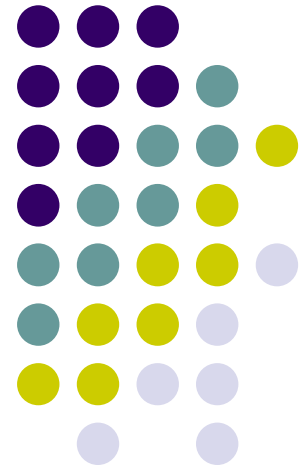


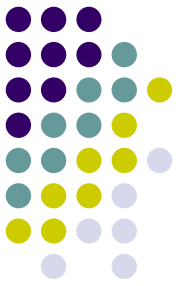
A Processing Account of Relativizer Omission

ICLC 10

Krakow
July 15th-20th, 2007



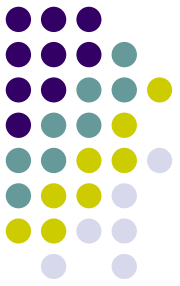
Daniel Wiechmann
Friedrich-Schiller-Universität Jena



Structure of the talk

- Introduction
- Main Hypothesis
- Data
- Method
- Results
- Discussion

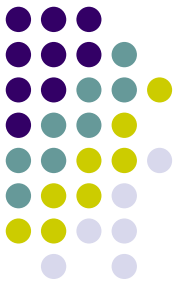




Structure of the talk

- Introduction
 - Phenomenon: Relativizer omission in Non-Subj RC
 - PLUS: *Super extra feature* (free too!):
Stance towards description and explanation
- Main Hypothesis
- Data
- Method
- Results
- Discussion

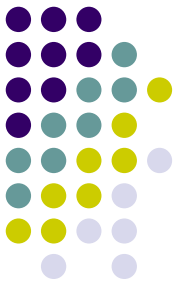




Structure of the talk

- ❑ Introduction
- ❑ Main Hypothesis
 - 'Complexity **A**nd **R**esource **L**imitation' hypothesis (CARL)
- ❑ Data
- ❑ Method
- ❑ Results
- ❑ Discussion

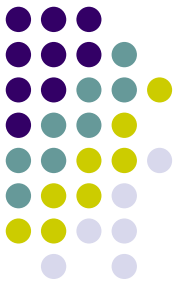




Structure of the talk

- ❑ Introduction
- ❑ Main Hypothesis
- ❑ Data
 - 200 RCC characterized as 30+ dimensional vectors
- ❑ Method
- ❑ Results
- ❑ Discussion

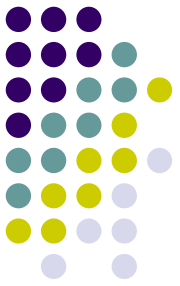




Structure of the talk

- Introduction
- Main Hypothesis
- Data
- Method
 - Discrete choice model: Logistic regression
- Results
- Discussion

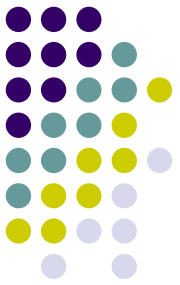




Structure of the talk

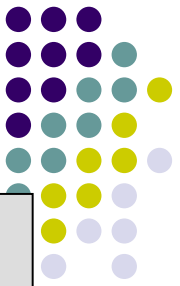
- Introduction
- Main Hypothesis
- Data
- Method
- Results
 - direct competitor: Wasow & Jaeger 2005+
- Discussion





Phenomenon





English relative clause constructions (RCC)

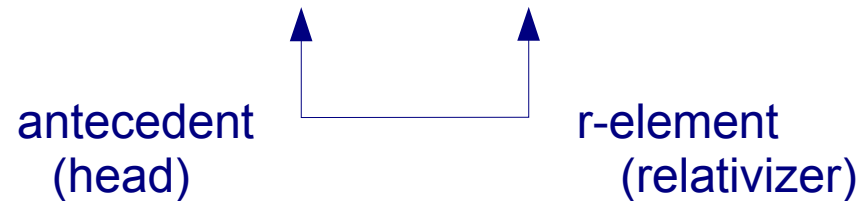
RCC minimally consist of two clausal elements:

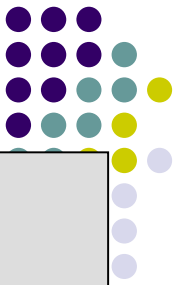
- a main clause and
- a subordinate clause that modifies some argument of the main clause (=RC)

RCs are usually introduced by an **element r (or relativizer)**, which

- is anaphorically related to the modified nominal (= **head** of RC)

Peter hates _{NP} [the **guy** _{RC} [who stole his car]].



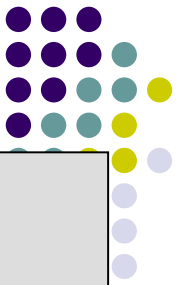


R-element omission in non-subject relative clauses

• Object RCs

- RCC_{NP} [*The man* $RC_{r\text{-element}}$ [*that|who*] *you met* ___ *yesterday*]] VP [*is* *JFK*]





R-element omission in non-subject relative clauses

• Object RCs

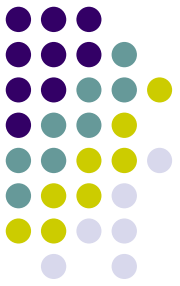
- ${}_{RCC-NP}$ [*The man* ${}_{RC}$ ${}_{r-element}$ [\emptyset] *you met* __ *yesterday*]]

${}_{VP}$ [*is*]]

• Subject RCs

- ${}_{RCC-NP}$ [*The man* ${}_{RC}$ ${}_{r-element}$ [*that|who*] __ *met you yesterday*] *VP*]]
- ${}_{RCC-NP}$ [**The man* ${}_{RC}$ ${}_{r-element}$ [\emptyset] __ *met you yesterday*] *VP*]]

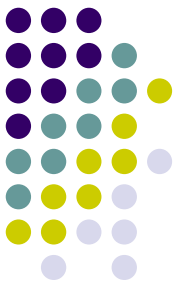




Guiding questions

- **Why** do people sometimes omit|produce non-obligatory elements?
- What are the **constraints** on relativizer omission?
 - What are the relative **weights** of these constraints?





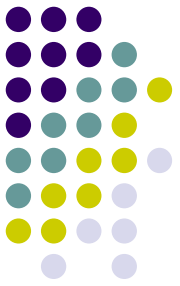
the complexity-and-resource-limitation hypothesis (CARL)

As-if approximation:

Language users are constantly looking for the most cost-efficient way to get across their communicative intentions.

But there are too many **dubious ontological commitments** here (e.g. the existence of intentions; beliefs, desires, and other **mental states**)





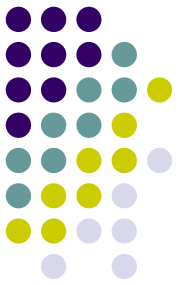
the complexity-and-resource-limitation hypothesis (CARL)

We might rather say that...

...there are limits to human symbolic processing and the **development and conventionalization of linguistic forms are functions of an optimization process geared to save energy.**

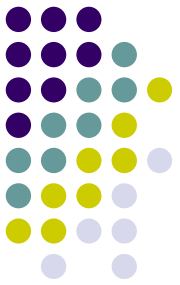


Short
version
#Slide 24



description & explanation





Extra content_{colloq} :

What do we expect **science** to do?

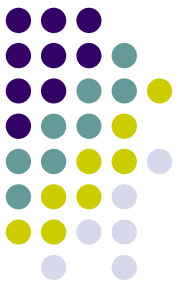
explore

predict

describe

explain



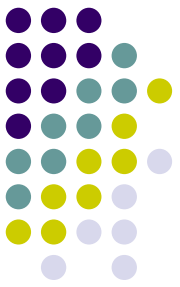


description

empiricist view

- All evidence must be empirical, observable, measurable
 - ➔ *find adequate operationalization*
 - *process of defining a concept as operations that will measure the concept through specific observations*

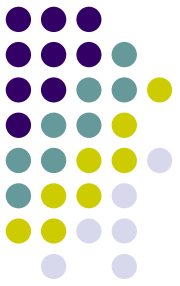




levels of description

- **computational level**
 - (function to be computed)
- **algorithmic level**
(algorithm which computes the function)
- **implementational level**
(hardware in which algorithm is realised)

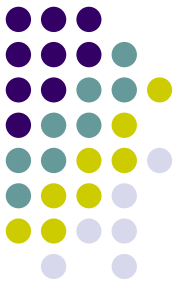




levels of description

- **computational level**
 - (functions which are involved in production and comprehension -> competence grammar)
- **algorithmic level**
 - (algorithm which computes these functions -> e.g. left-corner parser)
- **implementational level**
 - (hardware in which algorithm is realised -> brain or central nervous system)



