Modeling the emergence of pronominal paradigms

Sander Lestrade

Centre for Language Studies, Radboud University Nijmegen

s.lestrade@let.ru.nl

Pronoun development

phase 1		phase 2		phase 3	phase 4
N_A	⇒	pro_A	⇒	VM	VM
N_U	⇒	NM_U		NM_U	NM_U
		N_A	⇒	pro_A	**pro_A**
				pro_A + NM_U →	**pro_U**

Phase 1

All items are maximally long and fully specified, and role marking is ad hoc:

(1) a. *atolona otemoto tonenar*
1 otemoto.V tonenar
'I otemoto tonenar.'

b. *atolona* ***uturera*** *otemoto tonenar*
I **otemotoee** otemoto.V tonenar
'Tonenar otemotoes me.'

form	D1	D2	D3	D4	D5	D6	D7	D8	D9
atolono	1	1	1	1	1	.625	.75	.875	.625
uturera	1	1	0	0	1	.75	.75	.625	.125

Check: make sure you're understood:

(Grice 1975, Levelt 1983; Aristar 1997, de Swart 2011, Lestrade 2010)

(2) Apatani (after Abraham 1985: 38-40)

a. *mó sɨhɨnɨ pabɨne*
3SG cow killed
'He killed the cow.'

b. *sɨhɨnɨ mó* ***mi*** *alitubɨne*
cow 3SG **DAT** kicked
'The cow kicked him.'

Proximity: stand together, belong together:

(Givón 1995)

(3) *book yellow and man big*
'yellow book and big man' (not other way around)

Recruitment: extended use of available means:

(Zeevat 2007; Palancar 2002)

(4) *boy girl* ***from*** *hit*
'The girl hit the boy.' ("hitting is coming from girl")

Phase 2

Grammaticalization of pronouns and (incipient) case markers:

(5) ***ato suron*** *lonetan enarete*
1 U lonetan.V enarete
'Enarete lonetans me.'

form	D1	D2	D3	D4	D5	D6	D7	D8	D9
ato			1		1				
suron	0	0	0	0	0			.125	

Reduction: shortening of automatized forms:

(Jurafsky et al. 2001, Heine and Kuteva 2007)

(6) *eigenlijk* → *eik* 'actually'

Erosion and ***Bleaching***: actual storage of short forms and loss of meaning specificity:

(Nettle 1999; Bybee 2010, Heine and Kuteva 2007)

(7) *ēwa-haft* 'century-like' ⇒ *echt* 'real'

(van der Sijs 2010)

Modeling meaning

CAT: entity=T, animate=T, . . . , covering=fur, legs=4.

(Guiraud 1968, Wierzbicka 1996, Gärdenfors 2000)

form	D1	D2	D3	D4	D5	D6	D7	D8	D9
atadoso	1.00	0.00	1.00	1.00	0.00	0.75	0.25	1.00	1.00
nimator	1.00	1.00	0.00	1.00	1.00	0.38	0.38	0.62	0.88
umimota	1.00	1.00	0.00	0.00	0.00	0.62	0.50	0.25	0.62
isomera	1.00	0.00	1.00	0.00	1.00	1.00	0.12	1.00	0.62
enolate	0.00	0.00	1.00	1.00	1.00	0.00	0.25	0.75	0.00

Table 3: *First entries of noun lexicon (abbreviated)*

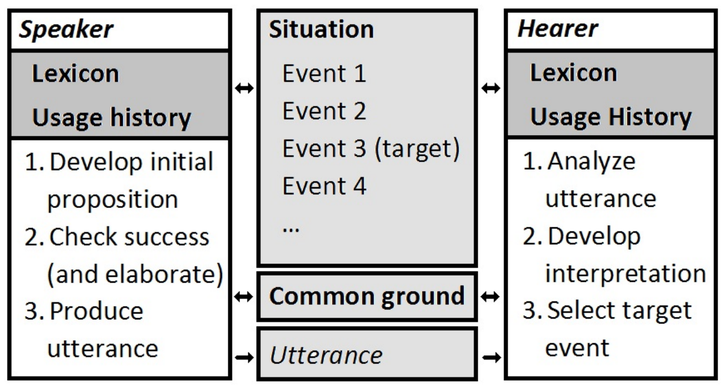

Figure 1: *Conversation procedure*

(Grice 70s, Levelt 80s, Steels 90s)

Phase 3

Development of verb marking and recruitment of new pronouns:

(8) ***atusom*** *usenama asumuro-****am***
1 usenama asumuro.V-**1**
'I asumoro usenama.'

form	D1	D2	D3	D4	D5	D6	D7	D8	D9
am			1		1				
atusom	1	1	1	1	1		1	.1	.375

Fusion: attachment of attenuated forms:

(Bybee 1985)

(9) *ik loop* ⇝ *'k=loop* 'I walk'

Referability: support attenuated pronouns:

(Ariel 1999)

(10) French (after Heine & Kuteva, 2002, 234)

a. *La jeune [. . .]. Elle est danseuse.*
'The girl [. . .]. She is a dancer.'

b. ***Ma femme*** *il=est venu.*
my:F wife 3=is come
'My wife has come.'

Note

Higher values are more prominent (for nouns) or agent-like (for roles) than lower values. Being prominent means being a good agent, and vice versa, and similarly for non-prominent undergoers.

