

Korpusbasierte Variablen in der Typologie

Balthasar Bickel

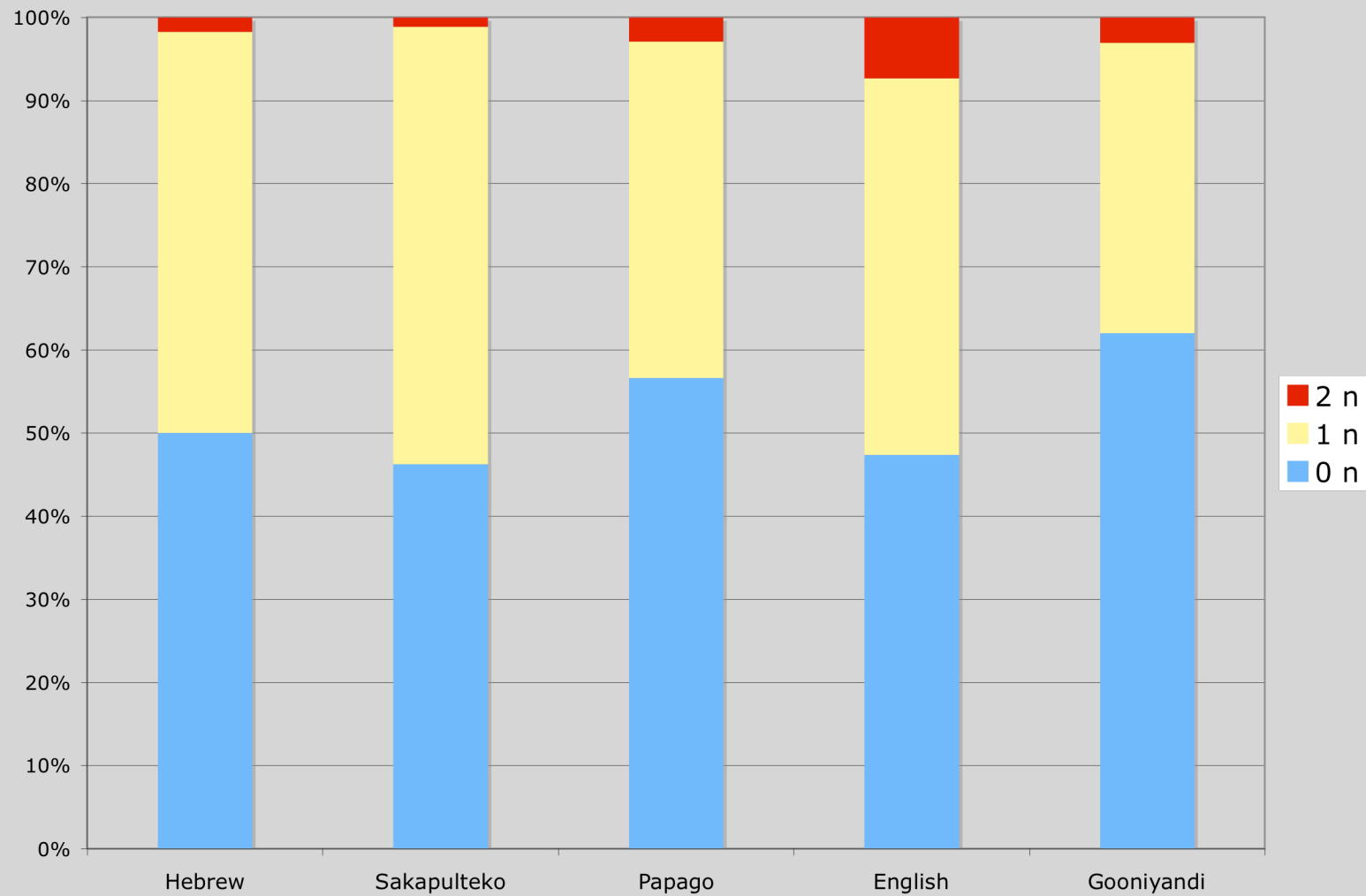
www.uni-leipzig.de/~bickel

Klassische Typologie

- Entdeckung universeller Diskurspräferenzen in den 80er Jahren, zum Beispiel:
- “Preferred Argument Structure” (PAS) (DuBois 1987, 2003)