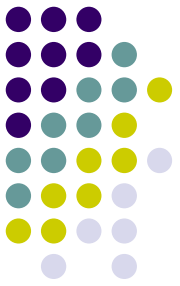


levels of description

- **computational level**
- **algorithmic level**
- **implementational level**

**job of cognitive
neuropsychologists and
cognitive neuroscientists**



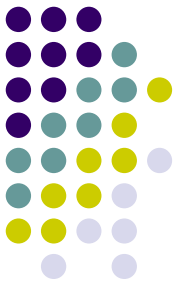


levels of description

- **computational level**
- **algorithmic level**
- **implementational level**

**our job as linguists and
psychologists of language**



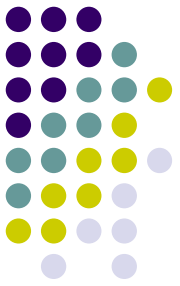


explanation

empiricist view

- *Explanation is just not part of the science*
- *Science's sole task is to describe*
 - ➔ *No room for 'inferences to the best explanation'*

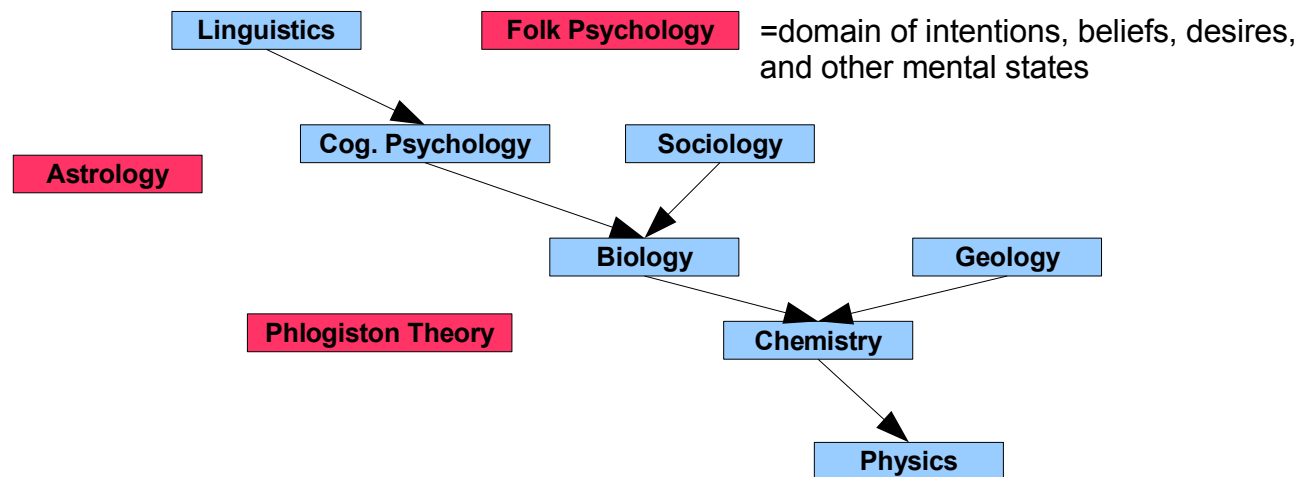




explanation

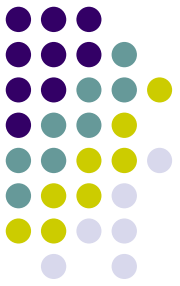
covering law view

- Explaining a phenomenon is to reduce the description of the governing theory to a more fundamental theory*



Irreducible theories (red) are isolated and need to be replaced



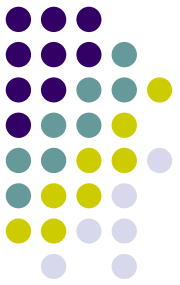


explanation

covering law view

- Explanations of linguistic events reduce the observed phenomenon to mechanisms of the processing system
 - ...and this leads to CARL





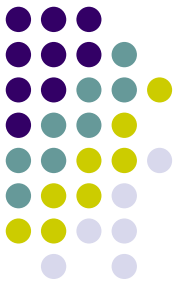
CARL-hypothesis

According to CARL the production of r-element is a function of the

- a) predictability of RC
- b) complexity of RCC
- c) memory load associated with RCC

that facilitates processing and/or signals difficulty





Interim summary

- Relativizer omission is to be explained by reducing the phenomenon to mechanisms of language processing
- The relevant mechanisms draw on various quantities (i.e. informational sources) to produce hypotheses concerning likely continuations
- These quantities figure in the operationalization of complexity, predictability and memory load

target level of analysis:

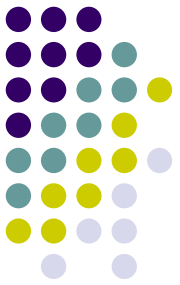
computational level

aim:

development of **statistical model** that
best **describes** the observed data
can **predict** new data

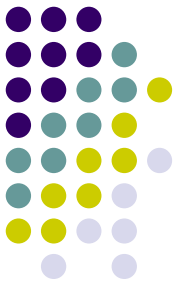
(prediction \equiv explanation)





Data





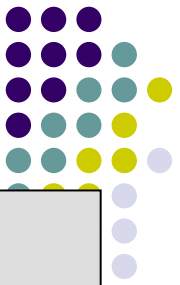
SAMPLE:

ICE-GB/ spoken part/ dialogue/ private/
direct conversations (=185193 words)

	Non-SubjRC	that (covert)	who	which	other R
N with R	1009	460	170	282	97
N without R	651	651	na	na	na
Free relatives	845	irrelevant	irrelevant	irrelevant	irrelevant
TOTAL	2505				

Should not be taken at face value:
manual inspection of data reveals
lots of false positives (e.g. complement clauses, fact-S constructions,...)





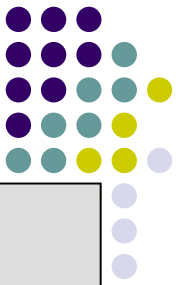
What?

current data set:

N=200 (100 overt + 100 covert non-subject RCs)

(proportions approximate -but do not exactly match- overall distribution of corpus sample)





What?

only declaratives of canonical form:

- no dislocations (e.g. locative/temp inversions, topicalization)
- no interrogatives
- no exclamatives
- no imperatives
- no 'fact-S' constructions
- no pseudo-embeddings

those details_i [that **you think|we feel**_s (_{-i} are most interesting)]

- no ambiguous attachments (PossP):

“**none**_i of the **students**_j in that **group**_k [that I was talking to _{i,j,k}] VP”

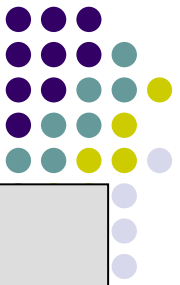
- no recursive RC-constructions

“I don't like _{NP}[people _{RC}[**who** underline _{NP}[books _{RC}[**that** you get from the university library]]]]
(coordinated RC have not been excluded though)

- no ellipsis within RC

“I would take the same view that Peter did ”





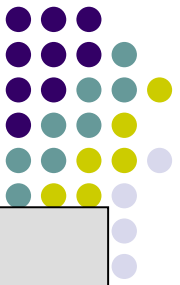
How?

Characterizing the data

Every relative clause construction_{token} is
characterized along 30+ dimensions

(but many of these are implied in others; e.g. abstract object → inanimate)





How?

variables of interest

variables

clause type (RC/EC)

internal & external role (head)

voice (RC(?)/EC)

theta head (RC/EC)

type of head/NP1

number head/NP1

determiner (head/NP1)

definiteness (head/NP1)

animacy (head/NP1)

animacy (head/NP1) scaled

scale of measurement

- factor [SV(C|A|O)...]

- factor [S|O|PN ...]

- boolean [active|passive]

- factor [Quirk et al. scheme]

- factor [bare NP|lexical NP...]

- boolean [singular|plural]

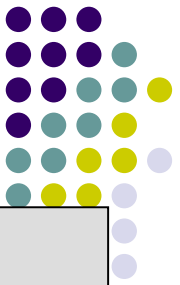
- factor [∅|a|the|some|...]

- boolean

- boolean [objective, not conceptual]

- factor [{animate(+,-),concrete(+/-)}]





How?

variables of interest (cont.)

variables

premodification (head/NP1)

uniqueness ADJ

contentfulness (head)

BNC frequency (head)

accessibility_{Ariel} (head/NP1)

perspective shifts

type of embedding

definiteness (head)

length of dependency domain

relativizer

scale of measurement

- boolean (y/n)

- boolean (y/n)

- boolean (low/not low)

- cont [count data]

- factor [degree of ~]

- factor [number of ~]

- boolean [center/right]

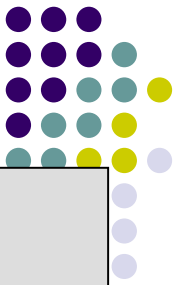
- boolean [y/n]

- cont [count in words]

- boolean [present/absent]



Note: By now the number of variables has grown larger; however not all variables should enter the analysis at the same time (more observations needed)



How?

