

# Referential Density in Typological Perspective

Balthasar Bickel

[www.uni-leipzig.de/~bickel](http://www.uni-leipzig.de/~bickel)

# An anthropological observation

*Belhare Pear Story* (Sino-Tibetan; Nepal; Bickel 2003)

ɣila .. a: .... ambibu      phighe                      kinahungo  
first                      mango              picked.from.above              and.then

otutui?=na      jhola-e ukthe  
quite.big=a      bag-in      took.down

inetnahungo      dhaki-e      leŋse  
and.then              basket-in      put

illam              il-lam              sassaba      leŋse      ʌni ...  
from.there      from.there      by.pulling      lead      and.then

riksha,              e:      saikil-lamma,      saikil-lamma              tahe      kinahungo...  
rickshaw              bicycle-on              bicycle-on              came      and.then

# An anthropological observation

*Maithili Pear Story* (Indo-European; Nepal; Bickel 2003)

ekṭā ām-ke gāch rahai. ā... a... a....  
one mango-of tree is

ām me ek, egoṭā chaurā ām torait rahai  
mango in one one boy mango plucking is

u ām toir-ke ṭokari me rakhne jāi chelai.  
s/he mango having.plucked basket in to.keep going was

omaharse egoṭā chaurā elai,  
and.then one boy came

laḍkā sāikal par caḍhne, ā...  
boy:HON bike on to.ride

## RD measurement

$$RD = \frac{N \text{ (overt argument NPs)}}{N \text{ (available argument positions)}}$$

RD measurements on Pear Stories (Chafe 1980),

10 speakers per language

7 languages (as of now, more in the works)

## RD measurement (cont'd)

- Exclude metapragmatic comments and reported speech verbs.
- Count available arguments per verb form, count complex predicates and serial verbs as one verb.

### *Belhare*

seu pheri leŋ-si un kha?-yu.  
apple again direct-SUPINE 3sNOM [3sS-]go-NPST

‘S/he/someone went again to put apples [into the basket].’

### *Maithili*

okrā lālac āb-ait chai.  
3hDAT greed come-IPFV.PART is

‘S/he/they is/are getting greedy.’

## RD measurement (cont'd)

- Exclude metapragmatic comments and reported speech verbs
- Count available arguments per verb form, count complex predicates and serial verbs as one verb.
- Count existentials and equationals as licensing one argument position.

## RD measurement (cont'd)

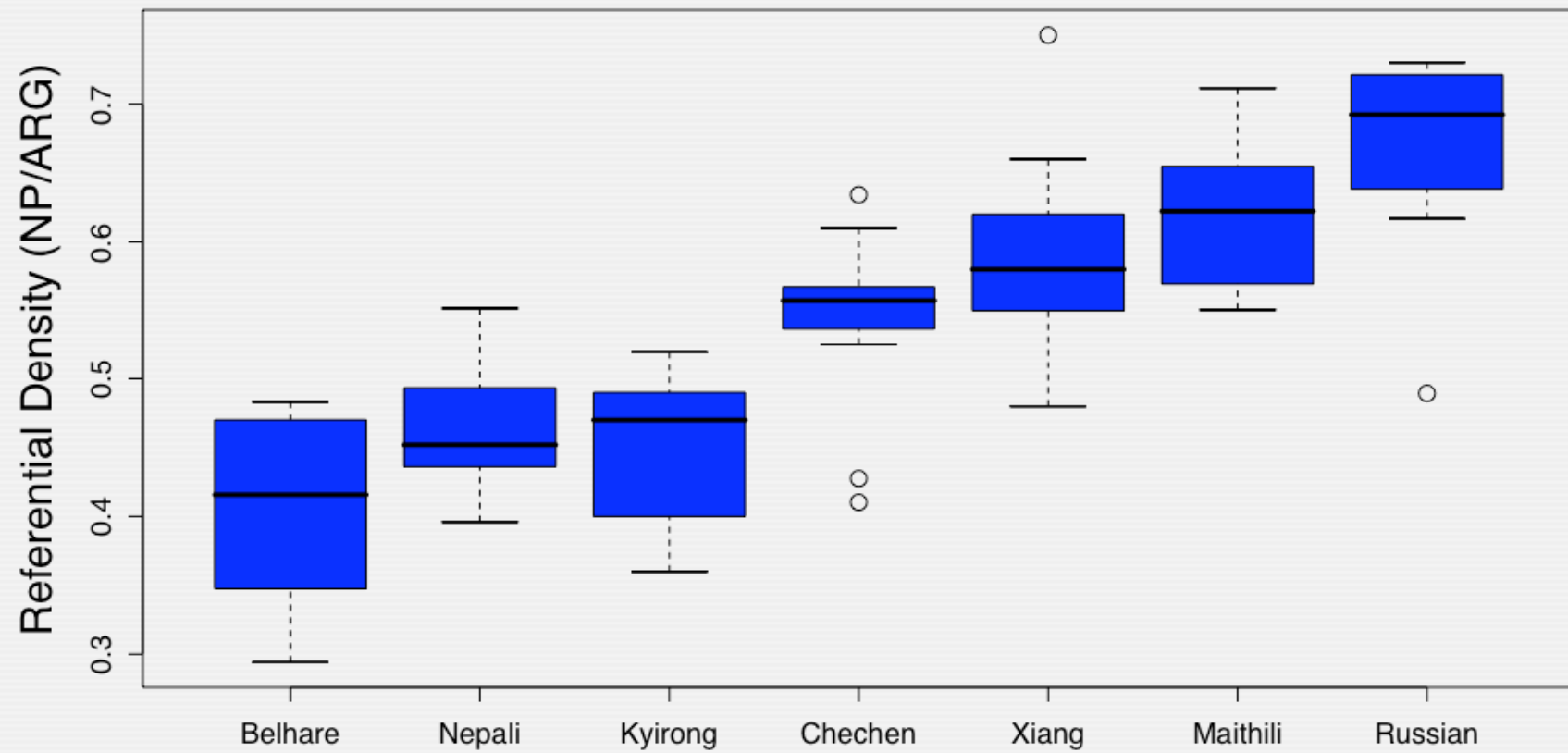
- Exclude metapragmatic comments and reported speech verbs.
- Count available arguments per verb form, count complex predicates and serial verbs as one verb.
- Count existentials and equationals as licensing one argument position.
- Count as overt NPs any NP (pronoun, demonstrative, noun, proper name etc.), including complex NPs with relative clauses.

