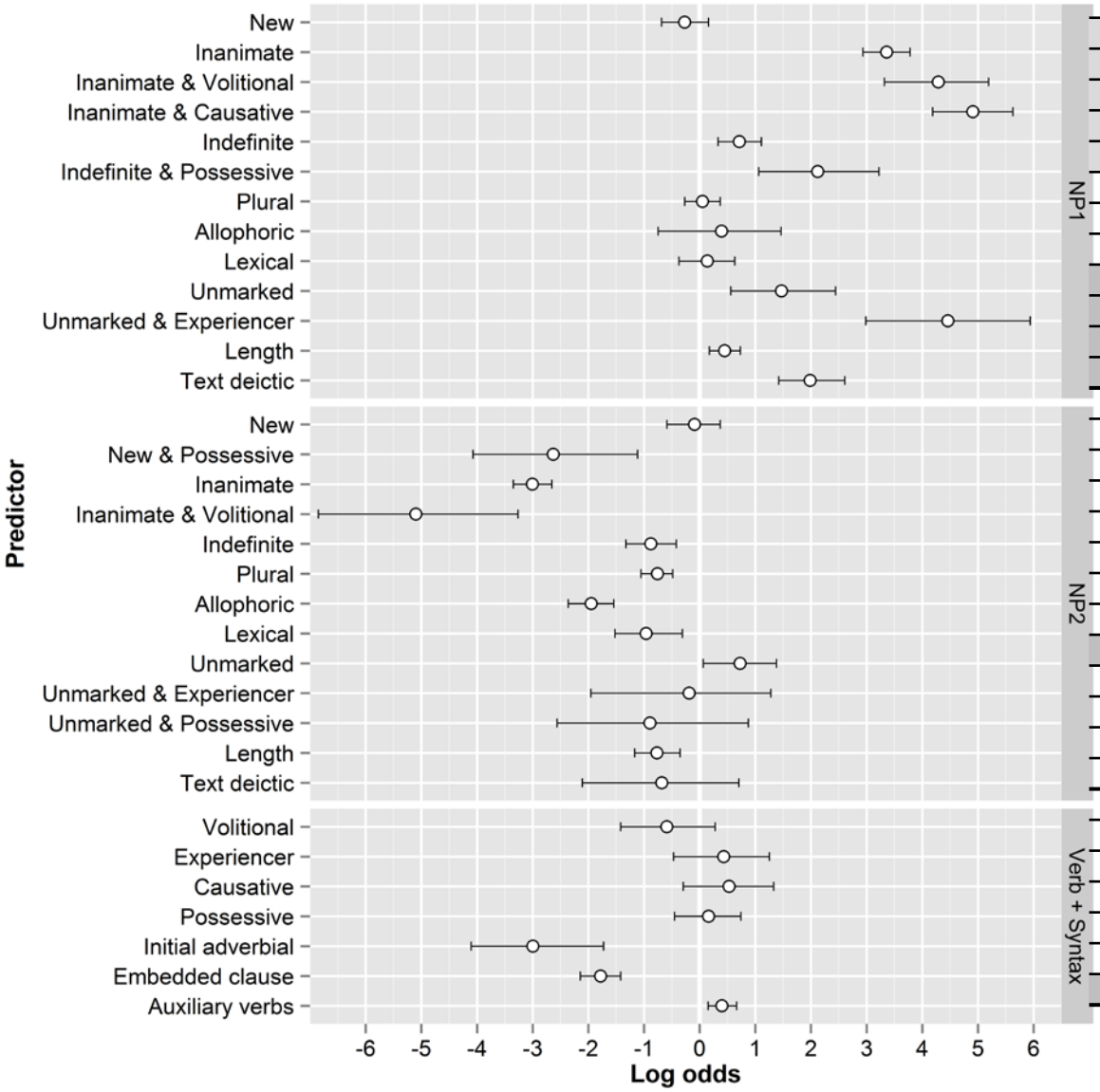
Van Valin, R. D. J. (2005). *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Zaharlick, A. M. (1982). Tanoan studies: passive sentences in Picuris. *State University Working Papers in Linguistics*, 26, 34–48.

Adaptation in written Swedish

Word Order	Property	NP1 > NP2	NP1 = NP2	NP1 < NP2	*p*
SVO	Givenness	59.5%	27.2%	13.3%	<.001
	Definiteness	41.2%	50.2%	8.6%	<.001
	Pronominality	49.0%	45.3%	5.7%	<.001
	Animacy	58.2%	38.6%	3.3%	<.001
	Case marking/Person	38.1%	57.4%	4.5%	<.001
OVS	Givenness	24.2%	36.0%	39.8%	<.001
	Definiteness	13.1%	66.4%	20.5%	<.001
	Pronominality	12.1%	45.2%	42.7%	<.001
	Animacy	0.3%	16.6%	83.1%	<.001
	Case marking/Person	0.1%	33.5%	66.4%	<.001

Model estimates of AIC strengths



	O.R. OVS	Lower	Upper	O.R. SVO	Lower	Upper
New	0,77	0,49	1,22	1,31	0,82	2,02
Inanimate	**28,84**	19,24	43,14	0,03	0,02	0,05
Inanimate & Volitional	**72,94**	31,92	175,31	0,01	0,01	0,03
Inanimate & Causative	**135,53**	64,26	270,90	0,01	0,00	0,02
Indefinite	**2,05**	1,38	3,08	0,49	0,32	0,73
Indefinite & Possessive	**8,35**	2,85	22,98	0,12	0,04	0,35
Plural	**1,05**	0,76	1,42	0,95	0,70	1,32
Allophoric	1,48	0,47	4,21	0,67	0,24	2,11
Lexical	1,15	0,72	1,95	0,87	0,51	1,39
Unmarked	**4,36**	1,79	10,85	0,23	0,09	0,56
Unmarked & Experiencer	**86,63**	22,86	333,39	0,01	0,00	0,04
Length	**1,57**	1,18	2,08	0,64	0,48	0,85
Text deictic	**7,28**	4,25	12,60	0,14	0,08	0,24
New	0,91	0,53	1,53	1,09	0,66	1,87
New & Possessive	0,07	0,01	0,31	**13,89**	3,25	66,94
Inanimate	0,05	0,03	0,07	**20,22**	14,11	28,89
Inanimate & Volitional	0,01	0,00	0,04	**163,81**	26,41	1190,24
Indefinite	0,42	0,26	0,67	**2,41**	1,50	3,82
Plural	0,47	0,35	0,65	**2,13**	1,54	2,84
Allophoric	0,14	0,10	0,21	**7,02**	4,74	10,23
Lexical	0,38	0,21	0,67	**2,61**	1,49	4,75
Unmarked	**2,07**	1,10	4,09	0,48	0,24	0,91
Unmarked & Experiencer	0,83	0,16	4,38	1,21	0,23	6,08
Unmarked & Possessive	0,41	0,09	2,26	2,44	0,44	11,43
Length	0,47	0,32	0,73	**2,15**	1,38	3,17
Text deictic	0,51	0,12	2,18	1,97	0,46	8,23
Volitional	0,55	0,23	1,26	1,80	0,79	4,37
Experiencer	1,54	0,63	3,84	0,65	0,26	1,59
Causative	1,70	0,74	3,71	0,59	0,27	1,35
Possessive	1,17	0,67	2,18	0,85	0,46	1,50
Initial adverbial	0,05	0,01	0,17	**19,99**	5,73	71,19
Embedded clause	0,17	0,12	0,25	**5,93**	4,05	8,61
Auxiliary verbs	**1,49**	1,14	1,93	0,67	0,52	0,88

Stockholms universitet