coding problems & principles

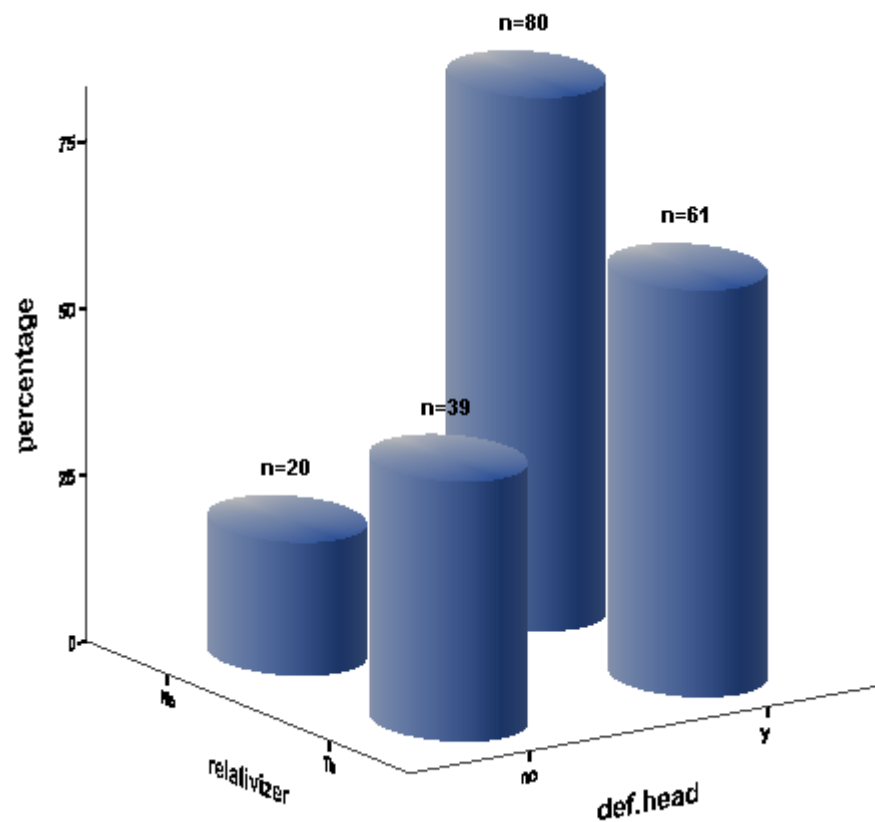
- 'when in doubt, leave it out'-maxim
 - ◆ drop data point if it is incomplete or ungrammatical
- avoid pragmatic inferences:
 - ◆ e.g. PossP ambiguities are sometimes easily resolved:
 - ◆ The **car** of the **man** that **has only three wheels** VP
 - ◆ A **car** of the **sixt fleet** that **was never famous for its quality** VP
- if expression is vague, go further down the hierarchy
 - ◆ *anything*: inanimate object -> concrete or abstract? decision: abstract
- theta roles have been assigned “literally”
 - ◆ “I felt close to **him**” --> *him* carries LOC role
- composite heads were considered modified
 - ◆ e.g. precision bombing -> head: bombing; premod:1

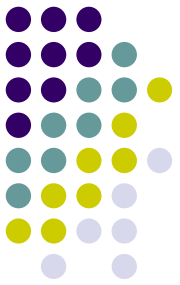




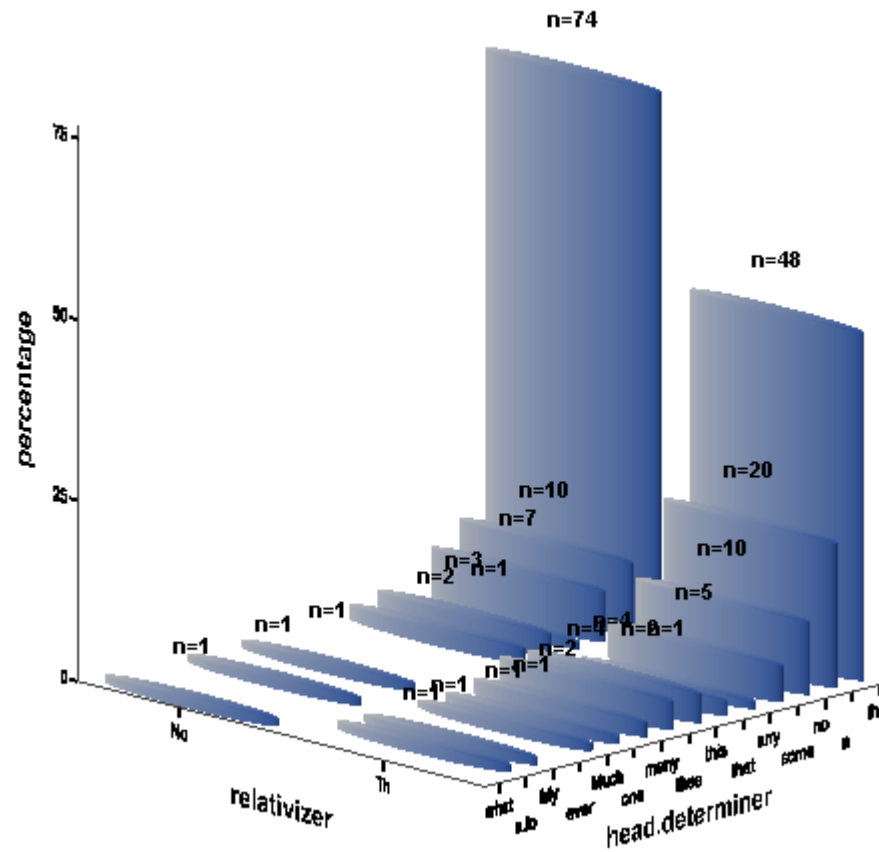
properties of the head: comparison

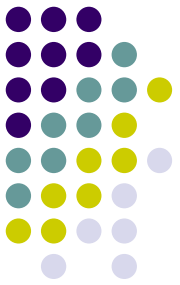
definiteness_{head}



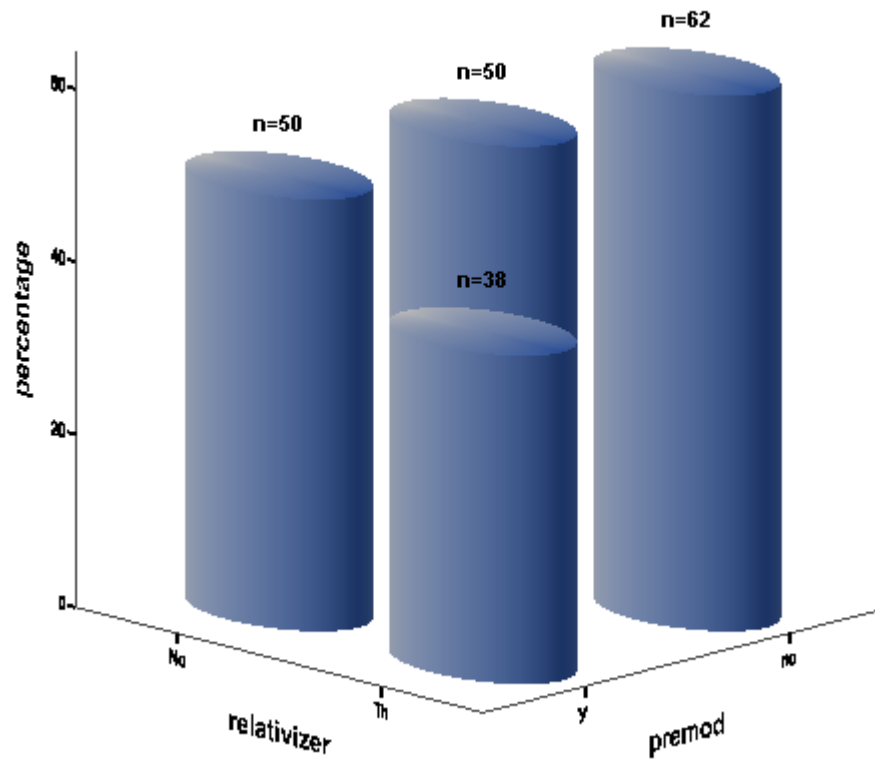


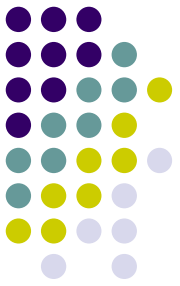
properties of the head: comparison head of DP incorporating head (fine-grained)