## RD measurement (cont'd)

- Exclude metapragmatic comments and reported speech verbs.
- Count available arguments per verb form, count complex predicates and serial verbs as one verb.
- Count existentials and equationals as licensing one argument position.
- Count as overt NPs any NP (pronoun, demonstrative, noun, proper name etc.), including complex NPs with relative clauses.
- Delimit arguments from adjuncts through semantics of role markers



# A survey



# What causes the differences?

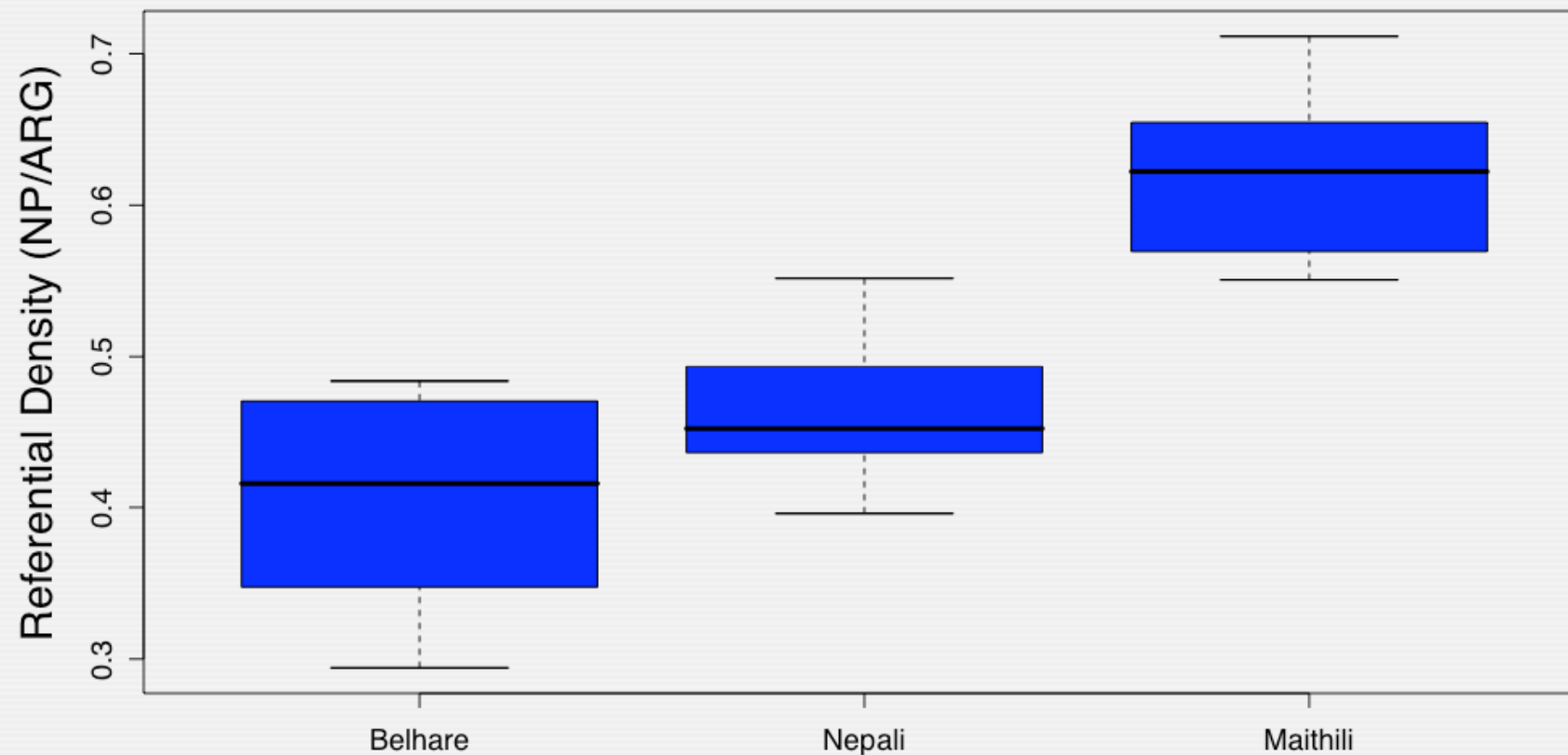
Plausible factors:

- Local discourse traditions
- Text length
- Sociology of communication
- Some structural property of grammar

## Local traditions

- The current sampling does not allow to build this into a statistical model as factor.
- But we can analyze a controlled sample of languages that share local traditions (stock of folklore, many conversational strategies, e.g. politeness strategies, rituals, myths, etc.):
  - Belhare
  - Maithili
  - Nepali
- Test for differences between these groups (Bickel 2003)

## A first hypothesis: local traditions



$F(2,27) = 37.0, p = 1.8e-08$  (cf. Bickel 2003)

## Factorial analysis

- For the other factors (text length, grammar, sociology), the sample is rich enough for factorial analysis.
- But first, what are the factors? What are the predictions?

## Text length

- Perhaps people who talk more use more NPs!
- Prediction from Factor “LENGTH”:
  - LENGTH ~ RD

# Sociology of communication

- Common observation in the Ethnography of Speaking: people who know each other ('close-knit society') can presuppose more information than strangers.
- This might create a conventional style of 'presuppositionalistic' discourse that is also used when talking about unknown referents, as in the Pear Story experiment.
- Predictions from Factor "SOC":
  - close-knit society → low RD
  - loose nets, large-scale society → high RD

## Structural properties of grammar

- First guess, although typologically suspect: languages with agreement morphology allow more pro-drop than languages without agreement morphology.
- Predictions from factor “M\_AGR”:
  - agreement → low RD
  - no agreement → high RD
- Another guess (Bickel 2003): structural priming effects from syntax to NP use



# Structural Priming Effects

A construction primes a structurally associated construction  
(e.g.  $V \rightarrow NP$ ; Lu & al 2001).

This construction primes its subsequent re-use  
(e.g.  $NP \rightarrow NP$ ; Bock 1986 etc.)

Long-term persistence and learning effects (Bock & Griffin 2000)

If frequent and strong, this could habituate  
speakers into a rhetorical norm.

