



Verb-based and noun-based languages

Johanna Nichols *

University of California, Berkeley

INTRODUCTION

Problems arise with the status of noun and verb in languages: No noun-verb distinction? rigid vs. flexible? unidirectional flexibility (contextual neutralization of noun/verb)? lexical class defined at different levels of grammar in different languages? flexibility, conversion, precategoriality? (See References.)

What typology needs:

- Turn these theoretical contributions into a workable typological variable.
- More multivariate variables.
- More fine-grained typological distinctions.
- Seek individual patterns, not whole-language types.

This project uses a wordlist approach to create two multivariate, easily replicable typological variables, and reports preliminary findings of interest to typology, historical linguistics, and language prehistory.

METHOD

Survey: 50-item wordlist covering various lexicosemantic subgroups of event-like, referent-like, and property-like lexemes (drawn from an earlier pilot 200-item list [Foley & Nichols] by choosing the cross-linguistically most variable lexemes). Using medium to large-sized bidirectional bilingual dictionaries (and consultation with grammars and/or expert linguists), find the closest fit to each wordlist item and determine its part of speech following the standard criteria for that language. 53 languages so far, covering northern Eurasia well.

Analysis: For each item determine the **derivational base** – the head of the derivational paradigm or word family – by seeking morphologically related words in the dictionary and looking up a standard set of likely related glosses: e.g. for *warm*, modifier 'warm', predicates 'be warm', 'get warm', causative 'heat up, make warm', noun 'warmth'. The base is whichever word(s) is/are underived or least derived. It may – or may not – be the etymological root.

Typologizing: For each language, count the number of wordlist items whose bases are verbs and the number with noun bases. Express that as a percent of the total items found. E.g. .:

Creek (Muskogean)	19% noun-based	76% verb-based	(verb-based language)
Ingush (Nakh-Daghestanian)	65% noun-based	17% verb-based	(noun-based language)

Language types: noun-based, verb-based, depending on which predominates.

Highly noun-based, highly verb-based: More than one standard deviation from the world mean (currently 46% and 30% for a small sample).

This typology is based on 50 separate word families of different types and is expressed as a percentage, so it is both multivariate and fine-grained.

EXAMPLES OF LANGUAGE TYPES

Highly noun-based languages: *Tajik* (Iranian) (67% noun-based, 15% verb-based). Simplex verbs are a closed class. New verbs are formed chiefly as denominals or light verb constructions where the lexical (heavy) piece is usually a noun or adjective (so these are noun-based or adjective-based):

<i>otaš giriftan</i>	(fire take)	'catch fire, flare up'
<i>otaš zadan</i>	(fire strike)	'burn, set afire'
<i>otaš dodan</i>	(fire give)	'light, set afire'
<i>qahr kardan</i>	(anger do)	'get angry'
<i>bad-qahs kardan</i>	(into-anger do)	'make angry, anger'
<i>sulfa kardan</i>	(cough do)	'cough'
<i>sulf-idan</i>	(denominal verb)	'cough'
<i>guš kardan</i>	(ear do)	'listen'
<i>parvoz kardan</i>	(flight do)	'fly' (incl. in airplane)
<i>par zadan</i>	(feather strike)	'fly' (of birds)
<i>par-idan</i>	(denominal verb < par 'feather')	'fly'

Light verb constructions are the standard way to form verbs from loans: *remont kardan* (repair do) 'repair' (Russian noun), *va?da dodan* (promise give) 'promise' (Arabic noun).

Very similar: *Nakh-Daghestanian* languages.

Highly verb-based languages: *Thompson River Salish*. (2%, ~98%). The great majority of words used referentially are deverbal (nominalized) or ('moon', 'eagle') are forms of verbs. These include words that are cross-linguistically basic nouns: body parts, kin terms, objects in nature, natural kinds:

'tongue'	<i>təf-e?</i>	'stick out (round flexible object)'
'wing'	<i>s-c'k=áXn</i>	s- nominalizer
'hair'	<i>s-/q'ép=qn</i>	s- nominalizer
'mother'	<i>s-/kix-ze?</i>	s- nominalizer
'father'	<i>s-/qac-ze?</i>	s- nominalizer
'moon'	<i>máƒ=xe-tn</i>	light:up=foot-INSTR
'sun'	<i>s-k'wák'wes</i>	s- nominalizer
'dog'	<i>s-/qáXa?</i>	s- nominalizer
'eagle'	<i>?es-/kwl-o?=?in</i>	STAT-yellow=head ('head' is a lexical suffix)
'tree'	<i>s-/ghep</i>	s- nominalizer 'erect (a pole)'

Fine print: Thompson is arguably flexible, not verb-based.