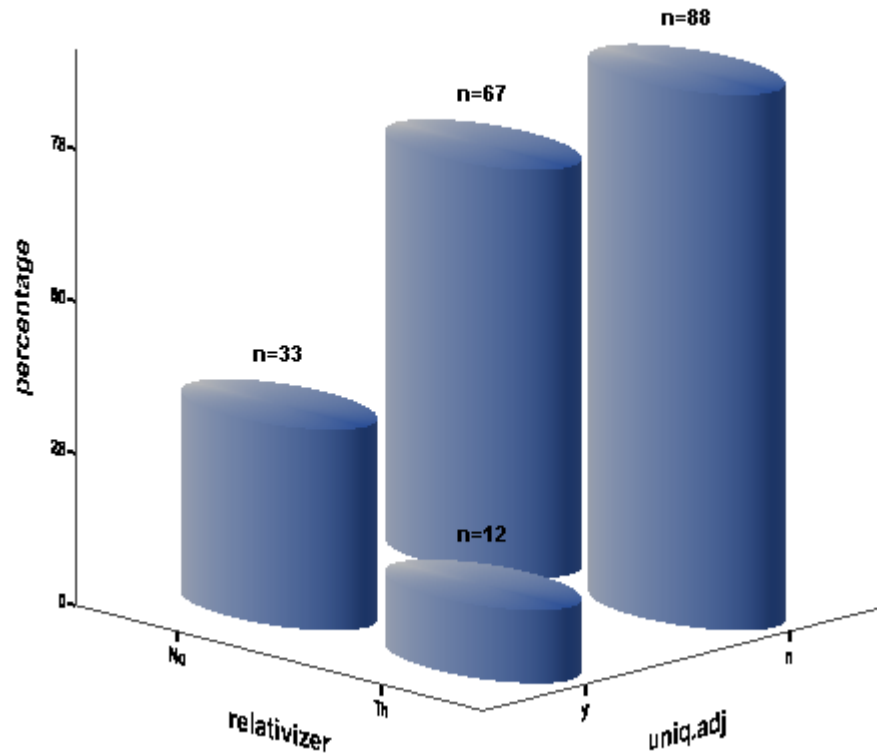


properties of the head: comparison additional premodification



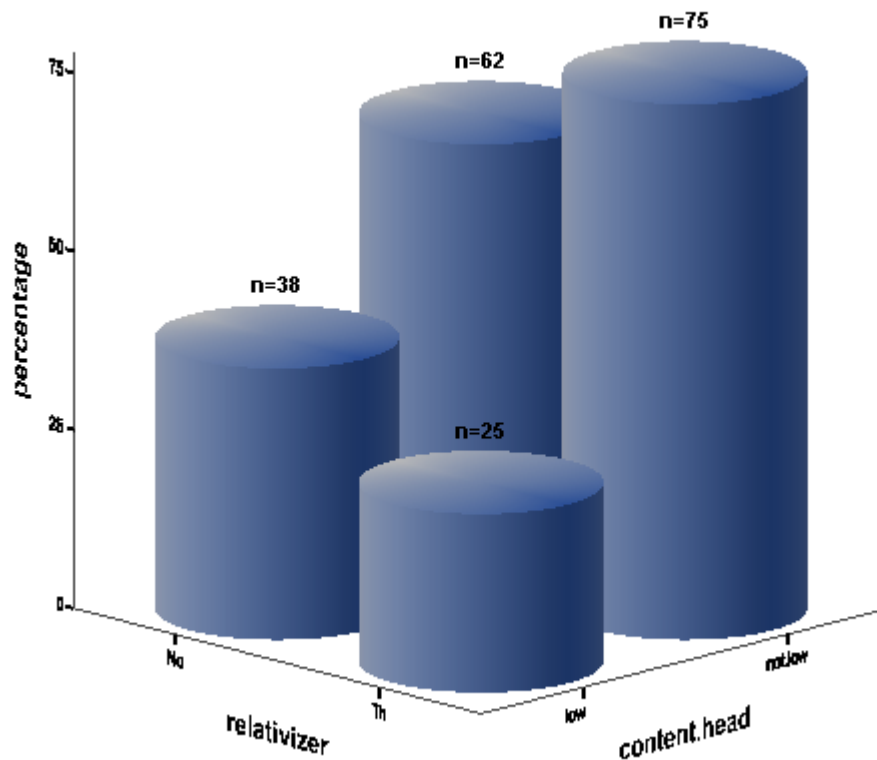


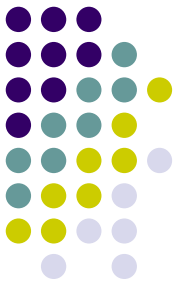
properties of the head: comparison uniqueness adjectives



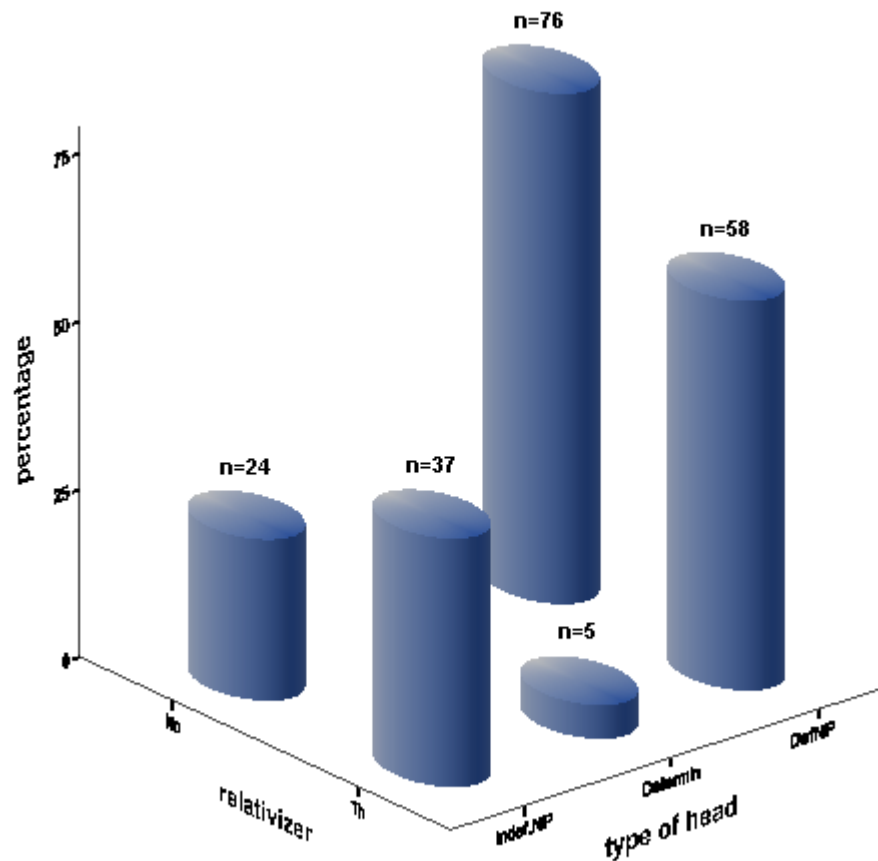


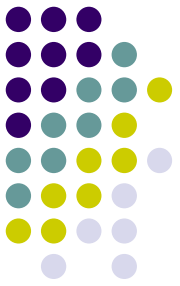
properties of the head contentfulness



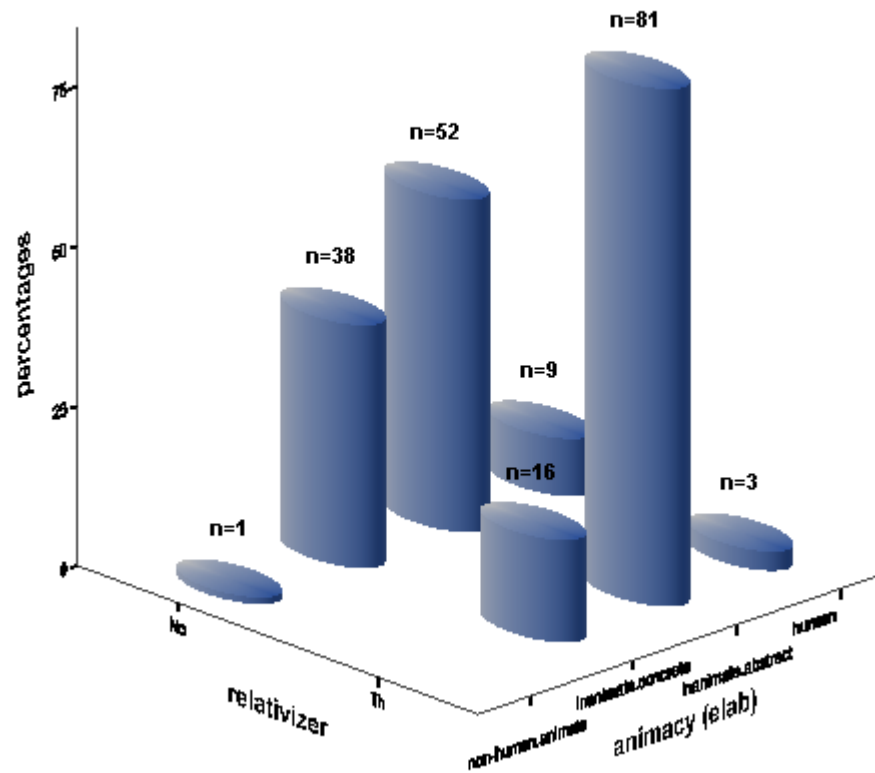


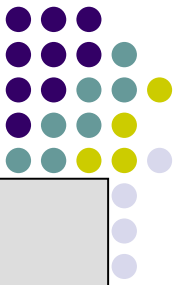
properties of the head: comparison type of head





properties of the head: comparison animacy(elaborated)_{head}





accessibility (Ariel 1990)

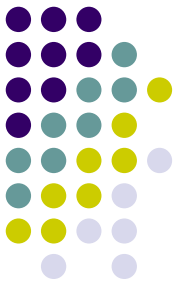
1st ps → 2nd ps → 3rd ps → Name_{prop} → def.NP → indef. NP → determinatives