## Structural Priming Effects (cont'd)

- For this to show, we need a construction that
  - primes NPs through structural association and that
  - is frequently used in discourse
- Such a construction should increase the use of NPs in discourse:

### **Case-based agreement**

# Structural Priming Effects (cont'd)

Case-based agreement in Maithili (Indo-European)

- a. (tũ)            bimār    ch-**æ**?  
    **2nhNOM**    sick      be-**2nhNOM**
- b. (torā)        khuśi    ch-**au**?  
    **2nhDAT**    happy    be-**2nhNONNOM**

# Structural Priming Effects (cont'd)

## Case-based agreement in Chechen (Nakh-Dagestanian)

- a. k'ant-as jo' 'exa j-o.  
boy-ERG girl[-NOM] lie FEM-do.NPST
- b. k'ant jo' 'exa j-ie-sh v-u.  
boy[-NOM] girl[-NOM] lie FEM-do-CVB MASC-be.NPST
- c. k'ant-as jo'-ana gho d-o.  
boy-ERG girl-DAT help D-do.NPST (= default agreement)
- d. k'ant jo'-ana gho d-ie-sh v-u.  
boy[-NOM] girl-DAT help D-do-CVB MASC-be.NPST
- e. k'ant-ana jo' go.  
boy-DAT girl[-NOM] see.NPST

## Structural Priming Effects (cont'd)

- No agreement
- Case-insensitive agreement in Belhare (Sino-Tibetan)

a. (han) khar-e-**ga** i?  
**2sNOM** go-PST-**2s** Q

b. (hanna) un lur-he-**ga** i?  
**2sERG** 3sNOM tell-PST-**2s** Q

c. ciya (hannaha) n-niña tis-e-**ga** i?  
tea.NOM **2sGEN** 2sPOSS-mind be.easy-PST-**2s** Q

## Structural Priming Effects (cont'd)

Prediction from Factor “SYN”:

Case-based AGR primes argument NPs because they are structurally associated and therefore regularly co-activated; if frequent and strong this should lead to a higher ratio of overt NPs per argument position (= Referential Density or RD):

**Case-based AGR → high RD**

NB: the hypothesis is a *unidirectional implication*, ergo:

**not case-based agr → low *or* high RD,**

depending on other factors, e.g. SOC

→ **Factorial Analysis**

→ **Expect interaction between SYN and some other factor(s)**

# Factorial analysis

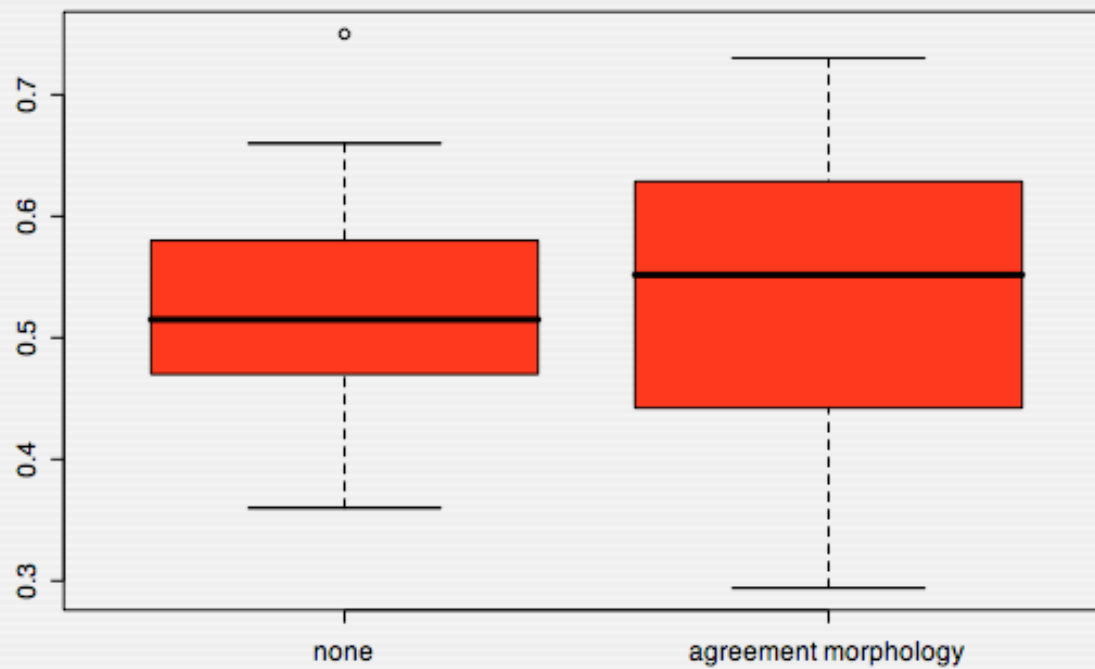
ANCOVA: SOC\*SYN\*M\_AGR\*LENGTH

	<i>Df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
SOC	1	0.09	14.55	0.001
SYN	1	0.05	8.07	0.006
LENGTH	1	0.00	0.00	0.97
M_AGR	1	0.04	6.82	0.011
SOC:SYN	1	0.06	8.97	0.004
SOC:LENGTH	1	0.00	0.59	0.45
SYN:LENGTH	1	0.00	0.44	0.51
LENGTH:M_AGR	1	0.00	0.40	0.53
SOC:SYN:LENGTH	1	0.00	0.00	1.00
Residuals	51	0.01		