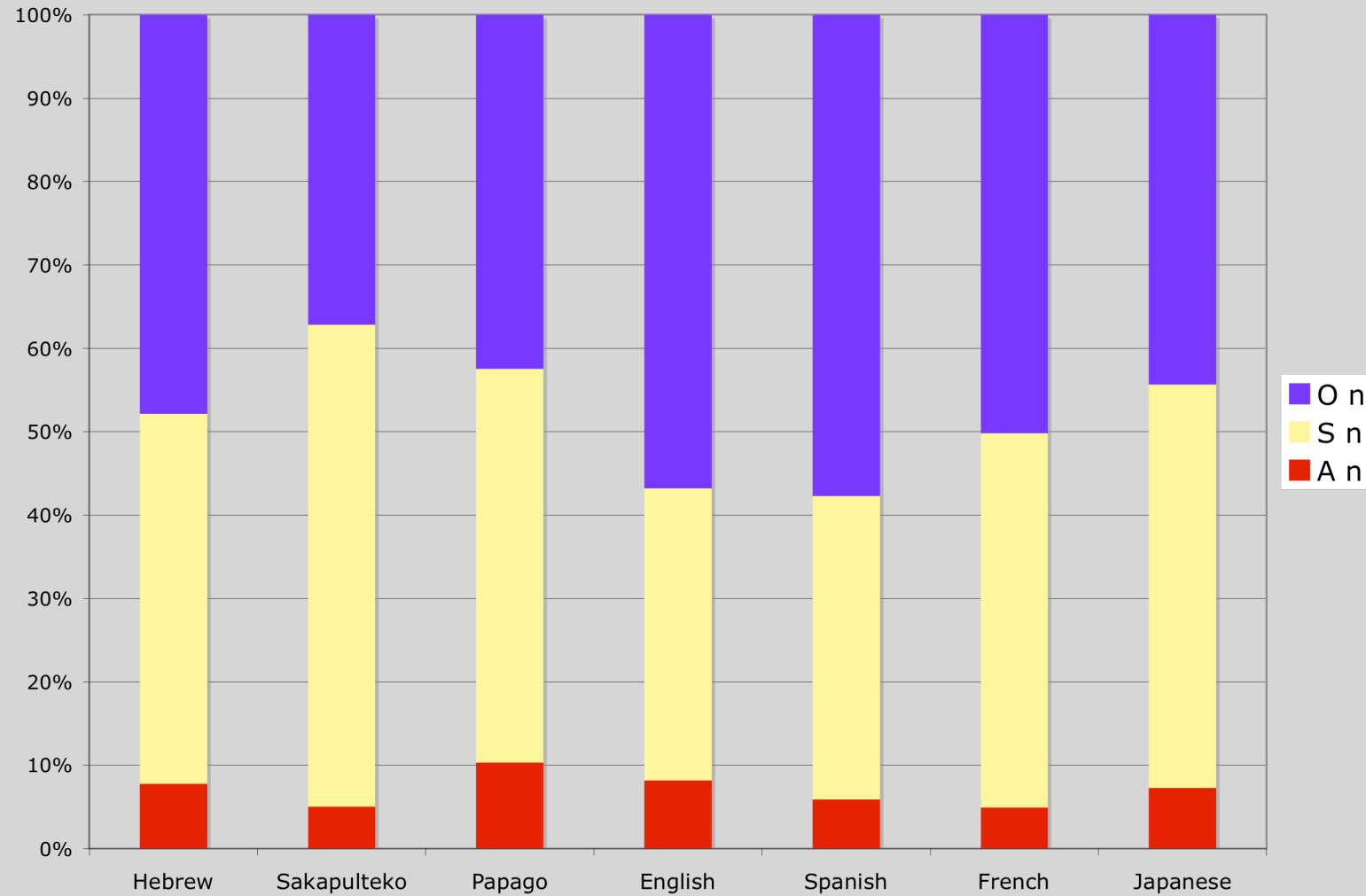
	Grammatik	Pragmatik
Quantität	Nicht mehr als einen lexikalischen Aktanten!	Nicht mehr als einen neuen Aktanten!
Rolle	Vermeide lexikalische A-Aktanten!	Vermeide neue A-Aktanten!