Table 4.1 Initial Accessibility marking

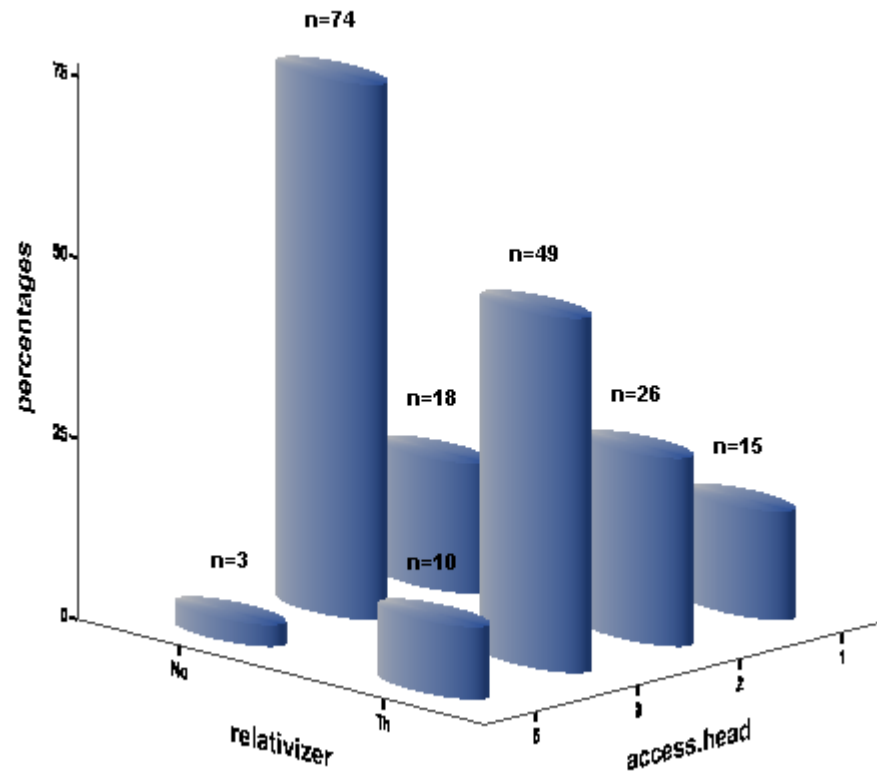
		Accessibility				
		Low ←				→ High
M E M O R Y	Long-term	KC: Joan Smith the president	Joan Smith	The president	Smith	Joan
		PC: This/that hat we bought last year.	That hat	This hat	I/You/That	This
	Short-term	LC: SHE	she	Hers	The former/ latter	Herself

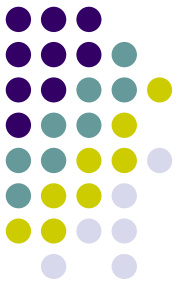
KC = General Knowledge Context, PC = Physical Context, LC = Linguistic Context





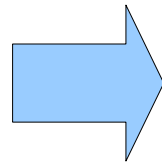
properties of the head: comparison accessibility_{head}





Semantic Roles: Quirk et al. 1985

aff(ected)
agent(ive)
attr(ribute)
 cog(nate)
event(ive)
 ext(ernal causer)
instr(ument)
 (prop) it
loc(ative)
 pos(itioner)
rec(ipient)
 result(ant)
temp(oral)



AGT > REC > EVENT > AFF > LOC/TEMP > INSTR/MAN

Possession:
She has a car

Attribution (non-human):
The show was interesting

Attribution:
She was happy

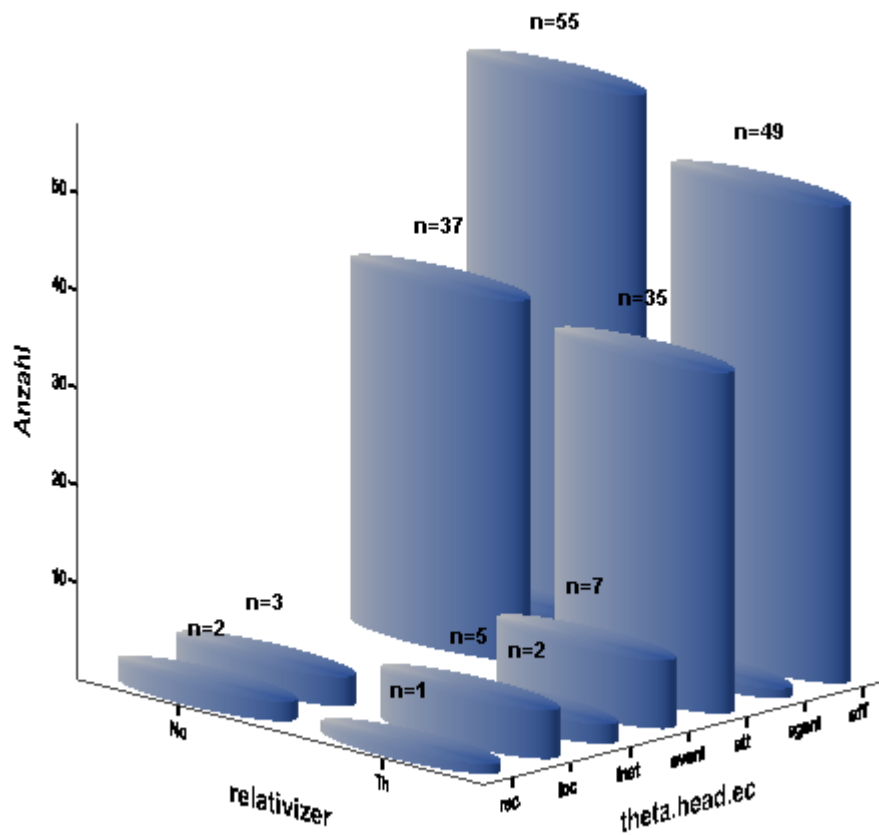
... > ATT

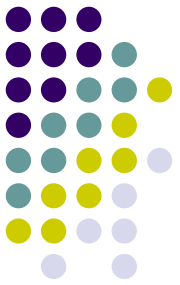
There BE NP
This BE NP
It BE NP





properties of the head theta role in embedding clause

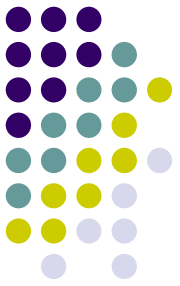




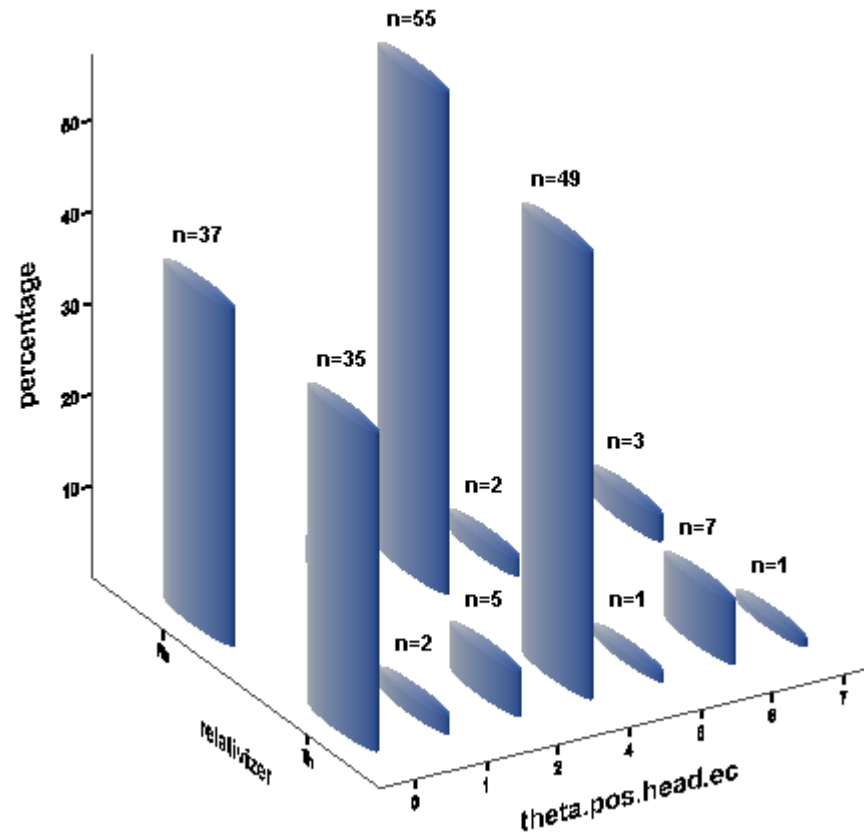
Thematic Role Hierarchies

Baker 1989	AG T > INST > THM > GOAL/LOC
Belletti & Ricci 1988	AG T > EXP > THM
Bresnan & Kanerva 1989	AG T > BEN > REC/EXP > INST > THM > LOC
Carrier-Duncan 1985	AG T > THM > GOAL/SOURCE/LOC
Fillmore 1968	AG T > INST > OBJ (G/LOC ranked before THM)
Givon 1984	AG T > BEN > REC/EXP > THM > LOC > INST
Grimshaw 1990	AG T > EXP > G/S/L > THM
Jachendoff 1972	AG T > G/S/L > THM
Jackendoff 1990	ACT > PAT/BEN > THM > G/S/L
Larson 1988	AG T > THM > G > Obj
Speas 1990	AG T > EXP > THM > G/S/L > MAN/TIME
Van Valin 1990	AG T > EFF > EXP > LOC > THM > PAT





