

A crosslinguistic investigation of referential density

Michael Noonan

University of Wisconsin-Milwaukee

(1) *Definition of referential density*

The percentage of overtly expressed arguments to possible [*i.e.* notional] arguments is referred to as ‘referential density’ [RD].

(2) *Baragaunle text*

- a. **kyi cik dan tu cik hœt-a rñak dñanpo dñanpo**
dog one and child one COP-NOM COP before before
‘There was once a dog and a child.’
- b. **deni bñalwa cik sœ-a rñak**
then frog one tend-NOM COP
‘[They] tended a frog.’
- c. **deni bñalwa-ku thœn-ti**
then frog-DET emerge-SEQ
- d. **so-ri**
go-SEQ
‘Then the frog came out [of the bottle] and [it] went.’
- e. **tshemu thœn-ti**
night emerge-SEQ
- f. **so-ri**
go-SEQ
‘one night, [the frog] came out [of the bottle] and [it] went.’
- g. **kyi pñal-ti**
dog sleep-SEQ
- h. **de-ti galan ni**
stay-SEQ time FOCUS
- i. **bñalwa-ku thœn-ti**
frog-DET emerge-SEQ
- j. **so-a rñak**
go-NOM COP
‘As the dog slept and [it] was still, the frog came out [of the bottle] and [it] went.’
- k. **deni bñalwa-ku thœn-ti**
then frog-DET emerge-SEQ
- l. **so-ri**
go-SEQ
- m. **tsal-tu**
search-PUR
- n. **so-a rñak**
go-NOM COP

'The frog having come out [of the bottle] and [it] went, [they], in order to search for [it], [they] went.'

o. **kyi-ku daŋ tu-ku tsal-tu**
dog-DET and child-DET search-PUR

p. **so-ri ni**
go-SEQ FOCUS

q. **ma-ŋet-a**
NEG-find-NOM

r. **tsal-tu**
search-PUR

s. **so-a rŋak**
go-NOM COP

'The dog and child, in order to search [for it], [they] went, [they] didn't find [it], [they] in order to search for [it], [they] went.'

(3) goals of the pilot study

- a. to find a crosslinguistically valid way to measure RD, one that could be applied mechanically by trained researchers;
- b. to sample major language families and linguistic areas, with a more detailed sampling of a single area/family;
- c. to determine if RD values are relatively consistent within a L, within a given genre type, etc. [*i.e.* if a range of RD values can be attributed to a L]; to determine appropriate sample size to determine characteristic RD for a given L in a given genre, etc.;
- d. to determine the range of characteristic RD values across languages; to determine what is 'high' RD and what is 'low' RD;
- e. to find what, if any, morphosyntactic features correlate with RD values;
- f. to determine the extent to which a propensity for particular RD values can be borrowed *vs* the extent to which they are relatively stable features of particular language families;
- g. to determine if the *Frog Story* narration, or some alternative mode of eliciting discourse, is an appropriate tool for measuring RD across languages:
 - i. Are FSs suitable [*vis-à-vis* some other technique]?
 - ii. Is *any* such device likely to provide a characteristic sample of RD values for a L? [*i.e.* are the narrative conventions employed in producing FS [or some similar] narrations similar enough to 'natural' discourse conditions to attribute the values obtained from FSs to the language?]

(4) URL for the *RD Handbook*:

<http://www.uwm.edu/~noonan/ReferentialDensity.html/>

(5) *What was measured*

RD1 measures directly the zero-anaphora phenomenon: the lower the number, the more zero-anaphora the language indulges in. A value of 1.00 would indicate that there were no instances of zero-anaphora – every possible argument was given nominal [or pronominal] expression; a value of 0.00 would indicate that there were no overt nominal references in the text.

RD1+ is a fuller measure of nominal reference in that it takes into account agreement phenomena in measuring nominal reference: verb-argument concord markers said in the absence of an overt noun/pronoun are added to the number of overt nominal arguments and then compared to possible argument slots. Here, too, possible values range between 1.00 and 0.00.

RD2 measures the verbiness/nouniness of the language: a number below 1.00 indicates that verbs [= semantic predicates] outnumber overt nominal arguments in discourse; a number above 1.00 indicates that overt nominal arguments outnumber verbs in discourse.