HIT: activity=T, duration=F, . . . , participants=2;
external: cause=T, volitional=T, . . . , means=hands;
internal: affected=T, control=F, . . . , result=pain.

form	type	D1	. . .	Ext1	. . .	Int1	. . .	Int9
rirunes	twoPlace	1	. . .	1	. . .	0	. . .	.00
amumali	twoPlace	1	. . .	1	. . .	0	. . .	1
emimano	twoPlace	1	. . .	0	. . .	1	. . .	.62
litaril	onePlace	0	. . .	0	. . .			
adasumu	twoPlace	1	. . .	1	. . .	0	. . .	.25

Table 4: *First entries of verb lexicon (abbreviated)*

V1	. . .	V9	A1	. . .	A9	123	U1	. . .	U9	123	tgt
1	. . .	.625	0	. . .	.375	3					0
1	. . .	.000	1	. . .	.875	3	1	. . .	.375	3	0
0	. . .	.625	1	. . .	.375	3	1	. . .	.250	3	0
0	**. . .**	**.875**	**1**	**. . .**	**.000**	**3**	**1**	**. . .**	**.750**	**1**	**1**
1	. . .	.375	1	. . .	.250	3	0	. . .	.375	3	0

Table 5: *First events of a situation (abbreviated)*

form	D1	D2	D3	D4	D5	D6	D7	D8	D9
atalama	0	1	0	1	0	.625	.750	.250	.625
surones	**0**	**0**	**0**	**0**	**0**	**.375**	**.500**	**.375**	**.625**
surones	**0**	**0**	**0**	**0**	**0**	**.000**	**.875**	**.000**	**.125**
unarema	0	0	1	1	1	.375	.250	.500	.750
sorusel	1	1	1	0	1	.750	.250	.000	.625

Table 6: *First entries of usage history*

Phase 4

Fusion of frequent combination of undergoer marker and actor pronoun:

(11) a. ***usalulu*** *semolom-am osetono*
1U semolom.V-1 osetono
'Osetono semoloms me.'

b. *amasu nonemeraa-am nelusum*
1A nomereraa.V-1 nelusum
'I nomereraa nelusum.'

c. *nemurol asutule-su asatole*
nemurol asutule-U asatole.V
'Nemurol asatoles asutule.'

form	D1	D2	D3	D4	D5	D6	D7	D8	D9
am			1		1				
amasul	1	1							
usalulu	0	0	0	0	0				
su	0	0	0	0	0				

Conclusion

Simulating language evolution in a computational model, it can be shown how:

⇝ pronouns (and case markers) develop from lexical expressions by erosion and bleaching,
⇝ case markers fuse with pronouns resulting in pronominal paradigms, and
⇝ attenuated pronouns fuse with the verb and recruit supporting expressions resulting in (local) verb agreement.

References and conventions

P.T. **Abraham** (1985), *Apatani grammar*. M. **Ariel** (1999), "The development of person agreement markers: From pronouns to higher accessibility markers". In *Usage based Models of Language*. A.R. **Aristar** (1997), "Marking and hierarchy. Types and the grammaticalization of case markers". *Studies in Language* 21 (2). D. **Bickerton** (1981), *Roots of language*. J. **Bybee** (2010), *Language, Usage and Cognition*. J. **Bybee** (1985), *Morphology. A study of the relation between meaning and form*. P. **Gärdenfors** (2000), *Conceptual Spaces: The Geometry of Thought*. T. **Givón** (1995), *Functionalism and grammar*. H.P. **Grice** (1975), "Logic and conversation". In *Syntax and semantics: Speech acts* 3. P. **Guiraud** (1968), "The semic matrices of meaning". *Social Science Information* 7 (2). B. **Heine** and T. **Kuteva** (2007), *The genesis of grammar. A reconstruction*. B. **Heine** and T. **Kuteva** (2002), *World lexicon of grammaticalization*. R. **Jackendoff** (2002), *Foundations of language: Brain, meaning, grammar, evolution*. D. **Jurafsky** et al. (2001), "Probabilistic relations between words: Evidence from reduction in lexical production". In *Frequency and the emergence of linguistic structure*. S. **Lestrade** (2010), *The space of case*, Ph.D. dissertation, Radboud University Nijmegen. W. **Levelt** (1983), "Monitoring and self-repair in speech". *Cognition* 14. D. **Nettle** (1999), *Linguistic Diversity*. E.L. **Palancar** (2002), *The origin of agent markers*. **R Core Team** (2014), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. N. van der **Sijs** (ed) (2010), *Etymologiebank*. L. **Steels** (1997), "Constructing and sharing perceptual distinctions". *Machine Learning*, ECML-97. P. de **Swart** (2011), "Sense and simplicity: Bidirectionality in differential case marking". In *Bidirectional Optimality Theory*. A. **Wierzbicka** (1996), *Semantics: Primes and Universals*.

⇒ erosion & bleaching; → fusion; *123* person; *A* actor; *D1:9* meaning dimensions, *Ext1:9* ∼ external role, *Int1:9* ∼ internal role; DAT dative; *N* noun; *NM* noun marker; *pro* pronoun; SG singular; *tgt* target event; *U* undergoer; *V1:V9* qualities of actions, *A1:A9* ∼ actors, *U1:U9* ∼ undergoers; *VM* verb marker.

Radboud University