Creek (Muskogean) has more basic nouns than Salish languages do, but deverbal nouns predominate.

'blood'	<i>c:at-a</i>	bleed-NZ
'moon'	<i>hali:ss-í</i>	(participle)
'screw'	<i>(i)š-im-aféy-ka</i>	< aféyy- 'twist'
'rifle, gun'	<i>icca-, ic-ha</i>	< icc- 'shoot'

TYPOLOGICAL CORRELATIONS

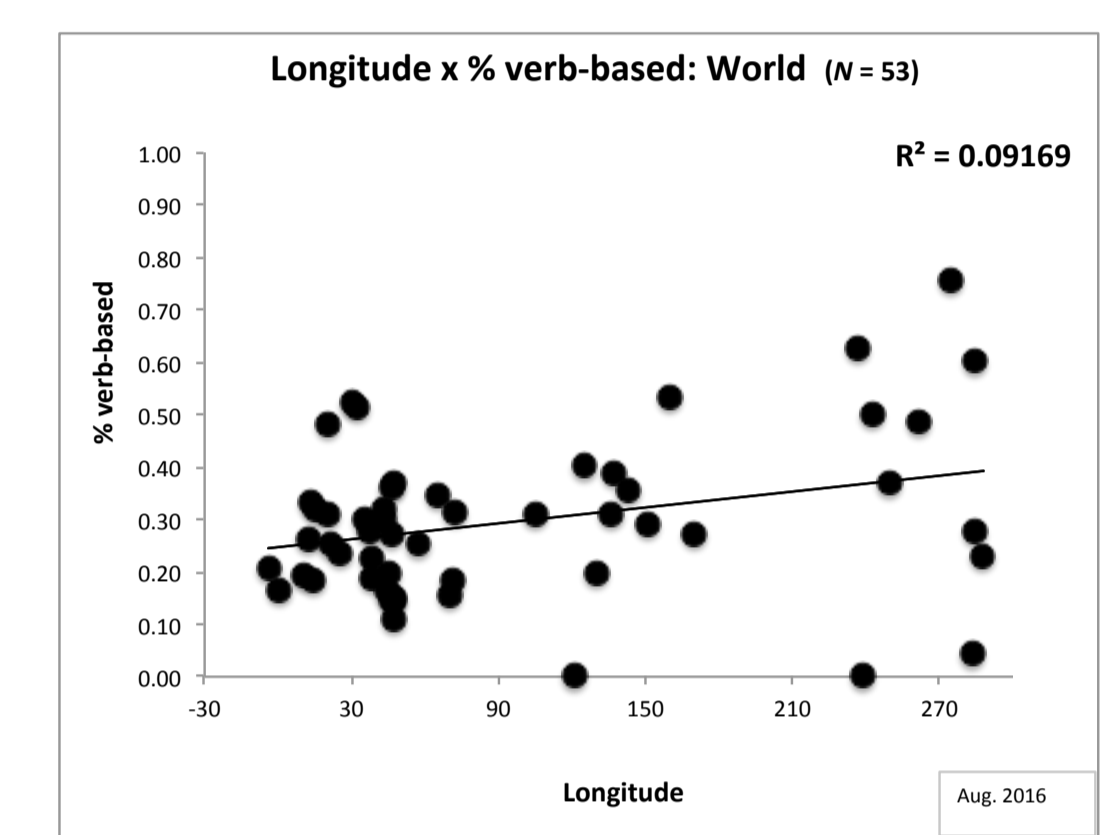
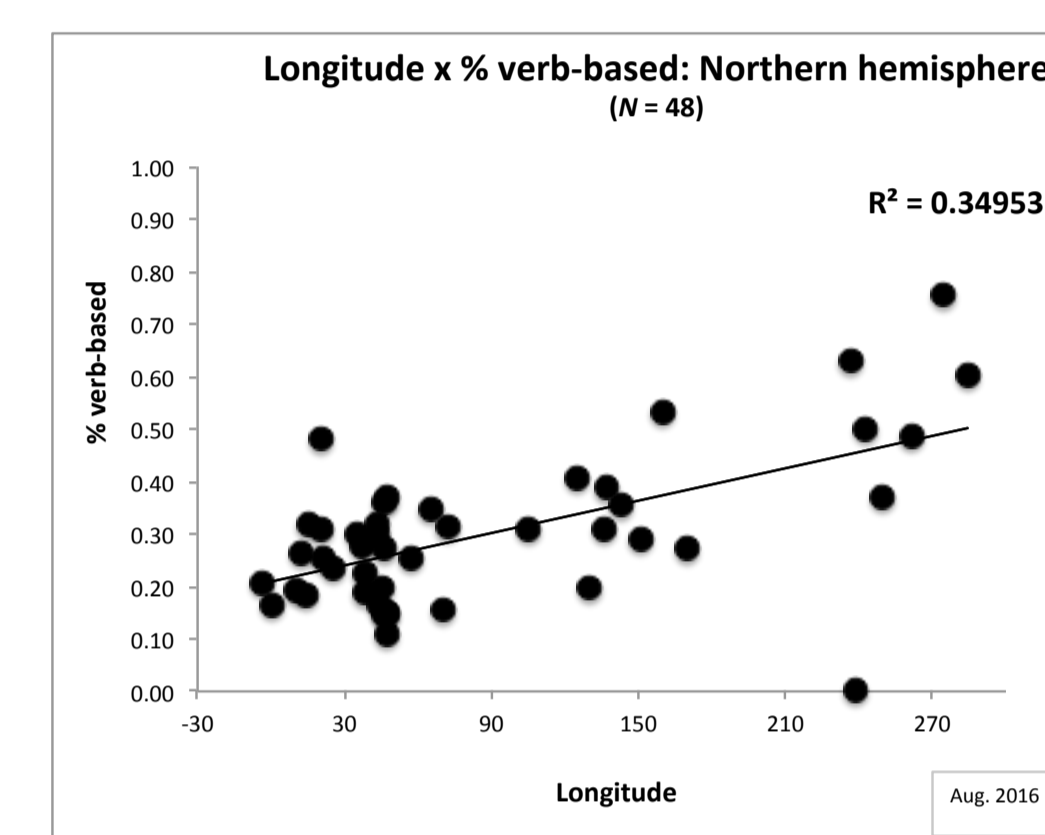
No significant correlations found with any of the commonly described typological variables: word order, alignment, head/dependent marking, basic valence orientation (transitivizing/detransitivizing), inclusive/exclusive, hierarchical patterning. Good news: **Independent typological variable.**