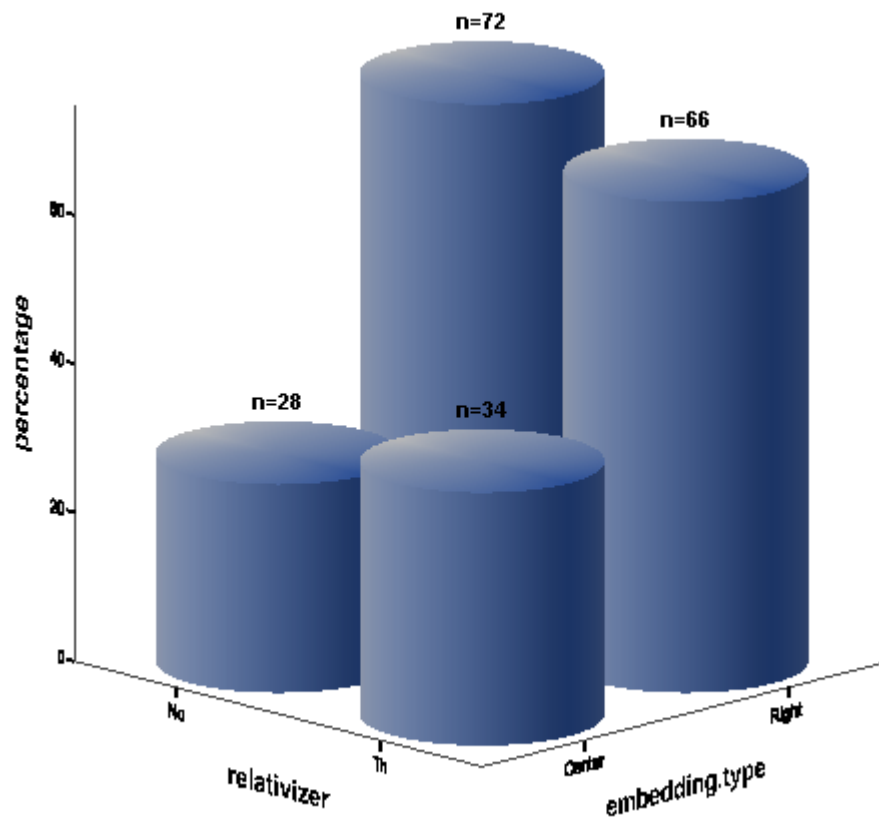
properties of head position_{theta hierarchy} in embedding clause

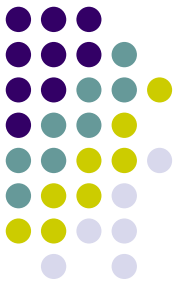




further comparisons

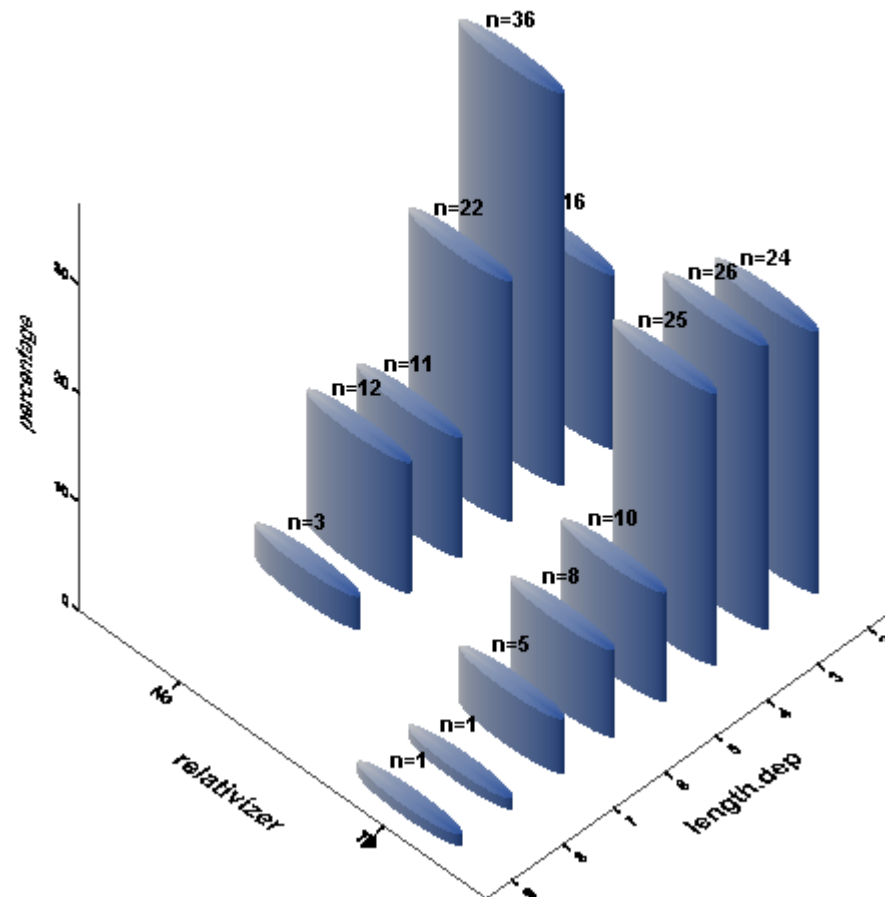
type of embedding

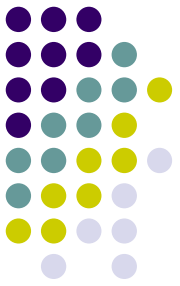




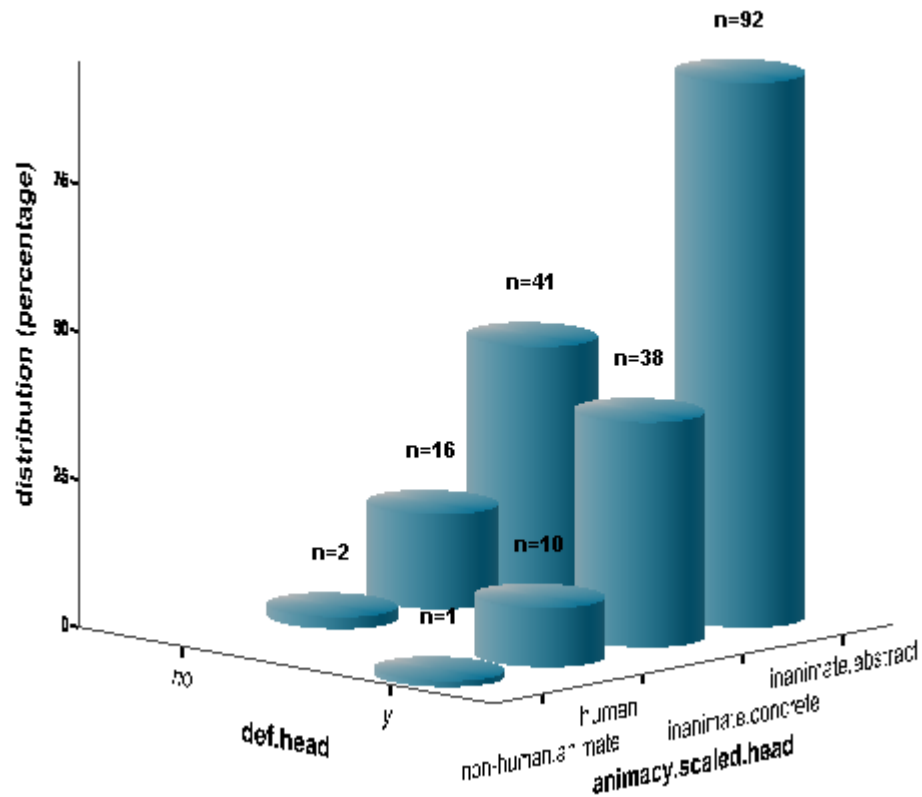
further comparisons

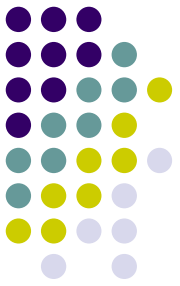
length of dependency domain (in words)



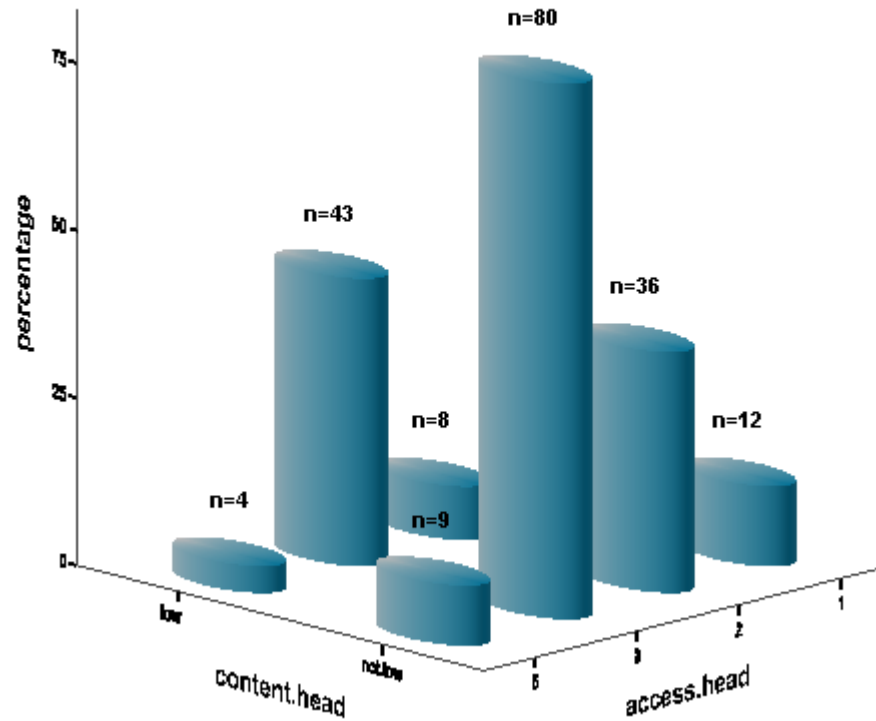


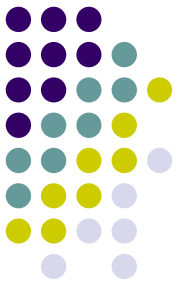
properties of the head animacy and definiteness





