

Headedness, again

Maria Polinsky

Introduction¹

Headedness is an intriguing feature of language design. On the one hand, headedness manifests itself very clearly; preposed relative clauses are visibly different from postposed ones, and postpositions are easily distinguished from prepositions. More generally, structural heads (the constituents which determine the category of their phrase) either precede or follow their dependents. On the other hand, there is room for disagreement and variation. For instance, the ordering of determiners and nouns can be assessed differently depending on what counts as the head, causing disagreements over the headedness of nominal constituents. Furthermore, even if all linguists agree on what counts as the head and what counts as a dependent, there is no requisite consistency in the way dependents and heads are ordered across different phrases within the same language. An otherwise dependably head-final or head-initial language may exhibit exceptions; the results are messy, and linguists get discouraged.

There is, as of yet, no good explanation for headedness. It is visible; it is rather robust; it seems easy to learn (Lupyan and Christiansen 2002; van Everbroeck 2006), but what is it? This question has puzzled many researchers, and we still do not know its answer. When dealing with something that is unfamiliar it is often tempting to just toss it out as unnecessary or superficial. Researchers now and again have suggested that headedness is no more than a trivial pattern-recognition device without much deep meaning and with no value in linguistic theory. Yet it is hard to dismiss a device that is so pervasive. This squib presents a new argument as to why theoreticians should still give headedness a chance. I am not prepared to explain headedness, but I will bring in a new dimension in which its effects are apparent: the proportions of lexical categories. If my results are on the right track they add further evidence in support of the conception that headedness is still important.

1 Starting point

The new dimension of language structure where headedness manifests its effects has to do with the number of verbs *vs.* number of nouns in a language, that is, the size of verb *vs.* noun class. To begin with a casual observation, many L2 learners of languages such as Japanese know from experience that being unfamiliar with a verb may not be the end of

¹ For helpful comments on this paper, I am grateful to Katryna Cadle, Keith Plaster, Kevin Ryan, and Barbara Stiebels. All errors are my responsibility.

the world. When at a loss, the learner takes a noun, combines it with the verb *suru* ‘do’ and can be understood, even though the result may not be authentic Japanese. These days *suru* mostly combines with English words, as in *beesu appu suru* ‘increase salary (from base up)’, *emayru suru* ‘email’, *guuguru suru* ‘Google’, and many others. There is the usual hand wringing about the Japanese vocabulary being destroyed by English, but the modern-day mourners of Japanese forget that before it was being destroyed by English, *suru* used to combine with Chinese words, yielding such compounds as *kenkyuu suru* ‘study’ or *gensyoo suru* ‘decrease.’ This initial observation suggests that Japanese has a rather small number of inflecting verbs and a large open class of complex predicates. Such complex predicates are created from non-verbal constituents combined with light verbs of which *suru* is the most common one.

Turning to less-known languages, Pawley (2006) discusses the northern Australian language Djamindjung (djd) and the PNG language Kalam (kmh), which both have a real paucity of verb roots—just over a hundred. These small, closed classes of verb roots occur as independent verbs, and all other verb meanings are expressed by complex predicates, as in Japanese. Pawley suggests that these languages are not unique, and that related Australian and PNG languages also have small, closed verb classes.

What do other languages do? English’s response to the need for new verbs is to make a verb out of pretty much anything using zero morphology (conversion), yielding *to ftp*, *to R the data*, *to KCCO a friend*, or, from the days of the Clinton White House, *to Linda-Tripp someone*. Languages encumbered by more morphology than English build new inflected verbs using verbal affixation; for instance, modern Russian, which has experienced a true Anschluss of English words has been creating verbs like *piarit* ‘to PR’, *parkovat* ‘park’, *postit* ‘post on a blog’, or *kopipejstít* ‘copy and paste’ in droves.

So the difference between English and Russian, on the one hand, and Japanese, Djamindjung, and Kalam, on the other, is that while English and Russian freely create new verb roots or stems to add new verbal concepts to the language, the latter three do not; instead, they rely on light verbs to produce new complex verbs. The three languages that utilize light verbs happen to be head-final and SOV. Is this an accident, or does that paucity of inflecting verbs have anything to do with headedness? This is the essence of the question that I will explore in this paper:

- (1) Does the noun-verb ratio differ across headedness types?