HOW MANY VARIABLES?

Percent noun-based and percent verb-based are independent of each other. Good news: **Two separate variables.** Percent flexible is not independent from verb-based or (especially) noun-based.

LINGUISTIC GEOGRAPHY

Percent verb-based forms a strong east-west cline across the northern hemisphere: low in the west, high in the east. It is strong enough (left below) to influence worldwide distributions (right below). No cline in the south.



It decreases to the south in the Americas and increases from Europe to Africa. Therefore, the bottom and peak are in the high northern latitudes. (Same distribution as causativizing valence type.)

No large-scale geography for percent noun-based.

HISTORY

Language families replicate internally the east-west cline: Uralic, IE, Iranian branch of IE. This reflects entries into the European cultural sphere. The evolution is visible over time in the Slavic written record.

Cross-type lexemes (noun-based verbs, verb-based nouns) appear to be less stable than same-type lexemes. Example: In Nakh-Daghestanian, a number of simplex verbs reconstruct as PND, but very few LVC's reconstruct.

Type is fairly stable in families.

PIE was evidently verb-based, so the best candidate for an ancient sister (from among the usual suspects) is Kartvelian (probably verb-based). Next: Uralic (mixed, flexible). Then: Yukagir. Very weak: Turkic, Mongolic (noun-based).