Klassische Typologie: PAS



(DuBois 2003)

Klassische Typologie: PAS



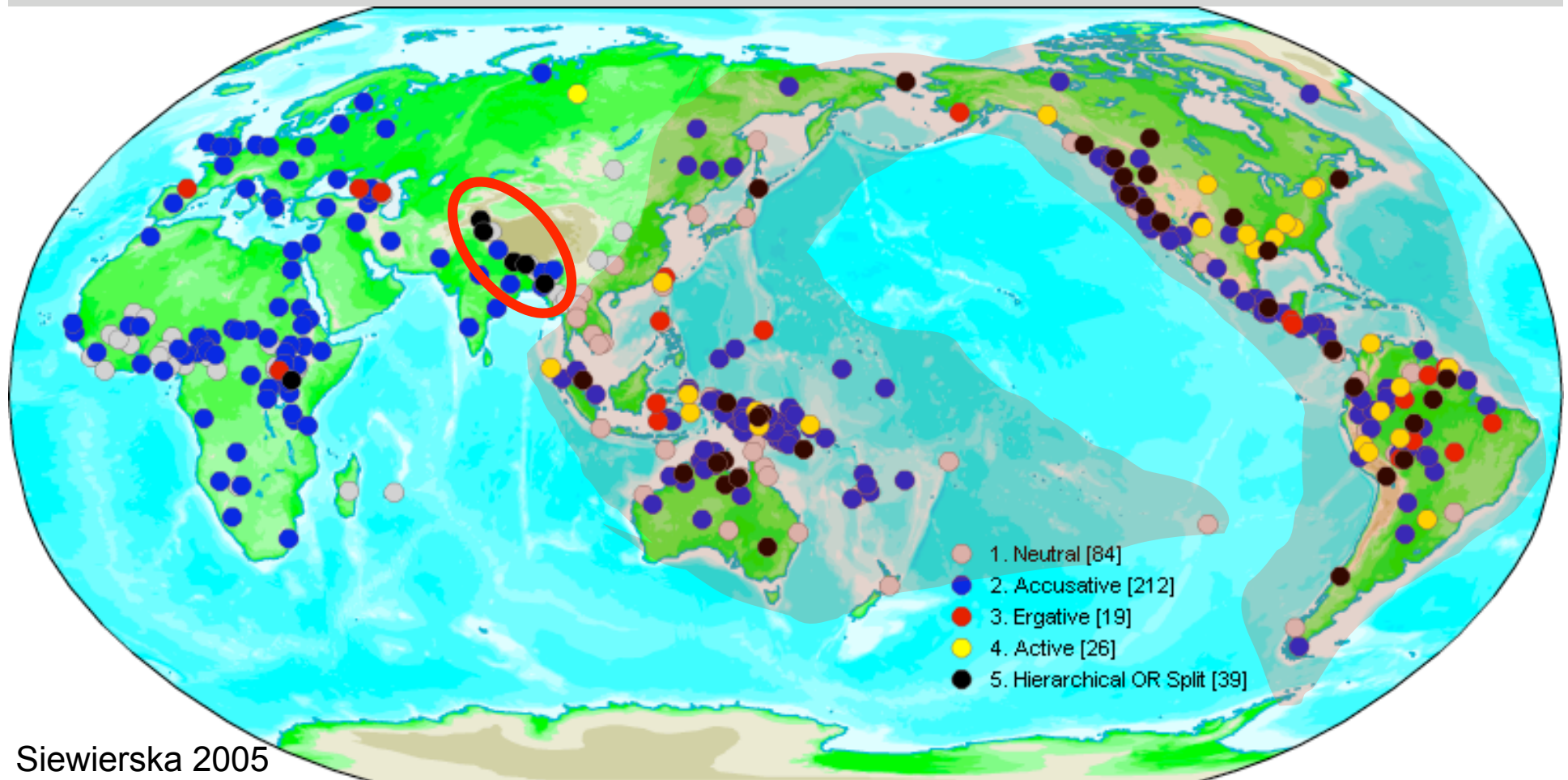
(DuBois 2003)