Adjusted  $R^2 = .339$

# Factorial analysis

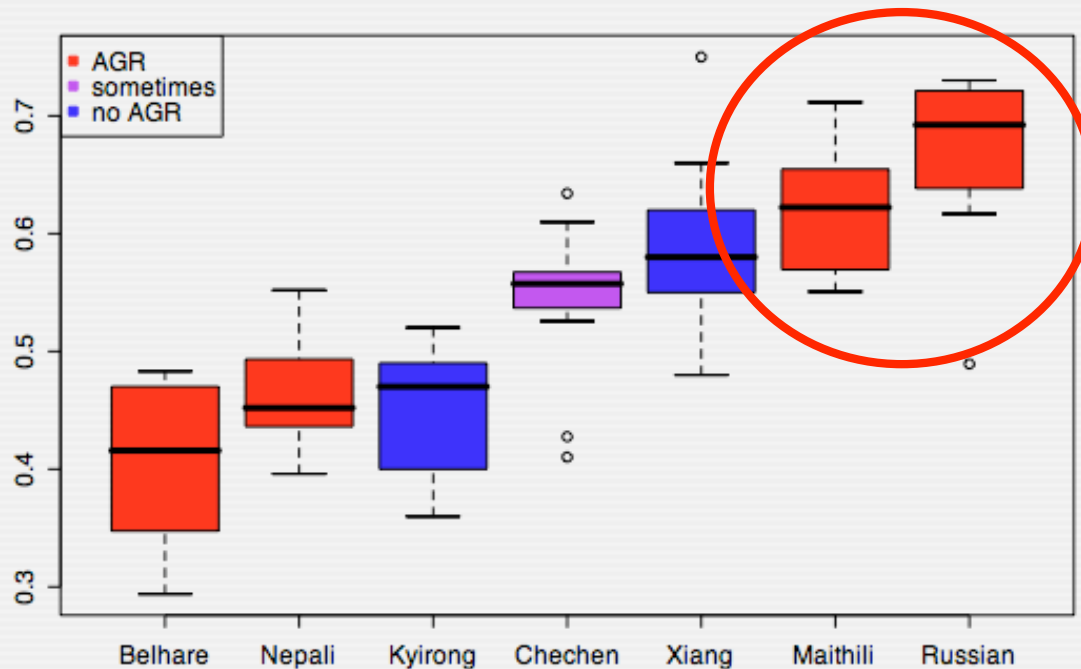
M\_AGR goes against the hypothesized direction...





# Factorial analysis

...and is probably induced by an accidental association of AGR-languages with the highest RD values:



→ Probably a fake effect; remove from model

# Factorial Analysis

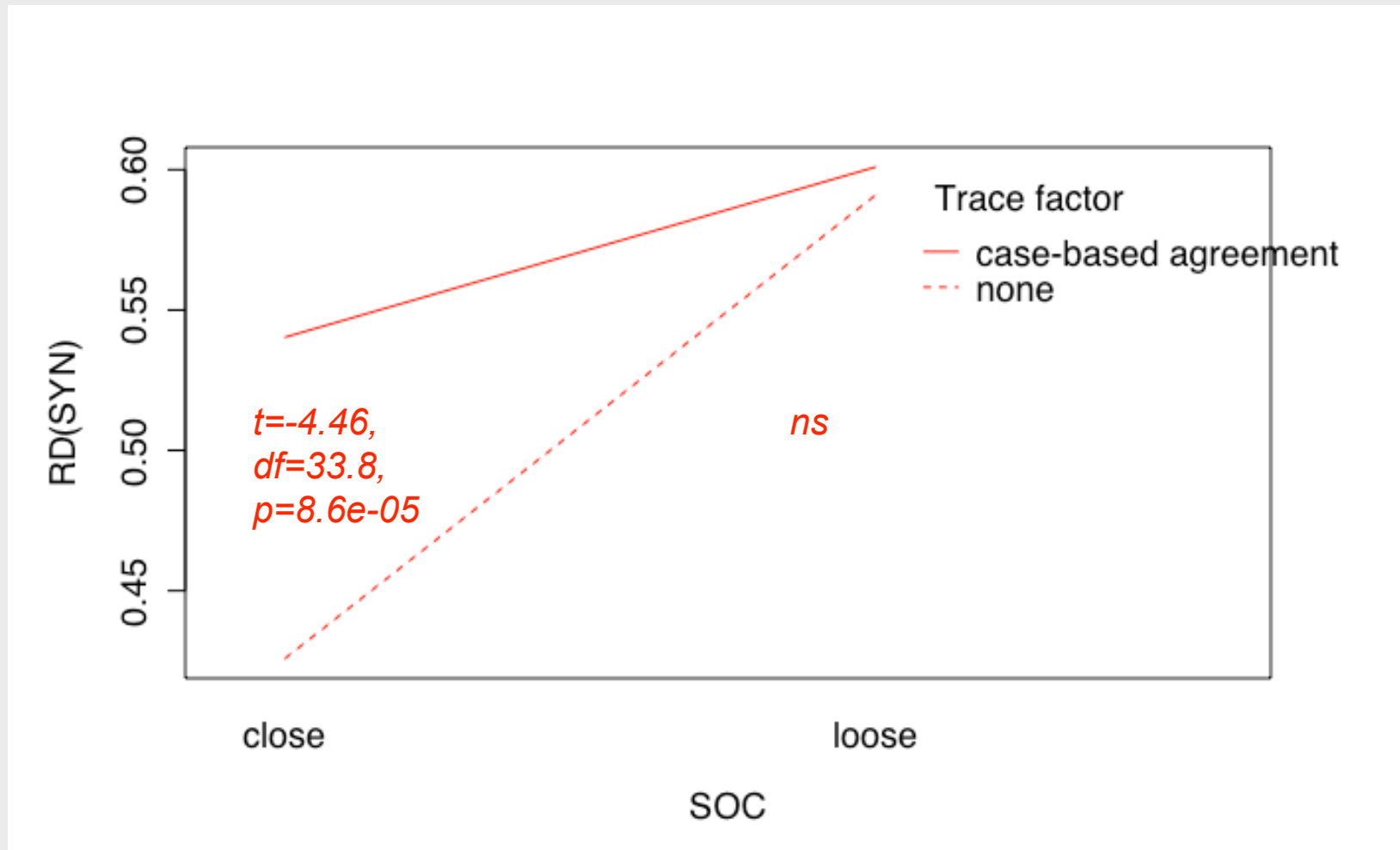
Improved model: SOC\*SYN\*LENGTH

	<i>Df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
SOC	1	0.09	14.61	0.00
SYN	1	0.05	8.11	0.01
LENGTH	1	0.00	0.00	0.97
SOC:SYN	1	0.09	14.46	0.00
SOC:LENGTH	1	0.00	0.24	0.63
SYN:LENGTH	1	0.00	0.71	0.40
SOC:SYN:LENGTH	1	0.00	0.09	0.77
Residuals	53	0.01		