In order to investigate (1) I first need to go over the main headedness types, and also clarify, even if only partially, what counts as a noun or a verb. The next two sections will address these issues.

2 Headedness types

As far as headedness goes, the main contrast is between head-final and head-initial languages. Within the head-final type, languages such as Japanese and Korean represent the “rigid head-final” type (cf. Kayne 1994; Siewierska 1997; Herring and Paolillo 1995 and references therein). In a way, they are dream languages because their heads consistently follow dependents in all types of phrases. Languages such as German or Persian can be considered exemplars of the non-rigid head-final type; their head-final property seems to be a violable constraint in an optimality design.

Rigidly head-final languages do not allow verb-medial or verb-initial orders, but at the other end of the headedness scale, head-initial languages (VSO, VOS) always seem to allow verb-medial orders. In fact, verb-initial languages that do not allow verb-medial SVO are either impossible or rare (Siewierska 1997).

Once we allow optionality, it can become confusing as to how to classify a given language. For instance, is Yucatec Mayan VOS or SVO? Its most frequent word order is SVO; all its genetic relatives are verb-initial, and it still uses a number of verb-initial orders. Understandably, researchers cannot agree; Briceño (2002) and Gutierrez-Bravo and Montforte (2008, 2009) classify it as SVO; Hofling (1984) and Durbin and Ojeda (1979) argue that it has two basic word orders, SVO and VOS, but with a secondary statistical preference for SVO, and finally, Gutierrez-Bravo and Montforte (2010) suggest that it is SVO with two-place predicates and VS in objectless clauses. This confirms that headedness is frequently inconsistent.

In establishing the subtypes for my query, I would like to balance the need to recognize different headedness subtypes and the desire to have as few types as possible. So the types I will be using are as follows:

(2) Basic headedness types and their examples

Rigid head-final	Non-rigid head-final	Clearly head-initial	SVO/head-initial	SVO, sundry
Japanese, Korean, Tamil	German, Persian, Latin, Tsez, Avar, Basque	Malagasy, Tongan, most Mayan languages, Irish	Indonesian, Yucatec Mayan	English, Russian, Romance languages, Bantu languages

With this very broad-based typology, I would like to examine the ratio of nouns vs. verbs in languages illustrating each type. This investigation is naturally limited by the available data; languages such as English and some other Indo-European languages are catalogued in WordNet (Miller et al. 1990) or CELEX Lexical Databases.² For other languages, the data are much more limited and surprisingly hard to come by (see also below).

In order to get a set of comparable data, I have limited my query to the ratio of nouns to verbs. This is a reasonable measure; if we added the two other lexical categories that are often included in the counts, adverbs and adjectives, we would start losing the strength of cross-linguistic comparison. While noun-verb distinctions may sometimes be subtler than we usually assume (an issue to which I will return in the next section), all languages have nouns and verbs. However, not all languages have easily *identifiable* adjectives and adverbs, another reason to exclude them.

Before discussing the absolute numbers, however, let me address the issue of noun-verb distinctions.

² For details on CELEX, see links and references at:
http://www ldc.upenn.edu/Catalog/readme_files/celex.readme.html#sources.

3 Nouns and verbs: Can we always tell?

Most linguists have historically agreed that all languages have some universal structural building blocks, among which are the lexical categories of nouns and verbs. However, “a persistent thread of research that maintains that there are languages that do not have ... familiar ... categories” (Chung 2012) has created serious doubts about this universality. While the division of the lexicon into nouns and verbs is likely universal, the diagnostics for lexical classes are language-specific, and may even be highly obscure or subtle. In general, the identification of nouns vs. verbs relies on formal patterns of inflection, morphological derivation, and syntactic distribution (Schachter 1985, Sasse 1993, Baker 2003, Kaufman 2009, Chung 2012).