RD2+ is a measure of verbs [= semantic predicates] to any overt reference to arguments: a number below 1.00 indicates that verbs outnumber overt nominal argument references in discourse; a number above 1.00 indicates that overt nominal argument references outnumber verbs in discourse.

(6) Hypotheses

No single factor is responsible for RD values. Relevant factors are:

1. individual preferences and goals
2. rhetorical styles, which are inevitably tied to
3. genre-type
4. morphosyntactic factors

(7) Morphosyntactic factors: assumptions & hypotheses

The last, morphosyntactic factors, predispose languages to higher or lower RD values:

Assumptions:

(a) Factors that lower RD1 have the effect of increasing the ratio of Vs to Ns, increasing the prominence of the internal structure of the event itself over the participants in the event.

(b) Factors that heighten RD1 have the effect of increasing the ratio of Ns to Vs, increasing the prominence of the participants in the event over the internal structure of the event itself.

(c) In discourse, speakers must divide their attention between the internal structure of the event and the participants in the event. Increased prominence of the internal structure of the event will result in increased attention to the event itself, resulting in a higher frequency of verbs in discourse; increased prominence of participants in the

event will result in increased attention to participants, resulting in a higher frequency of nouns in discourse.

Noun-prominence & Verb-prominence:

Languages can be characterized as being relatively *noun-prominent* [participant-prominent] or relatively *verb-prominent* [event-prominent].

General Hypothesis:

Languages which have heightening factors but no lowering ones will have relatively high RD values; languages which have both heightening and lowering factors will have moderate RD values; languages which have only lowering factors will have low RD values.

Specific Hypotheses

(i) Case-sensitive PSAs [Bickel's (2003) Hypothesis]:

The more constructions in a language are subject to case-sensitive PSAs [= 'privileged syntactic arguments' = controllers & pivots], the more often are case frames activated in syntactic processing and because case morphology is hosted by and structurally associated with NPs, speakers become primed for using overt NPs more often in discourse. Other things being equal, the expected result is a higher degree of referential density.

(ii) Syntactic PSAs

The more constructions a language has which involve syntactic PSAs, the more often argument arrays are activated in syntactic processing. Other things being equal, the expected result is a higher degree of referential density.

(iii) Verbal expression of core schema and supporting event:

If, for complex events, the core schema and the supporting event are both mapped onto verbals within a single, complex sentence, then, other things being equal, the expected result is a lower degree of referential density.

(iv) Verbal expression of spatial deixis:

Frequent use of verbs for spatial deixis results, other things being equal, in a lower degree of referential density.

[Note: there are many other construction-types that have the effect of lowering RD – e.g. the obligatory non-occurrence of coreferential arguments in clause combining, such as what was traditionally called equi-NP-deletion – but such constructions occur too infrequently in the sort of narrative discourse investigated in this study to have much impact on overall RD values.]

(8) Case-sensitive PSAs [Bickel's (2003) Hypothesis]:

“There are two crucial assumptions behind this hypothesis: (i) the assumption that speaking a language with many case-sensitive PSAs leads to frequent activation of case

frames in syntactic processing; (ii) the assumption that frequent case frame activation habituates speakers into using overt NPs in discourse.” [Bickel 2003]

(9) *Syntactic PSAs:*

The assumption here is that syntactic pivots and controllers require the activation of the relevant arguments in syntactic processing and that frequent activation of these arguments leads to speakers using more overt NPs in discourse.

(10) *Verbal expression of core schema and supporting event:*

This would apply to a subset of ‘verb-framed languages’ and ‘equipollantly-framed languages’ [in Slobin’s (2003) sense]. Because a single complex event is being described and shared arguments are not repeated among the two clause-mate verbals, this reduces RD by increasing the ratio of verbs to nouns.

(11) Examples illustrating mapping of the core schema and the supporting event onto verbals within a single, complex sentence follow:

Chantyal

- (a) **na-sə marak-ra latti-sə lhi-si-rə bən la-si pin-ji**
 I-ERG door-DAT kick-INST hit-ANT-SEQ closed do-ANT give-PERF
 ‘I kicked the door shut’

[KICK = supporting event, CLOSE = core schema] [The main verb **pin-ji** ‘gave’ signals that the event was done to the door’s detriment.]