Klassische und neoklassische Typologie

- *Hypothese*: Universelle Diskursstrukturen motivieren universelle Grammatikalisierungsmuster (Givón, DeLancey, Bybee etc.); aktuelle Renaissance in stochastischer OT (Bresnan, Clark) und spieltheoretischer Typologie (Jaeger)
- *Beispiel*: Gespaltene A-markierung
 - $F(\text{A-anim}) > F(\text{A-inanim})$ universell im Diskurs, ergo
 - $*\text{MarkA/anim} \gg * \text{MarkA/inanim}$ universell in der Grammatik, ergo
 - $F(\text{Sprachen mit MarkA nur auf [inanim]}) > F(\text{Sprachen mit MarkA nur auf [anim]})$, ergo z.B.
 - $F(\text{Sprachen mit distinkt markierter 1sgA})$ extrem selten
 - aber...

Probleme

- F (Grammatiken) ist anderen Faktoren als F (Sätze) unterworfen:



Probleme

- *F* (Grammatiken) auch Produkt der lokalen Geschichte, ergo
- offen für lokale Entwicklungen, z.B.
- *areales* “Markiere-Sprecher-als-A” Prinzip im Himalaya (Bickel 2000)
- Resultat: Sprachen mit distinkt markierter 1sgA *lokal* häufig
 - Grammatikpräferenzen können nicht ausschliesslich von universellen Diskurspräferenzen vorausgesagt werden.
 - Universelle Diskurspräferenzen sind nur *ein Teil* der relevanten Faktoren (wie auch in der klassischen, nicht aber der neoklassischen Typologie angenommen).

Probleme

- Ein zweites, grundsätzlicheres Problem: Diskurspräferenzen sind nicht alle universell.
- Deshalb ist das Erklärungsmodell “Grammatikpräferenzen wegen Diskurspräferenzen” nicht beliebig anwendbar.
- Evidenz ...

Typologische Varianz im Diskurs

Konvergierende Entdeckungen in den 90ern:

1. PFAD-Lexikalisierung → PFAD-Diskurs (Slobin)
2. KOORD-Grammatik → KOORD-Diskurs (Levinson &al)
3. QUOTATIV-Gammattik → “Dir. Rede”-Diskurs (Noonan)
4. GR-Grammatik → REFERENZ-Diskurs (Bickel)

PFAD-Kodierung

Grammatische Typologie (Talmy, Slobin):

Er rannte in den Wald.

PFAD-Kodierung

Grammatische Typologie (Talmy, Slobin):

*Er rannte **in den** Wald.*

PFAD

GANGART

Il est entré dans la forêt (en courant).

PFAD-Kodierung

Grammatische Typologie (Talmy, Slobin):

*Er rannte **in den** Wald.*

PFAD

GANGART

*Il est **entré** dans la forêt (en courant).*

PFAD-Kodierung

S-Sprachen, z.B.

Germanisch, Slavisch

Finnougrisch

Chinesisch

V-Sprachen, z.B.

Romanisch (ausser Rhätoromanisch)

Semitisch

Türkisch

PFAD-Diskurs

	S-Sprachen Bsp. Englisch	V-Sprachen Bsp. Spanisch
GANGART- Beschreibung	45% mean tokens 7 mean types	20% mean tokens 3 mean types
PFAD- Segmentierung	3 mean number of segments	2.1 mean number of segments
SZENEN- Beschreibung	8% of narrators	21% of narrators

Bsp: ***aus** dem Fenster **von** dem Fensterbrett **herunter***

(Slobin 1997, 2003, 2004; oral narratives “Frog Stories”)