The languages for which a stringent lexical division between nouns and verbs has been most doubted are characterized by a large class of roots that can be used either nominally or verbally, as in Tongan (Broschart 1997), Chinese (Chao 1968), Riau Indonesian (Gil 2005) or Mapuzungun (Malvestitti 2006). Such languages often have polysynthetic features (see Lois and Vapnarsky 2006 for Amerindian, Aranovich 2010 for Austronesian, Arkadiev et al. 2009 for Adyghe) or templatic morphology (Arad 2003), and include many multifunctional content words. A careful analysis of the categorially ambiguous content words usually shows fine-grained distinctions and thus leads to the desired differentiation of lexical categories. To cite a few examples, Chung’s (2012) meticulous study argued that the Chamorro language has noun, verb, and adjective categories. In Adyghe, only nouns proper but not derived nouns (e.g. nouns derived from verbs) can appear without overt case marking (Arkadiev et al. 2009: 51-56). Other studies that identify fine-grained distinctions between nominal and verbal roots include Arad’s (2003) semantic analysis of the relations between nominal and verbal roots in Hebrew (showing principled rules underlying root polyvalence), and Haviland’s (1994) analysis of roots in Tzotzil. This is not the place to defend the universality of the noun-verb distinction; much work in that direction will rely on better understanding the lexical semantics of complex word formation and category conversion.

For my purposes, the best I can do is to assume that the lexicon of a given language is divided into nouns and verbs based on language-particular criteria, including inflectional morphology, semantic correspondences (Arad 2003, Chung 2012), and syntactic distribution. In some of the languages cited below, most notably Zinacantec Tzotzil (Haviland 1994), the noun-verb division is established at the level of roots rather than lexical items.

All in all, the seemingly simple question of counting nouns and verbs is a quite difficult one; even obtaining data about the overall number of nouns and verbs proves to be an immense challenge. The ultimate consequence is that linguists lack reasonable tools to compare languages with respect to their lexical category size. Cooperation between theoreticians and lexicographers is of critical importance: just as comparative syntax received a big boost from the micro-comparative work on closely related languages (Romance; Germanic; Semitic), so micro-comparative WordNet building may lead to important breakthroughs that will benefit the field as a whole.

4 Results

Table (3) shows the calculation of the noun-to-verb ratio for some representative

languages.³

(3) Nouns and verbs across languages: Numerical comparison of lemmas⁴

	Nouns	Verbs	Noun-to-verb ratio
Japanese	86028	15346	5.6
Korean	89125	17956	4.96
Tamil	2403	423	5.6
Telugu	3489	521	6.69
Archi* (Kibrik et al. 1977)	2419	362	6.68
Tsez* (Xalilov 1999)	3508	506	6.93
Hungarian	31600	3300	9.57
Basque	23069	3496	6.59
Latin* (Aronoff 1994; Minozzi 2009)	4777	700	6.82
German* (Barbara Stiebels, p.c.)	72785	11201	6.49
Dutch (average of WordNet and CELEX)	59182	8549	6.9
English	82115	13767	5.9
Chinese* (Xu et al. 2008)	78764	13430	5.86
Polish	14131	3497	4.04
Czech	31029	5158	6.02
Greek	29782	7839	3.7
Romanian	56594	16122	3.5
Spanish	48323	12910	3.74
Swahili*	685	226	3.03
Hebrew	11961	4804	2.49
Vietnamese	6000	2500	2.4
Bahasa (Indonesian/Malay)	12429	5805	2.14
Zinacantec Tzotzil* (Haviland 1994)	1629	850	1.91
Halkomelem* (Galloway 2009)	967	916	1.05
Zapotec* (Long and Cruz 1999)	542	439	1.23
Irish (Modern, from 1800)	1850	890	2.07
Bukusu* (Larry Hyman, p.c.)	2879	1653	1.74
Bobangi* (Larry Hyman, p.c.)	3973	3324	1.19
Malagasy* (Diksionera 1973)	5436	3643	1.49
Maori* (Williams 1957)	2920	1656	1.76

The chart below shows the distribution of noun-verb ratios across these languages. The languages can be broken into three bins that show a strong correlation with headedness.

³ For languages marked with an asterisk the data come from dictionaries or published sources indicated in parentheses; all other numbers are from WordNet, CELEX and/or corpora. I am very grateful to Eneko Agirre, Francis Bond, Verena Hinrichs, Katia Kravtchenko, Sun-Hee Lee, Dan Tufis, Shuly Wintner for help with the counts.

