

Word order in zero-marking languages: A typological study

Kaius Sinnemäki

(kaius.sinnemaki@helsinki.fi)

University of Helsinki

ALT 8, Berkeley, July 23-26 2009

1. Introduction and background

A rather wide consensus that **zero marking of languages tend to correlate with SVO word order** (e.g. Lehmann 1978; Mallinson & Blake 1981; Jackendoff 1999). But:

- No typological studies dedicated to the topic. Studies from which data can be gleaned have about 10-20 zero-marking instances (e.g. Mallinson & Blake 1981; Siewieska & Bakker 1996), which is too few for for reliable conclusions.

Zero marking = the absence of overt morphological marking of core arguments of the predicate (Nichols and Bickel 2005). Core arguments = S, the more agent-like argument, and O, the more patient-like argument, of a two-place transitive predicate (Comrie 2005). Overt marking comes in many types:

Georgian (Kartvelian; Aronson 1991: 261)

- (1) *Bič'-ma c'ign-i da-mal-a.*
boy-ERG book-NOM he-hid-it
'The boy hid the book.'

Yelî-Dnye (Yele; Henderson 1995: 15)

- (2) *M:aa ngê Kaawa dê m.uu.*
Dad ERG Kaawa PUNCT.IND.IMM.PST.3.SBJ see
'Dad saw Kaawa.'

Welsh (Celtic; King 1993: 23)

- (3) *Collodd Siôn ddwybunt.*
lose.PST Siôn two pound
'Siôn lost 2£.' ([d] → [ð]: *dwy* → *ddwy*)

Word order refers here to the dominant arrangement of the core arguments and the main verb (Dryer 1997). At least the following word order types are attested for zero-marking languages.

Thai (Kam-Tai; Thailand; Iwasaki & Ingkaphirom 2005: 110)

- (4) *Lék tè nɔy.*
Lek kick Noy
'Lek kicks Noy.' (dominantly SVO)

Arára Karó (Tupian; Brazil; Gabas 1999: 153)

- (5) *Iyõm wat awe cape-t.*
father 1SG.POSS brother beat-IND
'Father beat my brother.' (dominantly SOV)

Quiegolani Zapotec (Zapotecan; Mexico; Black 2000: 45)

- (6) *W-ey Benit mël.*
COMPL-take Benito fish
'Benito took a fish.' (dominantly VSO)

Minangkabau (Sundic; Indonesia; Gil 2008: 123)

- (7) *Kartini cinto Ujang.*
Kartini love Ujang
'Ujang loved Kartini.' or 'Kartini loved Ujang.' (no dominant order)

2. Method and data

2.1 A diachronic approach to universals

Universals best understood as systematic structural pressure on how languages change over time (e.g. Greenberg 1978; Maslova 2000; Moravcsik (to appear); Bickel 2008).

- If there is universal pressure that favors the development and maintenance of zero marking in SVO languages and disfavors it in non-SVO languages, over time families that have a skewing to zero marking as well as SVO word order will have outnumbered those with non-SVO word order.
- If no universal, then the distribution of innovations will not be similarly skewed across families.
- The datapoints are genealogical groups with a certain response value at any taxonomic level.
 - The groups were divided into sub-groups based on possible difference in word order. For instance, SVO in one sub-group of Oceanic, as in Tungak (8), and verb-initial in another, as in Fijian (9).

Tungak (Oceanic; Fast 1990: 21)

- (8) *Ri aina ki la ta-taun ani keve pok.*
PL woman 3PL.SBJ PFV RED-COOK OBJ.INDF PL food
'The women cooked the food.' (SVO)

Fijian (Oceanic; Dixon 1988: 243)

- (9) *E rai-ca a gone a qase.*
3SG see-TR ART child ART old.person
'The old person saw the child.' / 'The child saw the old person.' (verb-initial)

- Such pseudo-groups can be used for testing if subgroups based on different word orders affect the distribution of morphological type (Bickel 2008). Skewing was determined in the following way:
 - Absolute skewing: all the sampled members of the group had the same morphological type.
 - In case of diversity, skewing was determined statistically by a Monte Carlo randomized permutation test.¹ A significant skewing occurred if < 5% of the permuted datasets had a greater deviation than the observed dataset, otherwise the group had a mixed response.

2.2 Statistical modeling

The relationship between morphological type and word order was studied with logistic regression, which estimates the probability of a particular outcome of the dependent variable (here morphological marking) on the basis of the independent variables (here word order and area) (e.g. Agresti 2002; Baayen 2008).