GR-Grammatik

Hypothesis (Bickel 2003, 2005): Case-based AGR primes argument NPs because they are structurally associated and therefore regularly co-activated; if frequent and strong this should lead, through recurrent priming effects, to a higher ratio of overt NPs per argument position (= Referential Density or RD):

Case-based AGR → high RD

GR-Grammatik

- 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**

- 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-niūa tis-e-**ga** i?
 tea.NOM **2sGEN** 2sPOSS-mind be.easy-PST-**2s** Q

GR-Grammatik und andere Faktoren

RD effects from case-based AGR can be visible only in languages

- where overt NPs are never syntactically required.
- where RD is not tied to sociological factors, e.g.

social networks ~ RD or literacy ~ RD

(Other typological factors were ruled out in a previous study:
presence of reference-tracking devices, biases in lexical choice: *in*
Lg 79 (2003))

GR-Grammatik und andere Faktoren

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

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

depending on other factors, such as, indeed, sociological factors.

→ **Analyze interaction effects**

→ **Stratify sample for loose vs. close social network and for literacy**

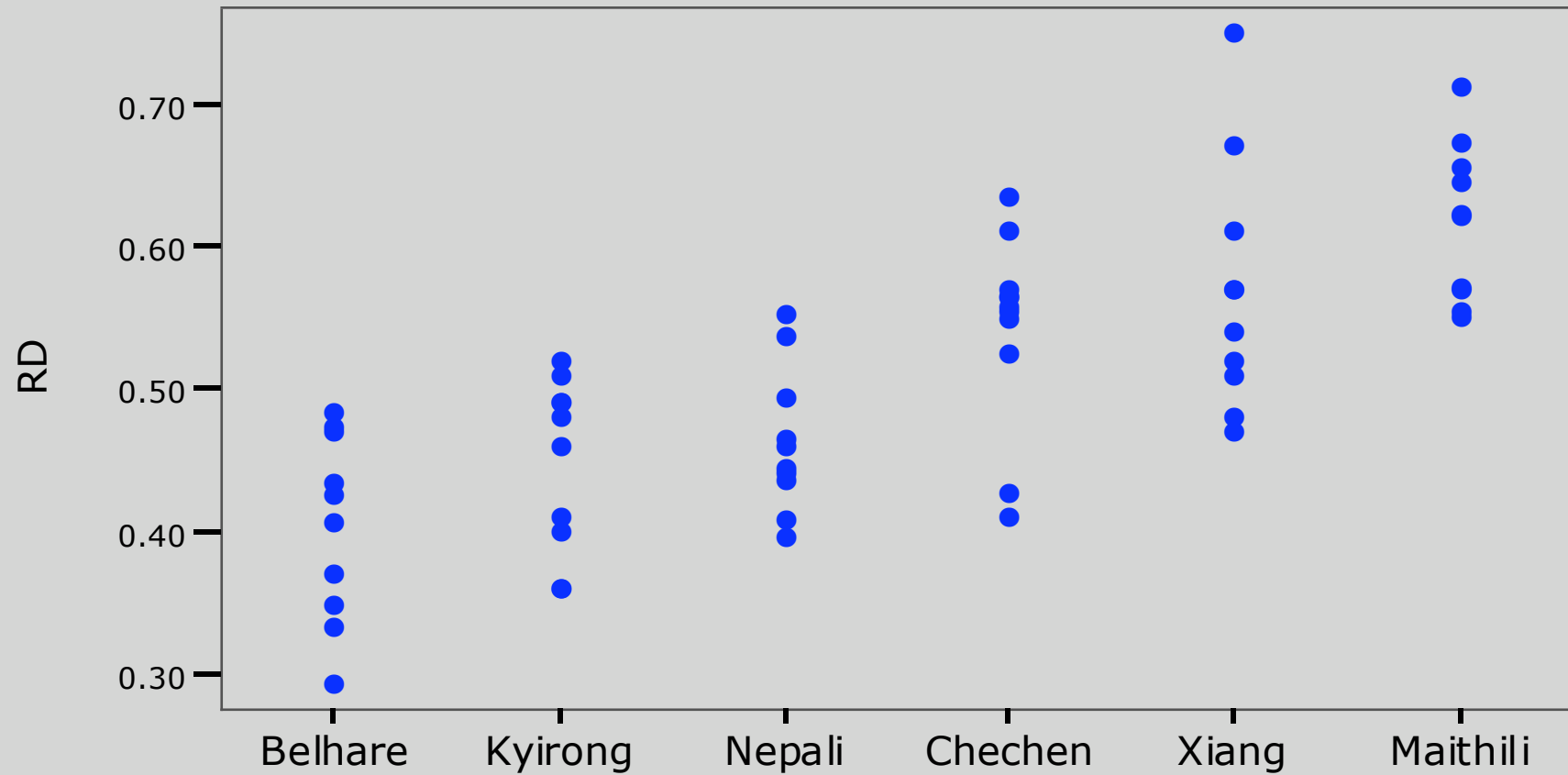
Sample

Language	Stock	Place	CASE-AGR	Subjects' social network	RD predicted
Belhare	ST	Nepal	-	close, rural	low
Kyirong			-	close, urban	low
Xiang		China	-	loose, urban	high
Nepali	IE	Nepal	+	close, urban	high
Maithili			+	close, urban	high
Chechen	ND	Cauc.	+	loose, urban	high

Cross-community variables:

- Literacy
- Text Length

Overall results



Community vs. Individual: $F(1,61) = 15.55, p < .001$

Referential Density (RD)

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

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

Referential Density (RD)

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

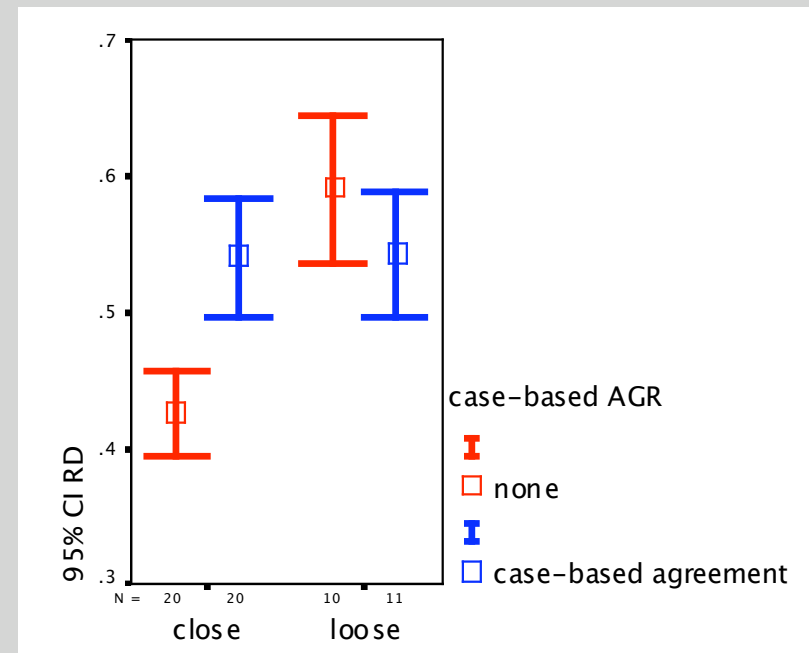
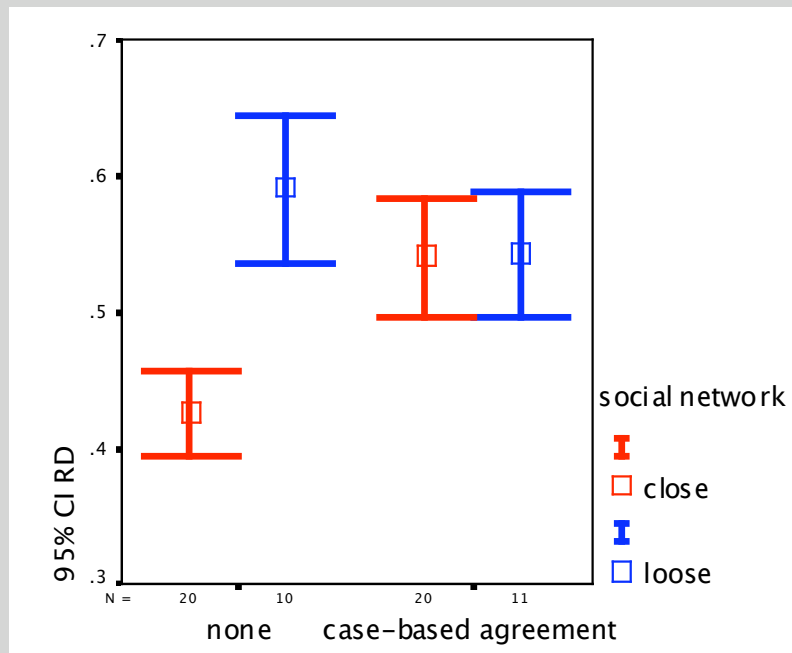
u ām toir-ke ṭokari me rakhne jāi chelai.
s/he mango having.plucked basket in to.keep going was