⁴ Where it is relevant, the counts exclude compound verbs formed using a light verb as in the Japanese examples above.

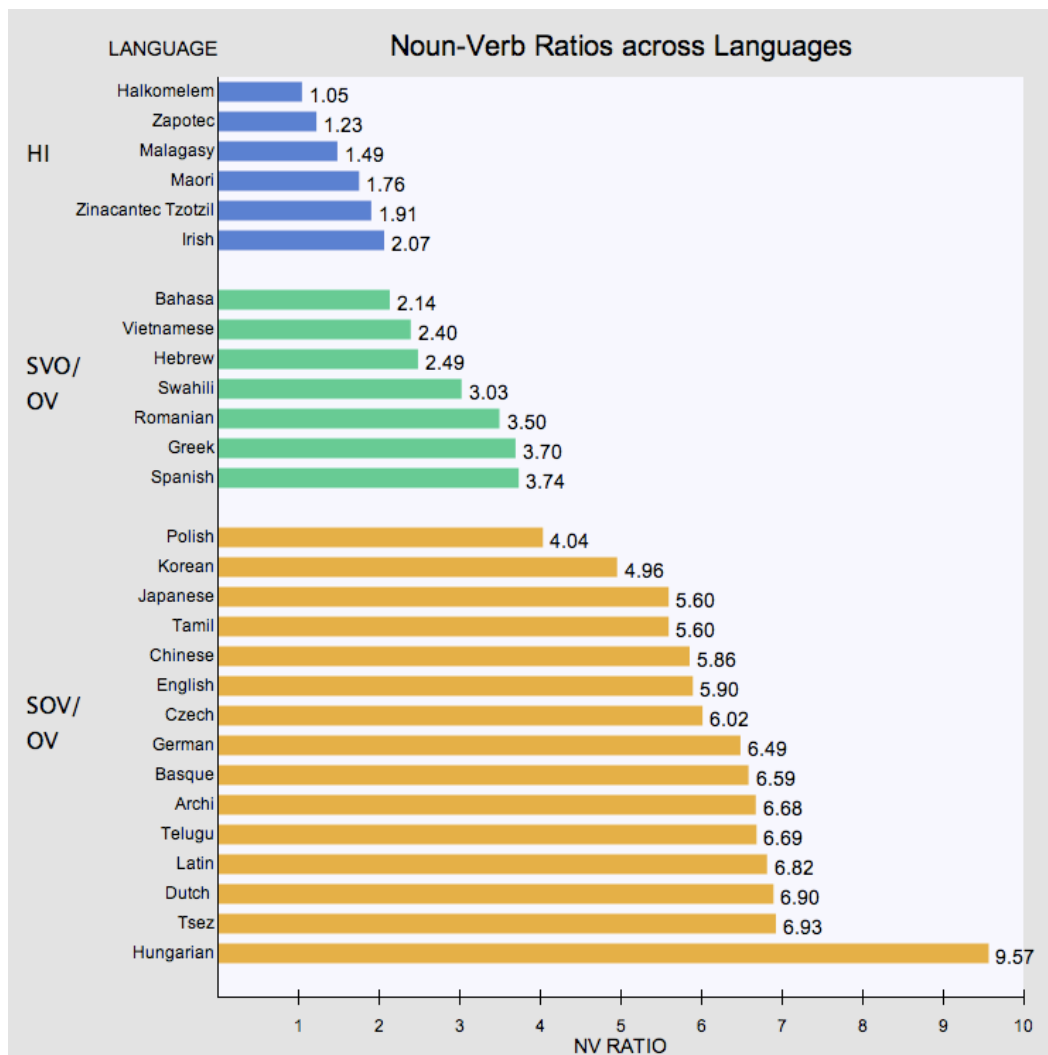


Figure 1. Noun-to-verb ratios in the sample languages

The group with the lowest noun-verb ratio includes Maori, Malagasy, Irish, Zapotec, Halkomelem, Tzotzil, and two Bantu languages. All these languages are head-initial (HI in the chart stands for ‘head-initial’). The two Bantu languages surveyed, Bukusu and Bobangi, differ from the rest of the group in that they are SVO; nevertheless they show particularly strong head-initial properties; their noun-verb ratios are very much like the ones observed in Irish or in the sampled Austronesian languages.