- (b) **khi ghāti-ri fiar həlji-si-rə si-i**
 he throat-LOC bone get.stuck-ANT-SEQ die-PERF
 ‘He choked to death on a bone’

[CHOKER = supporting event, DIE = core schema]

- (c) **na-sə məynbətti phur-si-rə bətti sar-ji**
 I-ERG candle blow-ANT-SEQ light kill-PERF
 ‘I blew the candle out’

[BLOW = supporting event; EXTINGUISH/GO OUT = core schema]

- (d) **na-sə kəmes khur-si-rə səffa la-i**
 I-ERG shirt wash-ANT-SEQ clean do-PERF
 ‘I washed the shirt clean’

[WASH = supporting event; MAKE CLEAN = core schema]

- (e) **na-sə phurbal latti-sə lhi-si-rə bakəs-nfiari wō-i**
 I-ERG ball kick-INST hit-ANT-SEQ box-INES enter-PERF
 ‘I kicked the ball into the box’

[KICK = supporting event; ENTER = core schema]

- (f) **na kadmandu-ri pfiar-gəy fiya-i**
 I Kathmandu-LOC walk-PROG go-PERF
 ‘I walked to Kathmandu’

[WALK = supporting event; GO = core schema]

(12) *Verbal expression of spatial deixis:*

Because a single complex event is being described and shared arguments are not repeated among the two clause-mate verbals, this reduces RD by increasing the ratio of verbs to nouns.

(13) Below are some examples illustrating the verbal expression of spatial deixis:

Nar-Phu

- (a) **târ kyû pi ni-w**
 in.order run move.fast go-IMP
 'Run in order [away from speaker]!'

[Note: *târ kyû pi-taw* is also possible, but doesn't specify direction.]
- (b) **târ kyû pi kha-w**
 in.order run move.fast come-IMP
 'Run in order [toward speaker]!'
- (c) **târ kyû pi tû-w**
 in.order run move.fast stay-IMP
 'Run in order in place [*i.e.* as in an army exercise]!'
- (d) **kræ kûr tû-w**
 head bend stay-IMP
 'Bow down!'
- (e) **thîm-re thô waŋ ni-se mû**
 house-DL arrive enter go-SEQ be
 '[She] arrived and entered the house [moving away from the speaker]'
- (f) **phâlpe cə baltin cãŋpe nhân-re cfaŋ-se thæ-se mû**
 toad DEF container small inside-DL put.in-SEQ keep-SEQ be
 'They put the toad in the small container so as to keep him there.'

(14) [Somewhat arbitrary] ranking of RD1 [and RD1+] values

<0.50 = low

0.50 – 0.70 = moderate

0.70 > = high

(15) *The 4 hypotheses and selected languages:*

	(i) CS-PSAs	(ii) S-PSAs	(iii) VECSSSE	(iv) VESD	RD1	RD1+
English	-	+	-	-	high	
Russian	+	+	-	-	high	high
Japanese	-	+	+	-	mod	
Nepali	+	+	+	-	mod	high
Lango	-	+	-	-	low	high
Yessan-Mayo	-	- (?)	?	+	low	
Nar-Phu	-	-	+	+	low	low



References

- Bickel, Balthasar. 2003. 'Referential density in discourse and syntactic typology.' *Language* 79/4.
- Noonan, Michael. 2003a. 'Motion events in Chantyal.' In Erin Shay and Uwe Siebert, eds. *Motion, Direction, and Location in Languages*. Amsterdam: John Benjamins. 2003.
- Noonan, Michael. 2003b. 'Recent language contact in the Nepal Himalaya.' In David Bradley, Randy LaPolla, Boyd Michailovsky & Graham Thurgood, eds. *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in Honour of James A. Matisoff*. Canberra: Pacific Linguistics.
- Noonan, Michael. *ms.* 'Direct speech as a rhetorical style in Chantyal.'
<http://www.uwm.edu/People/noonan/Papers.html>
- Slobin, Dan I. 2003. 'The many ways to search for a frog: linguistic typology and the expression of motion events.' In S. Strömquist & L. Verhoeven, eds. *Relating Events in Narrative: Typological Perspectives*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Talmy, Leonard. 1985. 'Lexicalization patterns: semantic structure in lexical form.' In Timothy Shopen, ed. *Language Typology and Syntactic Description*, Vol 3. Cambridge: Cambridge University Press. Pp36-149.
- Talmy, Leonard. 1991. 'Path to realization: a typology of even conflation.' *Proceedings of the Seventeenth Annual Berkeley Linguistics Society*, Pp480-519.