omaharse egoṭā chaurā elai,
and.then one boy came

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

Factorial analysis

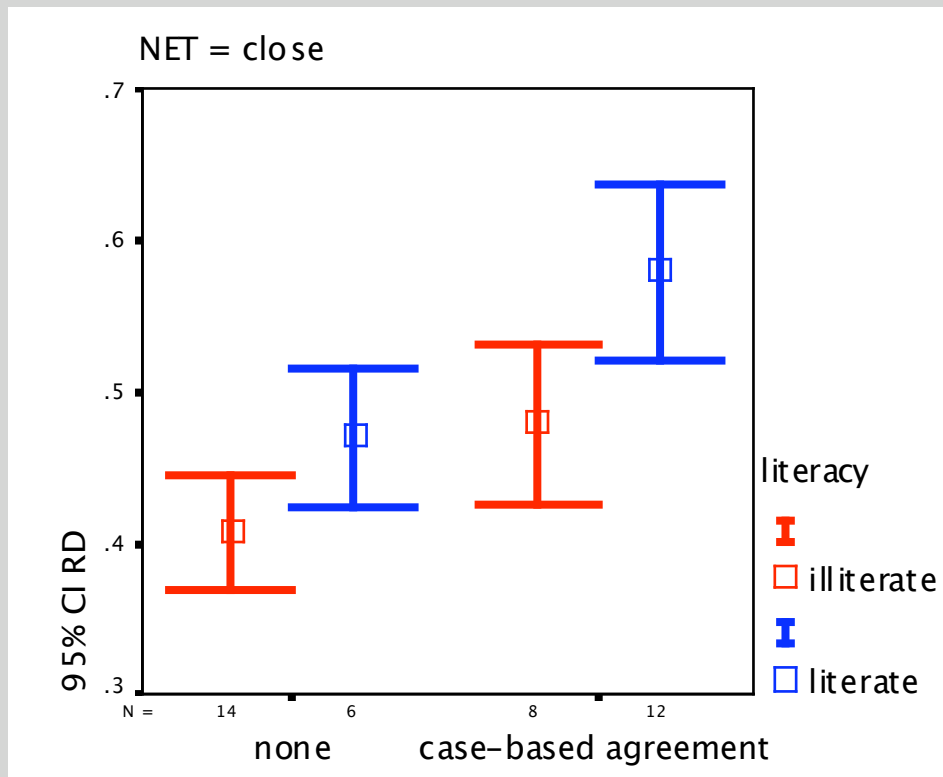
- Text length had no significant effect in a 4-way ANOVA (SYN * NET * LIT * LENGTH)
- 3-way ANOVA (SYN * NET * LIT):
 - main effect of LIT ($F(61,1) = 11.75, p = .001$)
 - interaction effect of SYN * NET ($F(61,1) = 11.22, p = .001$)



Factorial analysis (cont'd)