The intermediate group includes Greek, Romanian, Spanish, Swahili, Hebrew, Vietnamese, and Bahasa, languages that are all SVO with head-initial characteristics. The third group, with the highest noun-verb ratio includes the majority of the surveyed languages, which are rigidly verb-final and SVO/OV. Wilcoxon tests show that the three bins identified here and shown in different colors in the chart (head-initial, SVO/OV, and SOV/OV types) are real; the differences between the bins are highly significant ($p \leq 0.001$).

5 What this means

The results show a clear correlation between headedness and the proportion of verbs in the lexicon. Head-initial languages (Irish, Malagasy, Maori, Tzotzil, Bantu) have a particularly high proportion of verbs. In contrast, languages of the rigidly head-final type are verb poor. Hungarian seems almost an outlier, with the highest noun-verb ratio (9.57 as compared to the 5-6 ratio found for other head-final languages), but that could be an artifact of the incomplete corpus of Hungarian WordNet (Miháltz et al. 2008). This introduces an additional concern: we must question the adequacy of language corpora and dictionaries, whether they accurately identify nouns and verbs, and whether they accurately reflect the everyday life of a given language. Even if we accommodate for the variation in the sample we still see a significant clustering of verb-poor languages in the head-final type and of verb-rich languages, in the head-initial type.

The intermediate group includes the SVO languages, which must be probed more thoroughly to see what additional patterns may emerge. For now, I would like to offer two considerations. The first one takes into account the canonic idea that SVO languages are not uniform, comprising of OV and VO languages. Many researchers agree that OV and VO are simply representations of head-final and head-initial structures respectively (see Lehmann 1973, 1978, Venemann 1974, 1976, for the initial idea). Each subtype has significant structural corollaries; for instance, OV in an otherwise SVO language entails object shift, scrambling, final question particles, and head-final embedded structures—none of which is found in a VO subtype of SVO languages (Dryer 1991, Vikner 1994, Biberauer and Roberts 2005, 2009, a.o.). The noun-verb ratios reflect the division of SVO languages into OV and VO types quite well: Greek, Romanian, Spanish, Swahili, Hebrew, Vietnamese, and Indonesian/Malay all have independently documented VO characteristics and their noun-verb ratios are very close to the ones found in the bona fide head-initial languages such as Irish or Zapotec. At the other extreme, Chinese, a source of never-ending sorrow for advocates of well-behaved SVO languages, shows OV properties; its ratio is very close to the one observed in head-final languages in our sample. Indeed, Chinese has prenominal relatives, which is very unusual for SVO languages, as well as object shift and scrambling; as a result, researchers are often at a loss as to how to characterize it (see Dryer 1991: 447, 476 for different, often conflicting approaches). To take another example, Latin conforms to the OV stereotype with a high noun-verb ratio, even though its Romance offspring show VO properties. All these results add a novel argument to the general notion that SVO is no more than a shibboleth, and that the real distinction is between OV and VO language types.

A few languages do not fit into their expected slots, namely the Germanic and Slavic languages from my sample. Let us start with the three Germanic languages: German, Dutch, and English. According to the data in the table, German and Dutch seem more OV than even rigid OV languages. From all we know about its structure, English patterns with VO languages, but its ratio is like that of Chinese. As with the languages discussed in the preceding paragraph, such a pattern may be a side effect of the way English WordNet was built. For instance, if we count particle verbs as separate verbs that would inflate the verbal lexicon; counting obsolete or occasional nouns would inflate the nominal part of the English WordNet. For comparison, let's set the WordNet numbers aside and consider the ratio of nouns to verbs in child directed speech in CHILDES. In the corpus of parental speech addressed to Sarah we find 1403 nouns and 390 verbs, with

the resulting ratio of 3.5, which is much closer to the ratios in other VO languages.⁵ Assuming that the parental speech is a better representative of the actual language than the semantic web at WordNet, this is a welcome result.