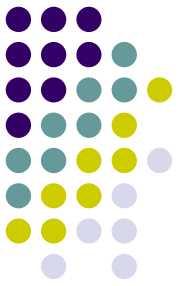
properties of the head accessibility and contentfulness





Method





Analytical Techniques

Data Mining: **K-optimal pattern discovery**

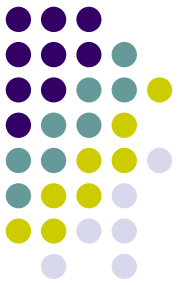
(maybe topic of my next talk)

Statistical Modelling: Binary Logistic Regression

- Frequent pattern discovery techniques find all patterns for which there are sufficient examples in the sample data (but arguably my data set is too small for that)
- In contrast, k-optimal pattern discovery techniques find the k patterns that optimize a user-specified measure of interest. The value k is also specified by the user.
- Software: Magnum Opus 4.1 (www.rulequest.com)

Webb, G. I. (1995). OPUS: An efficient admissible algorithm for unordered search. *Journal of Artificial Intelligence Research*, 3, 431-465.





binary logistic regression



What kind of statistical model?

- Regression

- Generalized linear model

- **Binary logistic regression model**

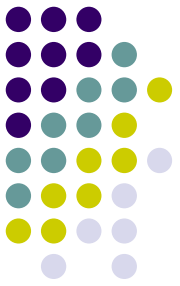
Why this one?

Glms are great when the response variable is:

- count data (expressed as proportions)
- binary
- variance is not constant (increases with mean)



www.daniel-wiechmann.net -> *presentations* for technical and conceptual details on logistic regression



binary logistic regression



Binary logistic regression is used to **predict** an **outcome** of dichotomous dependent variable from a set of explanatory variables.

-> Factor of interest, r-element omission, is modelled as a function of a number of explanatory variables

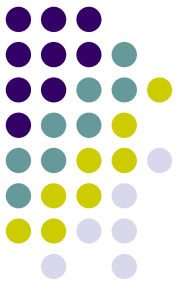


Properties:

- no delimiting distributional assumptions
- can deal with all types of explanatory variables
- number of required observed cases is relatively low (but of course dependent on number of explanatory variables)

(but of





binary logistic regression



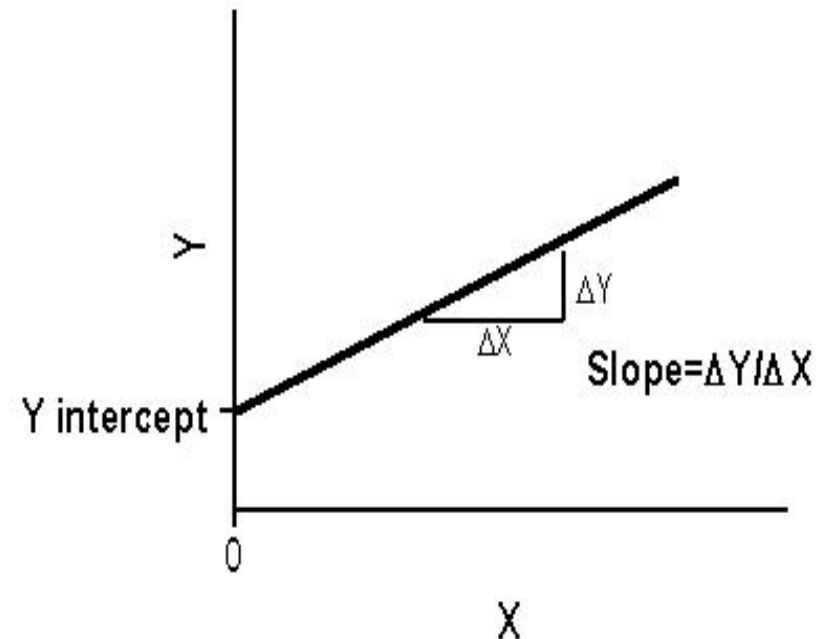
linear regression

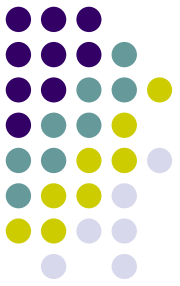
$$y = a + bx + \text{error term (e)}$$

outcome of y is predicted on the basis of x
(both var are continuous)

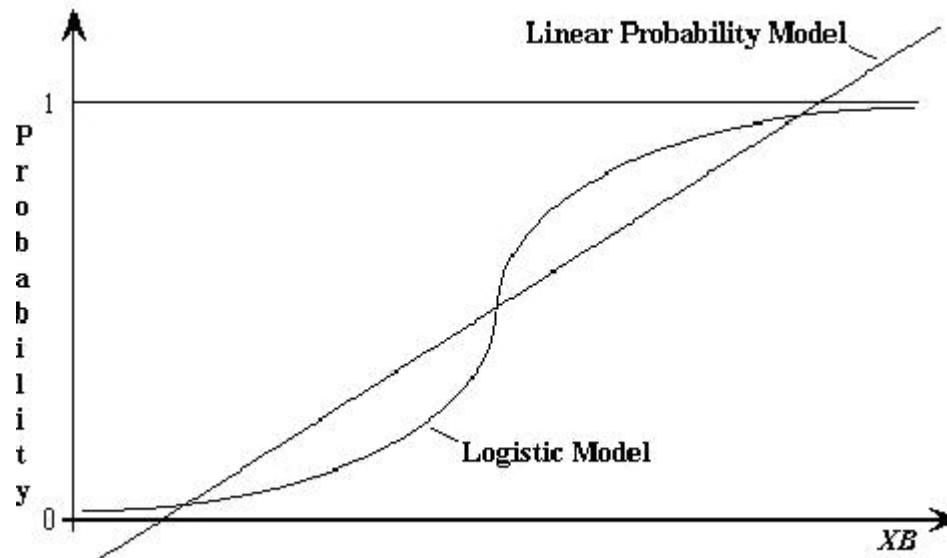
multiple linear regression

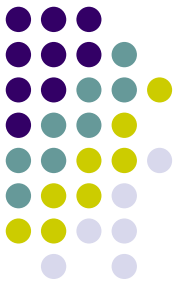
$$y = a + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k + e$$





binary logistic regression





binary logistic regression



We want to know:
probability (p) that r-element will be produced (y=1)

$$p(y=1)$$

probabilities vary between [0; 1], but we want to gather pos. and neg. evidence that is not so limited, i.e. we want $[+\infty; -\infty]$

$$\frac{p(y=1)}{(1-p(y=1))}$$

1st: openness: probabilities to **odds ratios**

$[0; +\infty]$

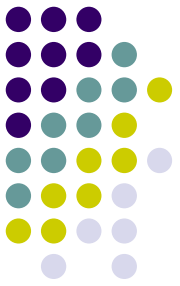
$$\ln\left(\frac{p(y=1)}{(1-p(y=1))}\right)$$

2nd: open up to negative space (rescale):
logarithmic transformation

$[+\infty; -\infty]$

logit

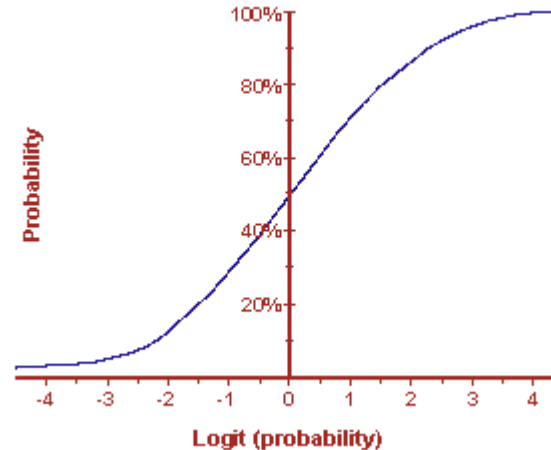


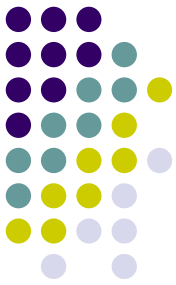


binary logistic regression



$$\ln\left(\frac{p(y=1)}{1-p(y=1)}\right) = \text{logit} = a + b_1 * x_1 + b_2 * x_2 + \dots + b_k * x_k + e$$





Results_{SPSS}: null model

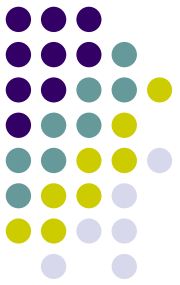
Klassifizierungstabelle^{a,b}

Beobachtet			Vorhergesagt		
			relativizer		Prozentsatz der Richtigen
			No	Th	
Schritt 0	relativizer	No	0	100	,0
		Th	0	100	100,0
	Gesamtprozentsatz				50,0

a. Konstante in das Modell einbezogen.

b. Der Trennwert lautet ,500



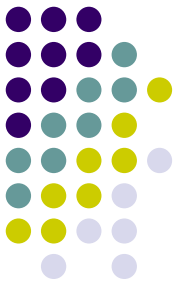


Results

Models have been built with **R** and model comparisons have been performed using **Akaike's Information Criterion**

For the purposes of this presentation I will show **SPSS** output since it tends to be easier to interpret.