- The effect of the independent variables is described as odds ratios and the expected response of the dependent variable is transformed via natural logarithm. The formula for our model is given in (10):

$$(10) \quad \log(p(\text{zero marking})/(1-p(\text{zero marking}))) = \alpha + \beta_1 \cdot \text{word_order} + \beta_2 \cdot \text{area} + \beta_3 \cdot \text{word_order} \cdot \text{area}$$

The effects scrutinized by maximum likelihood estimation. The likelihood ratio (LR) of a model with the variable of interest was compared to that of a simpler model without the variable of interest (Agresti 2002). The models were penalized by discouraging large coefficient values (Baayen 2008). The p-values deduced by permutation testing (10,000 permutations).² Encoding of the variables:

- Morphological type: skewed to zero marking vs. not skewed to zero marking.
- Word order: SVO vs. non-SVO.
 - In partial regressions: SVO vs. verb-initial / SVO vs. verb-final / SVO vs. no dominant order.
- For area, three different encodings were used:
 - 1) 6-way areal breakdown used by Dryer (1992).
 - 2) 3-way areal breakdown into the Old World, the Pacific, and the New World (Nichols 1992).
 - 3) Contrasting the hotbeds of zero marking with the rest of the world.

2.3 Data and sampling

No particular sampling method used, aiming at maximum genealogical coverage. Sample: 813 languages (from 448 genera); 102 zero-marking languages (from 50 genera; 13% of the languages).

1 A script developed by Bickel (2008) was used for the permutation test (<http://www.uni-leipzig.de/~autotyp/sample3.r>).

2 A script developed by Bickel (2008) was used for the permutation test (<http://www.uni-leipzig.de/~autotyp/rnd.lr.test.r>).

3. Results

3.1 Descriptive results

There were altogether 191 non-singleton genealogical groups with different responses. Single-member families excluded, since intra-family tendencies can be evaluated only if the family contains >1 member.

- More SVO groups were skewed to zero marking (15) than non-SVO groups (1) (Table 1).

	Languages		Genera		Genealogical groups	
	N	%	N	%	N	%
SVO	73	72	39	75	15	94
Verb-final	15	15	9	17	1	6
Verb-initial	11	11	4	7	0	0
No dominant order	3	3	2	4	0	0
Total	102	101	54	100	16	100

Table 1: Word order of zero-marking languages and genera as well as of genealogical groups skewed to zero marking

Besides the hotbed areas, zero marking occurs also in West Europe, New Guinea, Meso-America, and South America, but not at all in Eurasia (excluding W. Europe), Australia, or North America. SVO has the widest areal distribution, but non-SVO orders also occur scattered around the world.

3.2 The results of the statistical tests

Do the variables together explain the outcome of morphological type? Yes: the full models significantly different from the null models ($p < .0001$).

	Interaction		Word order			Area	
	LR	p	LR	p	Odds	LR	p
6 areas	4.5	.49	19.4	< .0001	14.8	14.8	.0051
3 areas	0.6	.83	20.0	< .0001	17.9	12.7	.0024
Hotbeds	2.9	.31	19.5	< .0001	29.8	27.8	< .0001

Table 2: Results of the regression modeling.

No effect existed for the interaction term in any areal breakdown ($p > .30$), but significant main effects existed for both word order ($p < .0001$) and area ($p < .006$) in all areal breakdowns (Table 2).

- The odds ratio for word order (SVO vs. non-SVO) was 14.8-29.8 → SVO groups about 15-30 times more likely to develop and maintain a skewing to zero marking than non-SVO groups.
- As for partial regressions, there was no effect for the interaction in any areal breakdown ($p > .20$), but significant main effects existed for word order ($p < .006$) and area ($p < .008$) in all areal breakdowns.
- All in all, word order has an effect that is independent of geographical areas.

4. Discussion:

If we can statistically show that diachronic change is affected by structural pressure independently of other factors, it is justified to generalize beyond observable data and logically infer that the same pressure has affected language change in the past and will affect it in the future (Bickel 2008).

- The data suggests that universal structural pressure has affected the way morphological type has changed in languages. This pressure is related to the dominant word order of languages and presumably to how motivated a particular word order is in relation to zero marking.

A more static approach was also tested: counting distinct values in genera as datapoints (Dryer 1992).

Family modeled with values Niger-Congo, Austro-Asiatic and the rest.

- A significant effect for the interaction between word order and area in the partial regression SVO vs. verb-initial order ($p < .05$) (6-way areal breakdown and the hotbed areal breakdown).
- No other effects for any other interactions. What does this mean?
 - 1) Non-SVO languages may **develop** zero marking but not **maintain** it for long, so that the whole group would develop a skewing to zero marking. Only one group that was skewed to zero marking

had a non-SVO word order, the Gur language Supyire (Childs 2003: 201-202).

- Compare to an SVO language: Old Chinese was zero marking 3000 years ago and modern Mandarin still is (Pulleyblank 1995).
- 2) Zero marking is motivated in verb-initial languages to some extent: the verb tells more about argument-predicate relations than the arguments do (e.g. about selectional restrictions, transitivity etc.); that information eases language processing (Hawkins 2004).

