

Population Size and the Encoding of Thematic Roles

(abstract for oral presentation)

Recent studies have begun to explore possible correlations between the structural properties of languages and the number of their speakers. Among these studies, several have concerned themselves with a putative correlation between population size and structural complexity, the claim being that languages spoken by small populations tend to be more complex in various ways than their counterparts with larger numbers of speakers (Trudgill 1996, McWhorter 2007, Dahl 2009 and others). This paper provides further empirical evidence for a connection between population size and structural complexity within the domain of thematic role marking, but shows that the correlation is in fact in the opposite direction: languages with small populations of speakers tend to have less elaborated encoding of thematic roles than do languages with larger populations.

This paper presents the results of an ongoing cross-linguistic psycholinguistic experiment, in which subjects are asked to judge the truth conditions of written sentences in relationship to pictorial stimuli. The experiment focuses on languages with reported SVO basic word order, and examines the availability of two main classes of constructions: (a) *Bare-Peripheral*, in which an expression associated with a peripheral or oblique thematic role in relation to the activity-denoting word occurs without any overt expression of its thematic role, as in (1); and (b) *Bare-Patient-Preceding*, in which an expression referring to the patient of an activity occurs in front of the activity-denoting word but without any overt coding of its thematic role, as in (2). So far, the experiment has been conducted on close to 1900 subjects, speakers of 29 languages/dialects, shown in (3).

The results of the experiment reveal a substantial amount of cross-linguistic variation with respect to the availability of Bare-Peripheral and Bare-Patient-Preceding interpretations, ranging from languages such as English and Hebrew in which these interpretations are generally unavailable, to languages such as Ju|'hoan and Meyah in which such interpretations are readily available. How might one account for such variation? One important factor is grammatical: as argued elsewhere, languages with more obligatory encoding of categories such as number, definiteness, tense and aspect tend to be less tolerant of such interpretations (even if the categories in question bear no direct connection with thematic roles). However, grammatical factors alone do not account for all of the variation; it is also necessary to take into consideration sociolinguistic factors. Specifically, this paper argues that Bare-Peripheral and Bare-Patient-Preceding interpretations are more readily available in languages with small numbers of speakers.

The correlation between availability of Bare-Peripheral and Bare-Patient-Preceding interpretations and small population size emerges from a statistical examination of the 29 languages tested so far. In particular, it can be vividly observed in pairwise comparisons of languages that are genealogically and typologically related, while differing substantially with respect to population size: Table 1 shows three such language pairs, within each of which the language with fewer speakers demonstrates a greater availability of Bare-Peripheral and Bare-Patient-Preceding interpretations in the experimental results.

Thus, with respect to the marking of thematic roles, languages with smaller populations tend to exhibit lesser complexity than languages with larger numbers of speakers. The paper concludes with some functionally-oriented speculations on why this correlation should obtain; the basic idea is that the larger the population, the more difficult it is to retrieve information concerning thematic roles from the shared context of the utterance, and hence the greater the need for such information to be encoded formally.

- (1) Bare-Peripheral Test Stimulus
Sentence (English): *The clown is drinking the book*
Picture: clown drinking from a glass while reading a book
- (2) Bare-Patient-Preceding Test Stimulus
Sentence (English): *The dog is drawing*
Picture: man drawing a dog
- (3) Test Languages:
European: English, Hebrew
Creole: Papiamentu, Sranan, Bislama
West African: Twi, Fongbe, Yoruba
Khoisan: Ju'hoan
Mainland Southeast Asian: Thai, Lao, Vietnamese
Sinitic: Mandarin, Cantonese
Malay/Indonesian: Standard Indonesian, Kuala Lumpur Malay, Kuching Malay, Riau Indonesian, Siak Malay, Bengkulu, Jakarta Indonesian, Kupang Malay, Papuan Malay
West Malayo-Polynesian: Mentawai, Minangkabau, Sundanese
Central-Eastern Malayo-Polynesian: Nage, Roon
Papuan: Meyah

<i>Group</i>	<i>Language</i>	<i>Population</i>	<i>Availability (%) of Bare-Peripheral Interpretations</i>	<i>Availability (%) of Bare-Patient-Preceding Interpretations</i>
Sinitic	Mandarin	873,014,298	31	10
	Cantonese	54,810,598	42	12
Tai	Thai	20,229,987	48	11
	Lao	3,188,577	54	24
Malayic	St Indonesian	180,747,607	29	11
	Minangkabau	6,500,000	74	57

Table 1: *Pairwise Correlations between Population Size and the Availability of Bare-Peripheral and Bare-Patient-Preceding Interpretations*

REFERENCES

- Dahl, Östen (2009) "Testing the Assumption of Complexity Invariance: The Case of Elfdalian and Swedish", in G. Sampson, D. Gil and P. Trudgill eds., *Language Complexity as an Evolving Variable*, Oxford University Press, Oxford.
- McWhorter, John H. (2007) *Language Interrupted: Signs of Non-Native Acquisition in Standard Language*, Oxford University Press, New York.
- Trudgill, Peter (1996) "Dialect Typology: Isolation, Social Network and Phonological Structure", in G. Guy et al eds., *Towards a Social Science of Language: Papers in Honour of William Labov, Vol 1: Variation and Change in Language and Society*, John Benjamins, Amsterdam.