CONCLUSIONS

- Noun-based and verb-based are good independent typological variables. Replicable, multivariate, fine-grained.
- Verb-based has an important geography. It is very similar to the geography of the causativizing (transitivizing) type, though there is no grammatical correlation. Implication: these two variables independently mark the same long-standing apex or spread center in the higher northern latitudes, apparently around Beringia.
- Derivational base opens new arenas for typology and historical linguistics, and brings morphosyntactic typology and etymology together.
- Lexical stability is relative to derivational type. This should make possible some fine-tuning of lexicostatistical measures.

The Swadesh lists are predominantly nouns. Nouns are less stable in verb-based languages. Does this mean that the (numerous) verb-based language families of the Americas are older than they appear (judged lexically)??

- Next steps: Expand the wordlist somewhat, expand the sample considerably (ongoing). Survey etymologies systematically. The survey is extremely labor-intensive and requires good lexical resources and a good description of word formation. But there is a high potential payoff for typology and historical linguistics.

REFERENCES

Baker, Mark C. 2003. *Lexical Categories: Verbs, Nouns, and Adjectives*. CUP.

Beck, David. 2013. Unidirectional flexibility and the noun-verb distinction in Lushootseed. In J. Rijkhoff and E. van Lier, eds., *Flexible Word Classes*, 185-220. OUP.

Bisang, Walter. 2008. Precategoriality and syntax-based parts of speech: The case of Late Archaic Chinese. In Umberto Ansaldi, Jan Don and Roland Pfau, eds., *Parts of Speech: Descriptive Tools, Theoretical Constructs*, 568-589. Amsterdam: Benjamins.

Bright, W.O. 2004. "Analyzability" of nouns in northwestern California. *Conference on Language Ecology*, UC Berkeley.

Evans, Nicholas and Toshiaki Osada. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology* 9.351-390.

Foley, William A. Forthcoming. *The Epidemiology of Language*.

Hengeveld, Kees, Jan Rijkhoff and Anna Siewierska. 2004. Parts-of-speech systems and word order. *Journal of Linguistics* 40.527-570.

Janhunen, Juha. 2001. Indo-Uralic and Ural-Altaic: On the diachronic implications of areal typology. C. Carpelan, A. Parpola, and P. Koskikallio, eds., *Early Contacts between Uralic and Indo-European: Linguistic and archaeological considerations*, 207-220.

Lehmann, Christian. 2010. Roots, stems, and word classes. In U. Ansaldi, J. Don and R. Pfau, eds., *Parts of Speech: Empirical and Theoretical Advances*, 43-64. Amsterdam: Benjamins.

Lehmann, Christian. 2013. The nature of parts of speech. *STUF* 66:2.53-92.

Martin, Jack B. and Margaret McKane Mauldin. 2000. *A Dictionary of Creek/Muskogee*. Lincoln: U. Nebraska Press.

Nichols, Johanna. 2010. Slavic indeterminate verbs are denominal. R. Perelmutter and V. Driagina-Hasko, eds., *New Approaches to Slavic Verbs of Motion*, 47-65. Amsterdam: Benjamins.

Nichols, Johanna, David A. Peterson, Jonathan Barnes. 2004. Transitivity and detransitivizing languages. *Linguistic Typology* 8:2.149-211.

Nichols, Johanna. 2010. Lexical semantic constraints on noun roots and noun borrowability. Ansaldi et al. eds. 183-200. Amsterdam: Benjamins.

Perry, John R. 2005. *A Tajik Persian Reference Grammar*. Leiden: Brill.

Rahimi, M. V. and L. V. Uspenskaja. 1954. *Tadžiksko-russkij slovar'*. Moscow: Gos. Izdat. Inostrannyx i Nacional'nyx Slovar'ej.

Thompson, Laurence C. and M. Terry Thompson. 1996. *Thompson River Salish Dictionary*. Missoula: U. Montana.

ACKNOWLEDGMENTS

* Work done partly with William A. Foley (U. Sydney)
 Wissenschaftskolleg zu Berlin
 NSF grants 9222294, 0966675
 Max Planck Institute for Evolutionary Anthropology, Leipzig

CONTACT:
 johanna@berkeley.edu