Sources of Texts

Anywa

Mechtild Reh. 1996. *Anywa Language*. Rüdiger Köppe: Köln.

Apuriña

Da Silva Sidney Facundes. 2000. *The Language of the Apuriña People of Brazil (Maipure/Arawak)*. State University of New York at Buffalo PhD dissertation.

Arrernte

David Wilkins

Baragaunle

Mary Brehm

Chantyal

Michael Noonan. 1999. *Chantyal Dictionary and Texts*. Berlin: Mouton de Gruyter

Dyirbal

R.M.W. Dixon. 1972. *The Dyirbal Language of North Queensland*. Cambridge University Press: Cambridge.

English

Michael Noonan

Fijian

R.M.W. Dixon. 1988. *A Grammar of Boumaa Fijian*. The University of Chicago Press: Chicago and London.

French

Michael Noonan

German

Sabine Stoll

Ghale

Holly Smith

Japanese

Shigekazu Hasegawa, Chikako Sakurai

Kilivila

Gunter Senft; Gunter Senft. 1986. *Kilivila: The Language of the Trobriand Islanders*. Berlin: Mouton de Gruyter.

Lango

Michael Noonan. 1992. *A Grammar of Lango*. Berlin: Mouton de Gruyter.

Limbu

George van Driem. 1987. *A Grammar of Limbu*. Berlin and New York: Mouton de Gruyter.

Lushootseed

Vi Hilbert & Thom Hess, in Barry F. Carlson, ed. 1977. *Northwest Coast Texts*. Native American Text Series, Vol.2, N. 3

Magar

Karen Grunow-Harsta

Manange

Kristine Hildebrandt

Meithei

Shobhana Chelliah. 1997. *A Grammar of Meithei*. Berlin: Mouton de Gruyter.

Mixtec

Monica Macaulay. 1996. *A Grammar of Calcatongo Mixtec*. Berkeley: University of California Press

Nar-Phu

Michael Noonan

Nepali

Michael Noonan

Qiang

Randy LaPolla. 1996. *Grammatical Sketch of the Qiang Language with Texts and Annotated Glossary*. Ms City University of Hong Kong.

Quechua

David Weber. 1989. *A Grammar of Huallaga (Huanuco) Quechua*. Berkeley: University of California Press.

Russian

Aida Martinovic-Zic

Sinhala

Phoebe Vitharana

Spanish

Michael Noonan

Squamish

Aert Kuipers. 1967. *The Squamish Language*. The Hague: Mouton.

Supyire

Robert Carlson. 1994. *A Grammar of Supyire*. Berlin: Mouton de Gruyter.

Taba

John Bowden. 1997. *Taba (Makian Dalam): Description of an Austronesian Language of Eastern Indonesian*. University of Melbourne PhD dissertation.

Tigre

Shlomo Raz. 1983. *Tigre Grammar and Texts*. Malibu: Undena Publications.

Wikchamni

Geoffrey Gamble. 1978. *Wikchamni Grammar*. Berkeley: University of California Press.

Yagua

Thomas Payne. 1993. *The Twin Stories: Participant Coding in Yagua Narrative*. Berkeley: University of California Press.

Yessan-Mayo

Velma Foreman. 1974. *Grammar of Yessan-Mayo*. Santa Ana: Summer Institute of Linguistics.

Yimas

William Foley. 1991. *The Yimas Language of New Guinea*. Stanford: Stanford University Press.

Yukaghir

Elena Maslova. 1999. *A Grammar of Kolyma Yukaghir*. Ms.