Adjusted  $R^2 = .342$

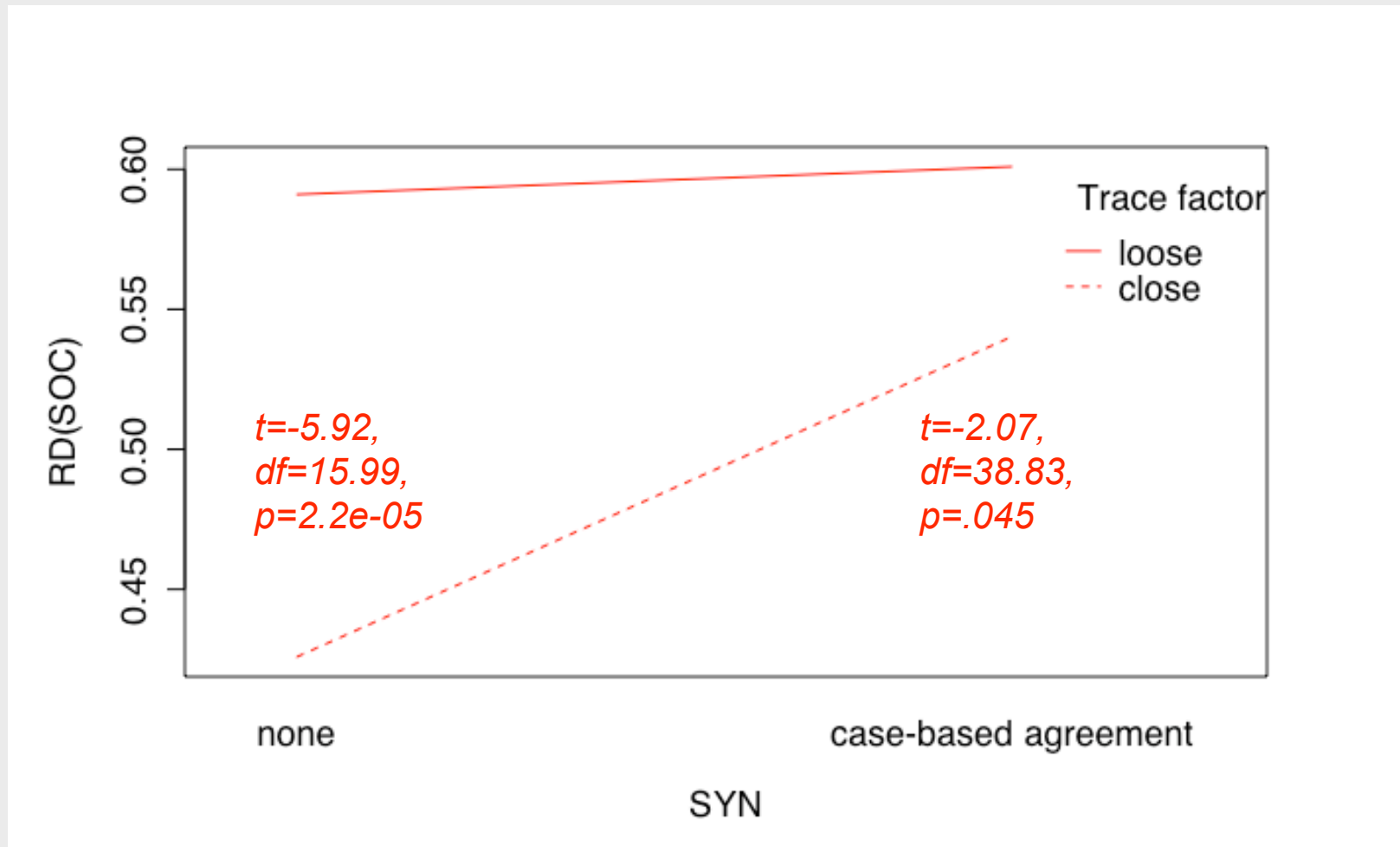
# Factorial analysis

Interaction effect SOC\*SYN:



# Factorial analysis

Interaction effect SYN\*SOC:



## Factorial analysis: summary

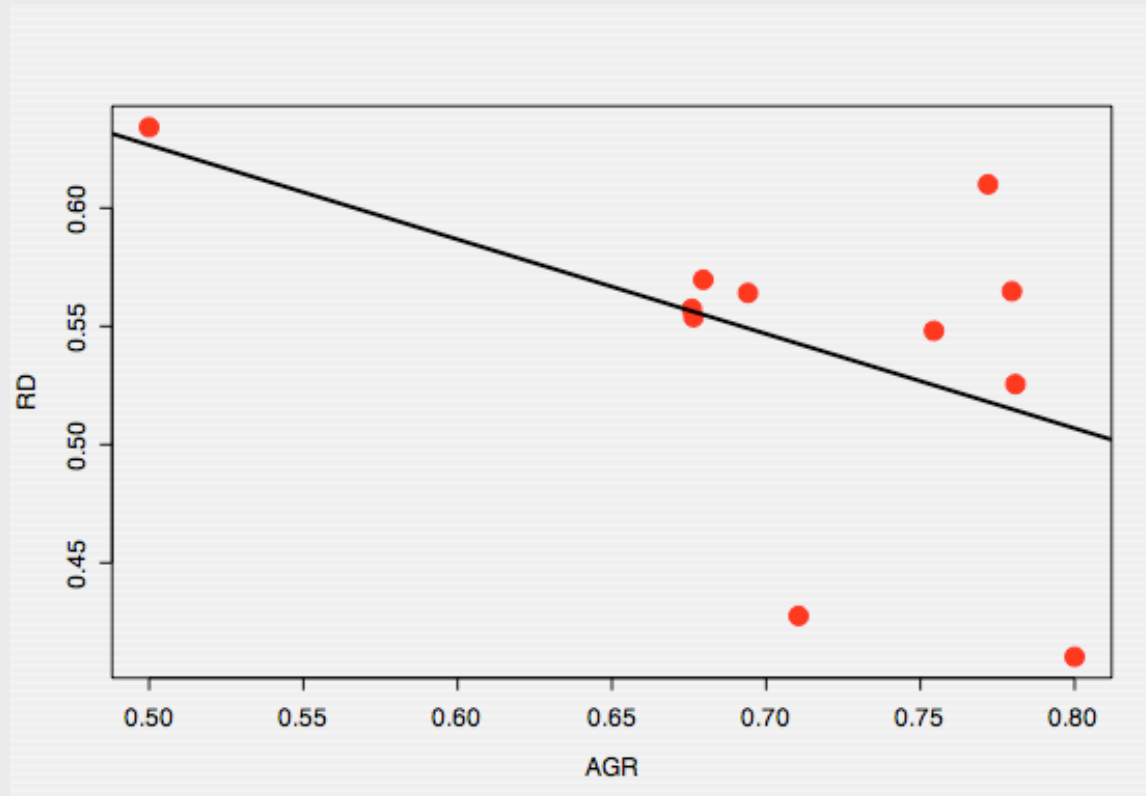
- RD varies systematically depending on both SOC and SYN factors
- Each factor is canceled out by the other under high RD levels, i.e.
  - If SOC yields high RD, SYN has no effect
  - If SYN yields high RD, SOC has no / a borderline effect
  - If SOC keeps RD low, SYN has an effect
  - If SYN keeps RD low, SOC has an effect