Results: inclusion; maximal model_{SPSS}

Omnibus-Tests der Modellkoeffizienten

		Chi-Quadrat	df	Sig.
Schritt 1	Schritt	166,238	78	,000
	Block	166,238	78	,000
	Modell	166,238	78	,000

Modellzusammenfassung

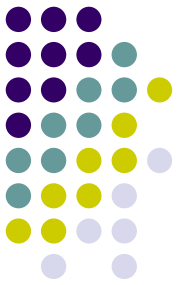
Schritt	-2 Log-Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
1	111,021 ^a	,564	,753

Klassifizierungstabelle^a

Beobachtet		Vorhergesagt		
		relativizer		Prozentsatz der Richtigen
		No	Th	
Schritt 1 relativizer	No	94	6	94,0
	Th	18	82	82,0
Gesamtprozentsatz				88,0

a. Der Trennwert lautet ,500



**most important factors:**

forward stepwise → minimal
adequate model_{SPSS}

The Magnificent Seven

Uniqueness Adjective

-> 60,5% correct classifications

Type of 1st NP in RC (simple)

-> 61%

Animacy of 1st NP in RC

-> 62,5%

Concreteness of 1st NP in RC

-> 63%

Definiteness of head

-> 66,5 %

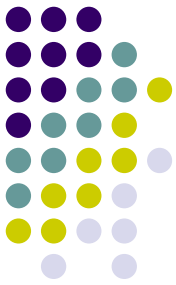
Theta role of head in main clause

-> 71 %

Syntactic role of head in main clause

-> 71% (but better ratio type1/type2 errors)





conditional stepwise deletion → minimal adequate model_{SPSS}

Modellzusammenfassung

Schritt	-2 Log- Likelihood	Cox & Snell R-Quadrat	Nagelkerkes R-Quadrat
1	111,021 ^a	,564	,753
2	111,021 ^a	,564	,753
3	111,021 ^a	,564	,753
4	111,021 ^a	,564	,753
5	111,021 ^a	,564	,753
6	111,237 ^a	,564	,752
7	111,239 ^a	,564	,752
8	111,247 ^a	,564	,752
9	112,266 ^a	,562	,749
10	112,506 ^a	,561	,748
11	112,944 ^a	,560	,747
12	113,079 ^a	,560	,747
13	113,079 ^a	,560	,747
14	114,719 ^a	,556	,742
15	115,207 ^a	,555	,740



conditional stepwise deletion minimal adequate model_{SPSS}



- a. In Schritt 2 entfernte Variablen: animacy.scaled.np1.
- b. In Schritt 3 entfernte Variablen: det.np1.
- c. In Schritt 4 entfernte Variablen: type.head.merge.
- d. In Schritt 5 entfernte Variablen: voice.ec.
- e. In Schritt 6 entfernte Variablen: role.rc.
- f. In Schritt 7 entfernte Variablen: length.dep.
- g. In Schritt 8 entfernte Variablen: plural.marking.
- h. In Schritt 9 entfernte Variablen: role.ec.gen.
- i. In Schritt 10 entfernte Variablen: content.head.
- j. In Schritt 11 entfernte Variablen: uniq.adj.
- k. In Schritt 12 entfernte Variablen: premod.
- l. In Schritt 13 entfernte Variablen: type.np1.merge.
- m. In Schritt 14 entfernte Variablen: embedding.type.
- n. In Schritt 15 entfernte Variablen: pshifts.



seit 1558

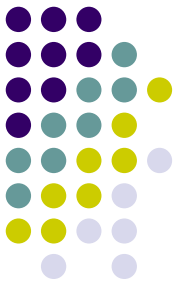
06/02/07

FSU Jena

Klassifizierungstabelle^a

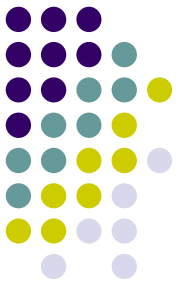
Beobachtet	Vorhergesagt			
	relativizer		Prozentsatz der Richtigen	
	No	Th		
Schritt 1 relativizer	No	9,4	6	94,0
	Th	1,8	8,2	82,0
	Gesamtprozentsatz			88,0
Schritt 2 relativizer	No	9,4	6	94,0
	Th	1,8	8,2	82,0
	Gesamtprozentsatz			88,0
Schritt 3 relativizer	No	9,4	6	94,0
	Th	1,8	8,2	82,0
	Gesamtprozentsatz			88,0
Schritt 4 relativizer	No	9,4	6	94,0
	Th	1,8	8,2	82,0
	Gesamtprozentsatz			88,0
Schritt 5 relativizer	No	9,4	6	94,0
	Th	1,8	8,2	82,0
	Gesamtprozentsatz			88,0
Schritt 6 relativizer	No	9,3	7	93,0
	Th	1,7	8,3	83,0
	Gesamtprozentsatz			88,0
Schritt 7 relativizer	No	9,3	7	93,0
	Th	1,7	8,3	83,0
	Gesamtprozentsatz			88,0
Schritt 8 relativizer	No	9,4	6	94,0
	Th	1,7	8,3	83,0
	Gesamtprozentsatz			88,5
Schritt 9 relativizer	No	9,3	7	93,0
	Th	1,8	8,2	82,0
	Gesamtprozentsatz			87,5
Schritt 10 relativizer	No	9,4	6	94,0
	Th	1,8	8,2	82,0
	Gesamtprozentsatz			88,0
Schritt 11 relativizer	No	9,3	7	93,0
	Th	1,9	8,1	81,0
	Gesamtprozentsatz			87,0
Schritt 12 relativizer	No	9,4	6	94,0
	Th	2,0	8,0	80,0
	Gesamtprozentsatz			87,0
Schritt 13 relativizer	No	9,4	6	94,0
	Th	2,0	8,0	80,0
	Gesamtprozentsatz			87,0
Schritt 14 relativizer	No	9,3	7	93,0
	Th	1,9	8,1	81,0
	Gesamtprozentsatz			87,0
Schritt 15 relativizer	No	9,2	8	92,0
	Th	2,1	7,9	79,0
	Gesamtprozentsatz			85,5

a. Der Trennwert lautet ,500



Discussion





DIFFICULTIES: spoken language is messy...

“<, > cos I mean **the company ones** _{RC} [that they pay you to go out you know like all the wine's on and all the food's free <, >”

Where's the gap?

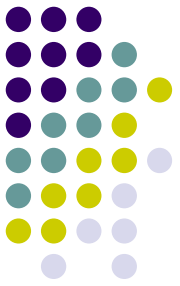
“<, > Well I started off applying to **jobs** _{RC} [that I **was kind of like** _ i in architecture]”

How long is the dependency domain?

“<, > and then there are **the really bland ones** [that I think oh come on]”

Is this actually a RC?





DIFFICULTIES: operationalization

classification scheme does not capture everything we might think influences complexity, e.g. conceptual complexity of predicates

“<, > I don't know what Jan Aarts would think uh would be able to commit ourselves to producing a text grammar uh uh $_{PP}$ [within $_{NP}$ [any foreseeable time limit $_{RC}$ [that that you *are likely to propose* __]]] <, >”

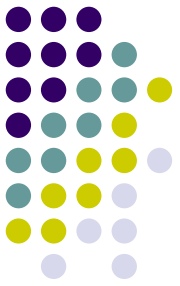
analytic predicates φ_{complex}

$\varphi_{\text{complex}}(\text{agent}_i, \varphi_2(\text{agent}_i, (x)))$

φ_{complex} e.g. IS.LIKELY(agent_i , PROPOSE(agent_i , p))

(alternatively modal logic: $\diamond P$ (it is possible 'that p'))





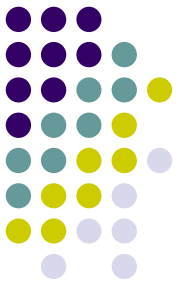
DIFFICULTIES: operationalization

**classification scheme does not capture everything...
unexpected production of relativizer:**

“I mean out of everyone_i [_{RC} **that** I know ____i] _{RC} **that** I went to college with ____i] ...

**Systematic?
Due to some kind of parallelism?**





Thanks a bunch.
See you next week.