LANGUAGE	RD1	RD1+	RD2	RD2+
Anywa	0.59	0.77	1.33	1.48
Apuriña	0.52	0.90	0.99	1.72
Arrernte FS	0.76	0.84	1.62	1.79
Baragaunle FS1	0.43		0.85	
Baragaunle FS2	0.40		0.80	
Baragaunle 3	0.40		0.83	
Chantyal FS1	0.53		0.98	
Chantyal FS2	0.60		1.13	
Chantyal 3	0.47		0.90	
Chantyal 4	0.44		0.83	
Chantyal 5	0.49		1.04	
Chantyal 6	0.49		0.97	
Dyirbal	0.61		1.23	
English FS1	0.84		1.60	
English FS2	0.81		1.51	
English FS3	0.80		1.60	
English FS4	0.83		1.62	
Fijian	0.62	0.64	1.21	1.25
French FS	0.76/0.68	0.84	1.54/1.38	1.71
German FS	0.80	0.92	1.59	1.84
Ghale FS1	0.60		1.20	
Ghale FS2	0.58		1.19	
Ghale FS3	0.69		1.47	
Japanese FS1	0.67		1.29	
Japanese FS2	0.68		1.31	
Japanese FS3	0.50		0.95	
Japanese FS4	0.53		1.03	
Japanese FS5	0.60		1.15	
Japanese FS6	0.66		1.33	
Kilivila FS1	0.26	0.63	0.51	1.26
<i>Kilivila 2</i>	0.55	0.86	1.09	1.74
Kilivila FS3	0.31	0.70	0.63	1.41
Kilivila FS4	0.34	0.68	0.68	1.36
Lango 1	0.48	0.78	0.92	1.52
Lango 2	0.38	0.72	0.79	1.50
Lango 3	0.38	0.72	0.77	1.46
Lango 4	0.34	0.73	0.68	1.47
<i>Lango 5</i>	0.66	0.80	1.35	1.54
Limbu	0.52	0.67	0.94	1.21
Lushootseed	0.52	0.64	0.87	1.07
Magar FS1	0.49		1.00	
Magar FS2	0.54		1.13	
Manange FS1	0.34		0.68	
Manange FS2	0.42		0.85	
Meithei	0.57		1.20	
Mixtec	0.50	0.79	0.97	1.52
Nar-Phu FS1	0.36		0.72	
Nar-Phu 2	0.44		0.87	
<i>Nar-Phu 3 [written]</i>	0.65		1.21	
<i>Nar-Phu 4 [non-nar]</i>	0.58		1.06	
Nepali	0.62	0.75	1.16	1.41
Qiang	0.40	0.43	0.79	0.84
Quechua	0.51	0.69	0.91	1.23
Russian FS	0.71	0.83	1.31	1.54
Sinhala FS	0.65		1.37	
Spanish	0.74/0.65	0.82	1.44/1.31	1.55
Squamish	0.53	0.77	0.99	1.45
Supyire	0.61		1.21	
<i>Taba</i>	0.52	0.82	0.97	1.53
Tigre	0.39	0.89	0.69	1.57
Wikchamni	0.62		1.21	
Yagua	0.41	0.80	0.78	1.54
Yessan-Mayo	0.43		0.88	
Yimas	0.38	0.73	0.68	1.31
Yukaghir	0.40	0.71	0.76	1.34

1 = RD1; 1+ = RD1+; (1) = RD1 for written, non-narrative, or non-third-person texts; 2 = two speakers with the same score; FS = Frog Story Narration; subscript letters, 1_a, indicate coindexing

	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				
Manange FS					1				1																																	
Nar-Phu FS						1																																				
Nar-Phu										1									(1)		(1)																					
Baragaunle FS								1	1																																	
Baragaunle								1																																		
Chantyal FS														1						1																						
Chantyal										1	1	2																														
Magar FS												1			1																											
Limbu													1																													
Ghale FS																				1	1																					
Qiang	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				
Meithei																	1																									
Nepali	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				
Sinhala FS																			1									1+														
Russian FS																									1																	
Spanish FS																					1 _a					1 _a																
French FS																						1 _a				1 _a																
German FS																													1													1+
English FS																																										
Japanese FS	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				
Yukaghir													1	1																												
Tigre	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				
Lango																																										
Anwya																																										
Supyire	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				
Dyribal																																										
Arrernte FS																																										
Yimas																																										
Yessan-Mayo																																										
Kilivila FS	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				
Kilivila	1 _a		1 _b	1 _c																																						
Taba																																										
Fijian																																										
Lushootseed	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00				

Squamish	1														1+																							
Wikchamni	1														1																							
Mixtec	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00
Yagua	1														1+																							
Quechua	1														1+																							
Apuríña	1														1+																							
	.26	.28	.30	.32	.34	.36	.38	.40	.42	.44	.46	.48	.50	.52	.54	.56	.58	.60	.62	.64	.66	.68	.70	.72	.74	.76	.78	.80	.82	.84	.86	.88	.90	.92	.94	.96	.98	1.00