## A closer look at SYN

- SYN is based on a priming effect of agreement constructions on structurally associated NPs.
- Does the RD effect of SYN directly depend on the actual occurrence of case-based agreement constructions?
- We can answer this in Chechen because agreement varies:
  - only auxiliaries and ~30% of lexical verbs show agreement (but all agreement is case-based)

## A closer look at SYN

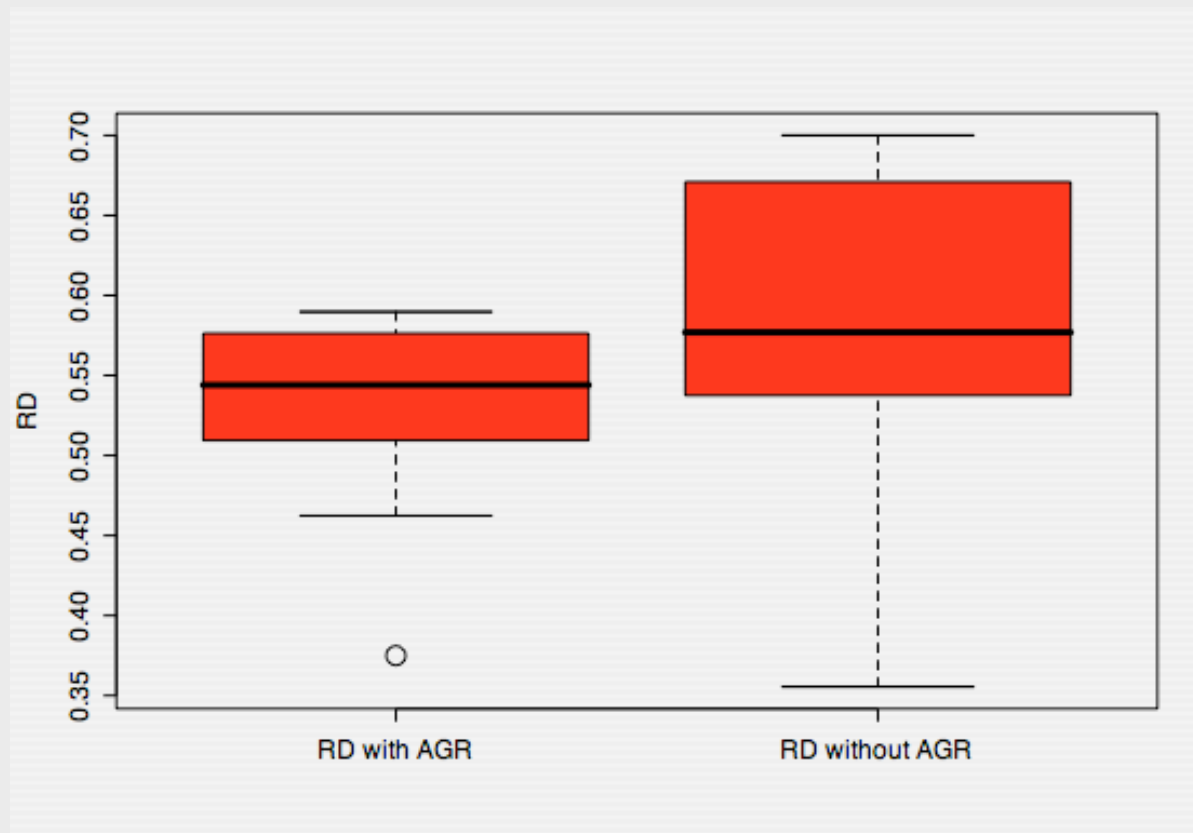
1. Test correlation between AGR and RD across the ten Chechen narratives.



*Coefficient AGR =  $-.399$ ,  $p = .12$ , Adjusted  $R^2 = .16$*

## A closer look at SYN

2. Compare RDs in clauses with agreement and those without (paired design):



$$t = -1.53$$

$$df = 10$$

$$p = .15$$

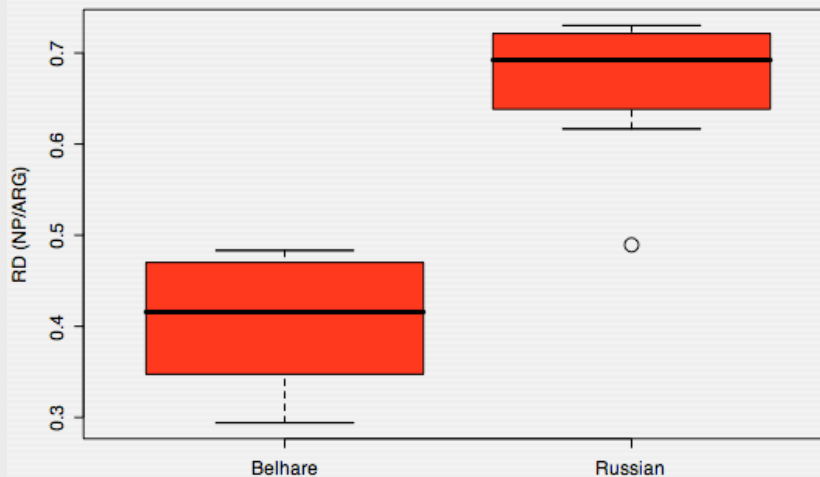


## A closer look at SYN: conclusions

- The SYN effect is not directly caused by processing agreement rules. Instead:
- Processing agreement rules → priming effects → habituation → conventionalization over time → discourse effects
- Independent confirmation: RD trends are amazingly stable in language contact:
  - Belhare vs. Nepali