2-way ANOVA (SYN*LIT) under the **NET = close condition**

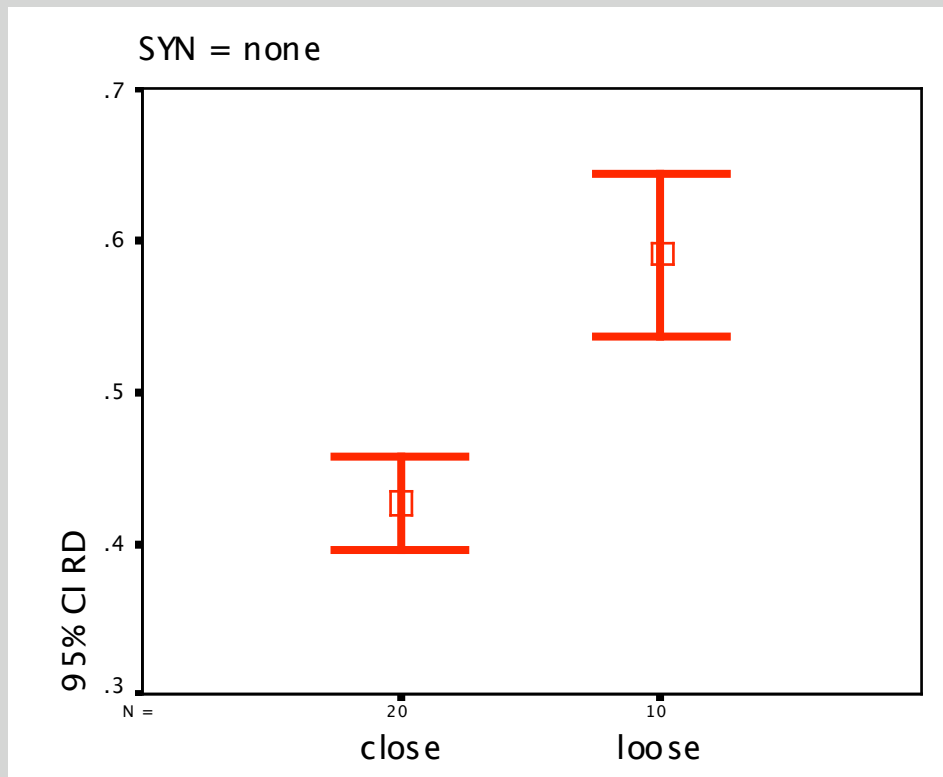
- main effect of SYN ($F(1,40) = 14.68, p < .001, \text{contrast} = .08^{**}$)
- main effect of LIT ($F(1,40) = 11.72, p = .002, \text{contrast} = .09^*$)



Factorial analysis (cont'd)

1-way ANOVA (NET) under the **SYN = none condition** (loose NET strictly implies LIT; hence LIT is excluded here)

- main effect of NET ($F(1,30) = 38.68, p < .001, \text{contrast} = .16^{**}$)



GR-Grammatik und RD

Conclusions from the RD experiment

1. In the absence of sociological factors leading to high RD (honorifics or loose social network), the frequent use of an NP-priming construction (case-based AGR) systematically increases RD.
2. This effect suggests that typological differences in agreement syntax directly affect discourse style.
3. Small effects during processing can yield large cultural differences, perhaps even beyond discourse!

Konklusionen

- Es gibt interessante typologische Variablen in Diskurspräferenzen
- Wir kennen erst sehr wenige davon:
 - PFAD
 - REFERENZDICHTHE
 - RAUMKOORDINATEN
 - DIREKTE REDE
 - VALENZPRÄFERENZEN (Bickel 2003, Nichols 2004)
- Aber...

Konklusionen

- diese Variablen sind wichtig, weil
 - sie universalistische diskursbasierte Erklärungen von Grammatikpräferenzen in Frage stellen
 - Evidenz für Relativitätseffekte im Diskurs liefern, möglicherweise auch darüber hinaus (*Thinking for Speaking*)
 - sie präzisere und zugleich realistischere typologische Messverfahren definieren
 - ihre Verteilung statistisch besser auswertbar sind als unsere traditionelle binären Parameter
 - sie zur komputationelle Sprachidentifikation beitragen (?)