

Predictions across clauses: Coding asymmetries in complex sentences from a typological perspective

An important claim of the so-called usage-based approach to language structure is that interlocutors constantly draw on their ‘mental corpus’ (Taylor 2012) in planning and comprehending utterances, and that we exploit this statistical knowledge in order to render communication more efficient: Predictable information tends to be formally reduced when directly compared to less predictable counterparts. Psycholinguistic studies have provided ample evidence of such predictability-based reduction processes (e.g. Gahl & Garnsey 2004; Wasow et al. 2011; Kurumada & Jaeger 2015, among many others), and predictability has also been invoked to explain cross-linguistically recurrent asymmetries of grammatical coding, such as differential zero coding of case, number, person or voice (e.g. Givón 1979; Comrie 1981; Croft 2003; Hawkins 2004, 2014; Haspelmath 2008, 2017).

The present paper continues along these typological lines and investigates an area for which such coding asymmetries are yet to be established systematically, viz. that of clause combinations of various types. I specifically focus on reduction effects at clause boundaries, such as the differential use of relativizers and complementizers (cf. (1)), and on the length of directly contrasting markers of referential continuity (same-subject versus different-subject marking, cf. (2)):

- (1) Barbareño (Chumashan; Wash 2001: 89)
S-ʔip-waš (hi=)ʔal-saʔ-aktiña.
3-say-PST COMP=STAT-FUT-come
‘She said (that) she was going to come.’
- (2) Jamul Tiipay (Cochimi-Yuman; Miller 2001: 240–256)
- a. *Juan may we-yiw xemaaw [xenu-ch kenaach].*
Juan NEG 3-come not be.sick-SS because
‘Juan didn’t come because he is sick.’
- b. *[[Servees me-si-x-pu] m-aa-chm] uuyaaw.*
beer 2-drink-IRR-DEM 2-go-DS know
‘I know you went there to drink beer.’

Based on a growing typological database of languages from diverse families and areas, I show that the grammatical data square well with the findings from the psycholinguistic and corpus-linguistic literature: Optional marking (and grammaticalized zero marking) of clause boundaries occurs in proportion to the degree to which the occurrence of the subordinate clause in question is predicted by (elements of) the matrix clause (as in (1)); and where two overt markers contrast in the expression of a cross-clausal relation, the less common scenario shows at least as much coding material as the more common one, but not less (as in (2)). These two coding principles are fleshed out and tested in more specific terms, and I also briefly outline which other factors (apart from predictability) may yield similar effects.

References

- Comrie, Bernard (1981). *Language universals and linguistic typology*. Chicago: The University of Chicago Press.
- Croft, William (2003). *Typology and universals*. 2nd ed. Cambridge: Cambridge University Press.
- Gahl Susanne & Susan M. Garnsey (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80(4). 748–775.
- Givón, Talmy (1979). *On understanding grammar*. New York: Academic Press.
- Haspelmath, Martin (2008). Creating economical morphosyntactic patterns in language change. In Jeff Good (ed.), *Language universals and language change*, 185–214. Oxford: Oxford University Press.
- Haspelmath, Martin (2017). Explaining alienability contrasts in adpossession constructions: Predictability vs. iconicity. *Zeitschrift für Sprachwissenschaft* 36(2). 193–231.
- Hawkins, John A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hawkins, John A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.
- Kurumada, Chigusa & T. Florian Jaeger (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language* 83. 152–178.
- Miller, Amy W. (2001). *A grammar of Jamul Tiipay*. Berlin, New York: Mouton de Gruyter.
- Taylor, John R. (2012). *The mental corpus: How language is represented in the mind*. Oxford: Oxford University Press.
- Wash, Suzanne (2001). *Adverbial clauses in Barbareño Chumash narrative discourse*. PhD dissertation, University of California at Santa Barbara.
- Wasow, Thomas, T. Florian Jaeger & David M. Orr (2011). Lexical variation in relativizer frequency. In Horst Simon & Heike Wiese (eds.), *Expecting the unexpected: Exceptions in grammar*, 175–195. Berlin: Mouton de Gruyter.