## But how deep are RD differences?

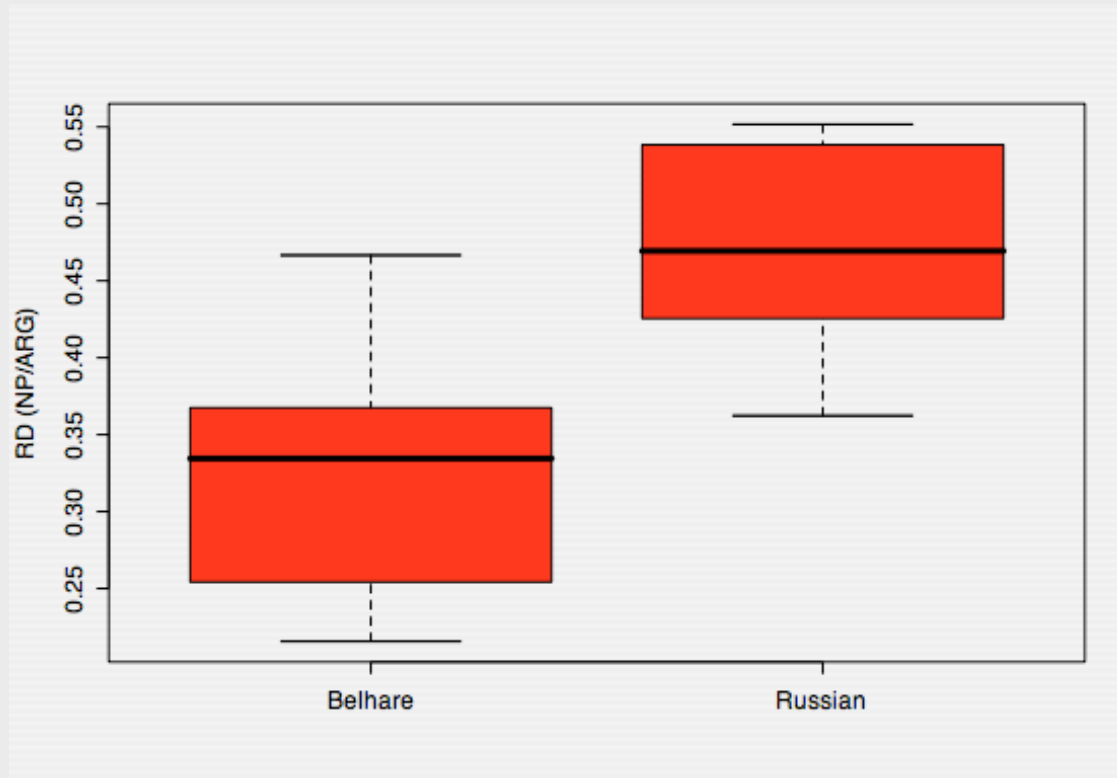
- Is it just about pronoun use? Or does it affect more informative NPs as well?
- Scale of informativeness:  
lexical NP > generic NP > pronoun >  $\emptyset$
- Study by Stoll & Bickel (2006) contrasting the most extreme languages in the sample:



$t = -8.49,$   
 $df = 18,$   
 $N = 20,$   
 $p < .001$

# How deep are RD differences?

Comparing RD among lexical NPs only



$t = 4.49,$   
 $df = 18,$   
 $N = 20,$   
 $p < .001$

Corollary: no significant difference in the pronoun/noun ratio.

## How do they do it?

### High-RD style in Russian

‘A garden. Near the river a village is visible. The owner is collecting pears in the garden. (He) collected one basket. Another man with a goat appears. The goat is baaing. They went by. The man with the goat went by. And the owner of the garden went to collect the second basket. Here, a boy came towards them on a bicycle, probably his son.’

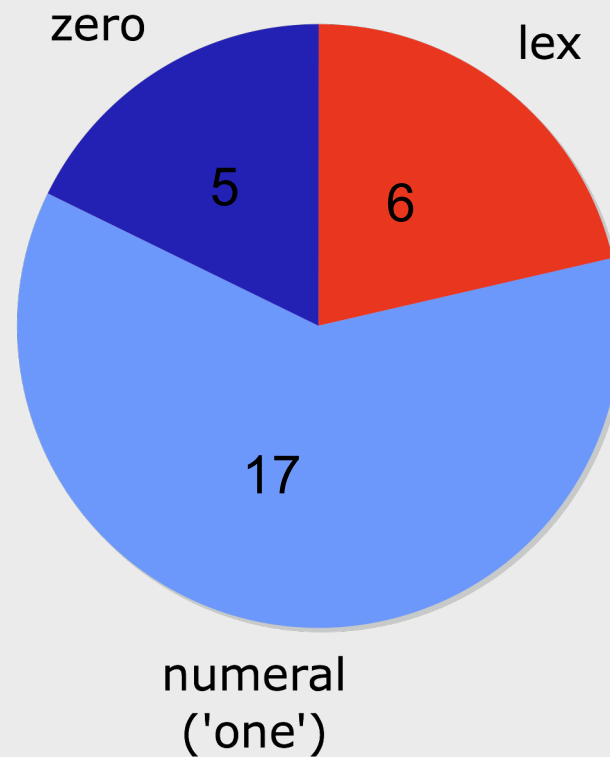
## How do they do it?

### Low-RD style in Belhare

‘Now, first, (someone) picked apples. (S/he) picked apples and filled (them) into a basket. Then... then, one came along pulling a goat. (S/he) came pulling a goat and took (it) over there across; (s/he) took (it) over there across like that ((gesturing)). (S/he) took away that goat. And then, a bicycle} ... one came on a bicycle.’