5. Conclusions

The results suggest that there is universal correlation between zero marking and SVO word order.

Although zero marking clusters areally, the correlation was independent of it.

- Although zero marking motivated in verb-initial languages, this only shows up in static snapshots of the phenomenon, but not when approaching it more dynamically.

Abbreviations

1 first person, 3 third person, ABS absolutive, ART article, COMPL completive, DECL declarative, DEF definite, ERG ergative, IMM.PST immediate past tense, IND indicative, INDF indefinite, NARR narrative auxiliary, NOM nominative, OBJ object, PFV perfective, PL plural, POSS possessive, PST past tense, PUNCT punctual, RED reduplication, SBJ subject, SG singular, TR transitive.

References

- Agresti, A. 2002. *Categorical data analysis* (2nd edn). New York: Wiley-Interscience.
- Aronson, H. 1991. Modern Georgian. *The indigenous languages of the Caucasus (vol. 1): The Kartvelian languages*, A. Harris (ed.), 219-312. Delmar: Caravan Books.
- Baayen, R.H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: CUP.
- Bickel, B. 2008. A general method for the statistical evaluation of typological distributions. Manuscript, <http://www.uni-leipzig.de/~bickel/>.
- Bickerton, D. 1981. *Roots of language*. Ann Arbor: Karoma.
- Black, C. 2000. *Quiégolani Zapotec syntax: A principles and parameters account*. Dallas: SIL and University of Texas at Arlington.
- Brown, M. 2004. *Topics in Iquito syntax: Word order, possession, and nominal discontinuity*. MA thesis, University of Texas at Austin.
- Childs, T.G. 2003. *An introduction to African languages*. Amsterdam: John Benjamins.
- Comrie, B. 2005. Alignment of case marking. Haspelmath et al. (eds.), 398-405.
- Dixon, R.M.W. 1988. *A grammar of Boumaa Fijian*. Chicago: University of Chicago Press.
- Dryer, M. 1992. The Greenbergian word order correlations. *Language* 68: 81-138.
- Dryer, M. 1997. On the six-way word order typology. *Studies in Language* 21: 69-103.
- Fast, L. 1990. Tungak grammar essentials. *Two grammatical studies*, J. Roberts (ed.), 1-72. Ukarumpa: SIL.
- Gabas, N. 1999. *A grammar of Karo, Tupi (Brazil)*. PhD dissertation, University of California at Santa Barbara.
- Gil, D. 2008. How complex are isolating languages. *Language complexity: Typology, contact, change*, M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), 109-31. Amsterdam: John Benjamins.
- Greenberg, J. 1978. Diachrony, synchrony and language universals. *Universals of human language I: Method and theory*, J. Greenberg, C. Ferguson & E. Moravcsik (eds.), 61-92. Stanford: Stanford University Press.
- Haspelmath, M., M. Dryer, D. Gil & B. Comrie (eds.). 2005. *The world atlas of language structures*. Oxford: OUP.
- Hawkins, J.A. 2004. *Efficiency and complexity in grammars*. Oxford: OUP.
- Henderson, J. 1995. *Phonology and grammar of Yele, Papua New Guinea*. Canberra: ANU.
- Iwasaki, S. and P. Ingkaphirom 2005. *A reference grammar of Thai*. Cambridge: CUP.
- Jackendoff, R. 1999. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3: 272-9.
- King, G. 1993. *Modern Welsh: A comprehensive grammar*. London: Routledge.
- Lehmann, W. 1978. English: A characteristic SVO language. *Syntactic typology: Studies in the phenomenology of language*, W. Lehmann (ed.), 169-222. Austin: University of Texas Press.
- Mallinson, G. & B. Blake. 1981. *Language typology: Cross-linguistic studies in syntax*. Amsterdam: North-Holland Publishing Company.
- Maslova, E. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4: 307-33.
- Moravcsik, E. (to appear). Explaining language universals. *Oxford Handbook of Typology*, Jae Jung Song (ed.). Oxford: OUP.
- Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.
- Nichols, J. & B. Bickel. 2005. Locus of marking in the clause; Locus of marking in possessive noun phrases; Locus of marking: Whole-language typology. Haspelmath et al. (eds.), 98-109.
- Pulleyblank, E. 1995. *Outline of Classical Chinese grammar*. Vancouver: University of British Columbia Press.
- Siewierska, A. & D. Bakker. 1996. The distribution of subject and object agreement and word order type. *Studies in Language* 20: 115-62.
- Sinnemäki, K. 2009. Word order in zero-marking languages. Manuscript, <http://www.ling.helsinki.fi/~ksinnema>.