## How do they do it?

First mentioning of referents across all 10  
Belhare speakers



## How do they do it?

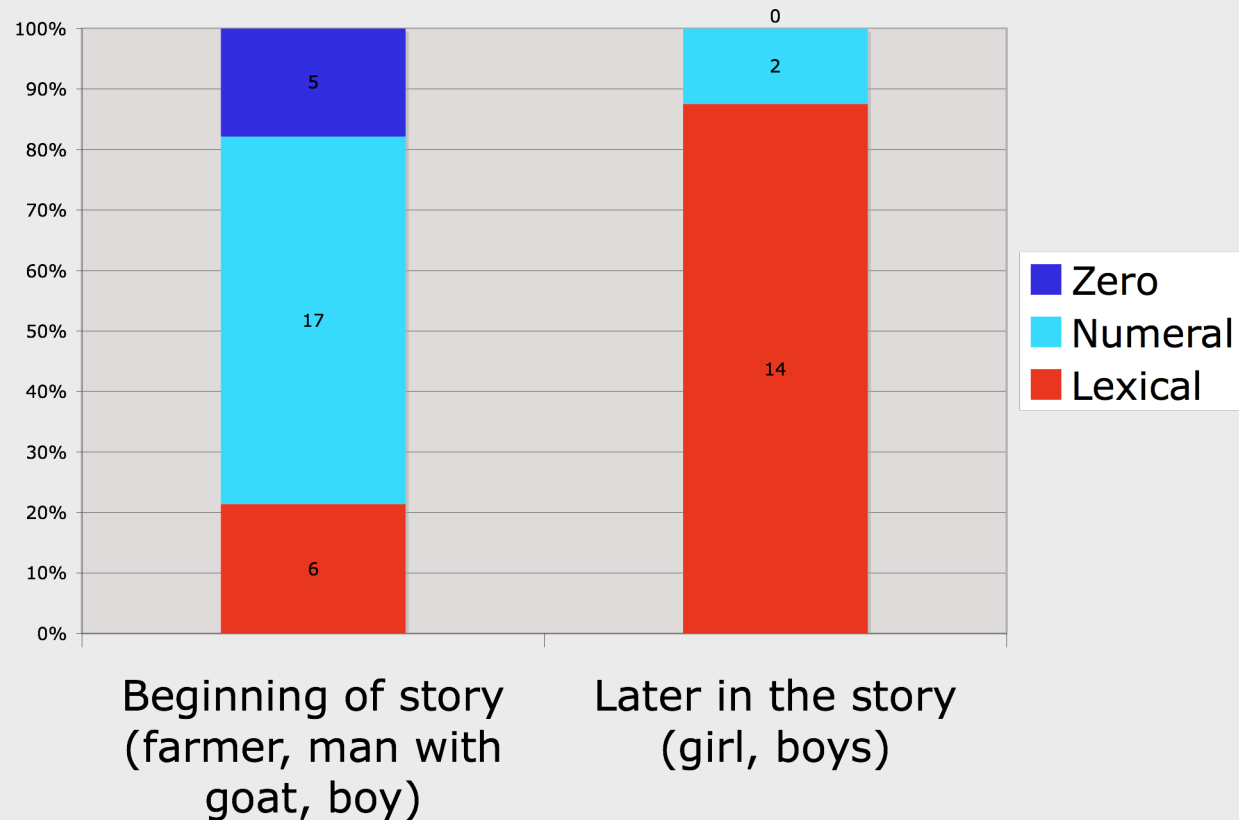
- In high-RD languages, the names introduced at the beginning serve as reference trackers:  
‘Here came a boy towards them on a bike, probably his son. (He) took a basket, put it on the bicycle and carried it away. (He) carried it away. Then, a girl arrived, also on a bicycle. (She) took the second basket. And so. They stumbled on the bicycles. The boy with the basket, with the baskets, with the bags, fell. (Some) guys showed up, who helped the boy to collect the fruits into the basket. (They) lifted it up. And he carried (it) away. So. During the time of collision, when the boy fell, his hat fell off. The guys, who had helped him this one to collect the fruits. went further.’ ’

## How do they do it?

- Information management in a low-RD language  
'(S/he) came on a bicycle and then (s/he) filled the apples (into the basket) on the bicycle. (S/he) filled in the apples and then went away. (S/he) went away and then again came. Then, another girl came, ... on a bicycle. Then, they crashed their bicycles (into each other) and he fell} they fell. They fell and ... the girl went over there. As for the boy, he patted his knee right there (where he was). He patted his knee and then four} four (guys) picked up apples for a while, and then one lifted up the bicycle and then (s/he) went away.'



# Information management with low RD



- All subjects pooled together: Fisher Exact Test,  $p = .00007$
- Mean (Lexical) per subject: Wilcoxon Test,  $p = .004$

## Information management with low RD

- Hold back information as long as possible
  - Introduce referents only when needed
- ‘(S/he) came on a bicycle and then (s/he) filled the apples (into the basket) on the bicycle. (S/he) filled in the apples and then went away. (S/he) went away and then again came. **Then, another girl came, ... on a bicycle. Then, they crashed their bicycles (into each other) and he fell} they fell. They fell and ... the girl went over there. As for the boy, he patted his knee right there (where he was).** He patted his knee and then four} four (guys) picked up apples for a while, and then one lifted up the bicycle and then (s/he) went away.’

# Conclusions

- Typological variance in both RD and RD<sub>lex</sub>
- This reflects differences in information management:
  - free info vs. info only when needed.
  - cf. collaborative information management reported in the Ethnography of Speaking literature (e.g. Besnier 1989)
- Perhaps RD differences also have an impact on attentional balance between events and participants.
- Main causes of the typological variance:
  - degree of presuppositionalism induced by size of social network
  - conventionalized style originating from structural priming effects triggered by the way agreement works
- Experimental demonstration of a 'deep' relativity effect from syntax on discourse.

# Acknowledgements

- Belhare: Lekhbahādur Rāi
- Kyirong: Brigitte Huber
- Nepali: Mādhavprasād Pokharel, Karṇākhar Khativaḍā, Rāmrāj Lohani, Bhimnārayaṇ Regmi
- Maithili: Yogendra Yādava, Bindeśvar Yādav
- Chechen: Zarina Molotchieva, Alena Witzlack, Johanna Nichols
- Xiang: Yan Luo; data from Mary Erbaugh ([www.pearstory.org](http://www.pearstory.org))
- Russian: Mixail Lurie, Tatjana Krugljakova
- Swiss National Science Foundation, 1998-2002
- University of Leipzig, 2002-2004