Turning now to the two Slavic languages, Czech and Polish, one would expect them to be more similar to each other. The differences may be due to the Slavic-specific issues that arise in the construction of dictionaries, WordNet, or other databases. In their discussion of the Czech Wordnet, Pala et al. (2008: 371) explicitly address the outstanding issues that Slavic lexicographers need to address: verb aspect; reflexive verbs; verb prefixation (single, double, triple); diminutives (noun derivation by suffixation), and noun derivation by suffixation. For instance the number of verbs could go up or down depending on how the lexicographer approaches Slavic aspectual pairs: does one count verbs in the perfective and imperfective as separate lemmas or as members of the same lemma? Counting all verbs twice of course inflates the size of the verbal lexicon. Similarly, counting nouns in the diminutive as separate lemmas or as part of the same lemma as the corresponding non-diminutive would affect the size of the nominal lexicon. These two factors alone are more than sufficient to force an even greater discrepancy than the one we observe.

Conclusion

Initially, I asked whether the noun-verb ratio differs across headedness types. I collected simple numerical data on the noun-verb ratio across a sample of languages, chosen more or less opportunistically as a “convenience” sample, focusing on languages for which I was able to find or recover numerical data on the number of nouns and verbs.

The results may be surprising: there is a robust correlation between headedness and the proportion of verbs in the lexicon. Head-final, OV languages have a relatively small percentage of *simple* verbs, whereas head-initial languages have a considerably larger percentage of simple verbs. The OV/VO difference with respect to noun-verb ratios also reveals itself in SVO languages; some languages, Chinese and Latin among them, show a strongly OV ratio, whereas others, such as Romance or Swahili, are VO-like in their noun-verb ratios.

Another way of looking at these results is to tie them to the possible and/or preferred derivational methods used by a given language. In that case, the correlation is between headedness and choice of derivational method. Looking back at the examples used in this paper, English happily zero-derives verbs, Russian adopts new verbal roots with or without a derivational suffix, but head-final languages prefers to use light verbs.

On either approach, the proportion among lexical classes emerges as a new linguistic characteristic that is correlated with headedness. Further verification is needed, and assuming that further studies confirm this new generalization, the next step is to explain why this pattern exists.

References

Arad, M. 2003. Locality constraints on the interpretation of roots: The case of Hebrew

⁵ I am grateful to Robyn Orfitelli for help with the CHILDES statistics.

denominal verbs. *Natural Language and Linguistic Theory* 21, 737-778.

Aranovich, R. 2010. Incorporation, pronominal arguments, and configurationality in Fijian. Ms. UC Davis.

http://linguistics.ucdavis.edu/People/raranovi/papers-and-manuscripts/Fijian_NI_Lang10.pdf/view

Arkadiev, P., Y. Lander, A. Letuchiy, N. Sumbatova, and Ya. Testelelets. 2009. Vvedenie. Osnovnye svedeniya ob adygejskom jazyke. In Ya. Testelelets (ed.) *Aspekty polisintetizma: Očerki po grammatike adygejskogo jazyka* [Aspects of polysynthesis: Essays on Adyghe grammar], 17-120. Moscow: Russian University for the Humanities.

Aronoff, M. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.

Baker, M. C. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.

Biberauer, T., and I. Roberts. 2005. Changing EPP parameters in the history of English: accounting for variation and change. *English Language and Linguistics* 9, 5-46.

Biberauer, T., and I. Roberts. 2009. The return of the Subset Principle. In P. Crisma, and G. Longobardi (eds.). *Historical syntax and linguistic theory*, 58-74. Oxford: Oxford University Press.

Briceño, F. 2002. Topicalización, enfoque, énfasis y adelantamiento en el maya yucateco. In V. Tiesler Blos, R. Cobos and M. Greene Robertson (eds.). *La organización social entre los mayas prehispánicos, coloniales y modernos*, 374-387. Mexico City/Mérida: INAH/UADY.

Broschart, J. 1997. Why Tongan does it differently: Categorical distinctions in a language without nouns and verbs. *Linguistic Typology* 1, 123-65.

Chao, Y.-R. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chung, S. 2012. Are lexical categories universal? The view from Chamorro. *Theoretical Linguistics* 53.

Diksonera 1973. *Diksonera Malagasy-Englisy*. Antananarivo: Trano Printy Loterana.

Dryer, M. 1991. SVO languages and the OV:VO typology. *Journal of Linguistics* 27, 443-482.

Durbin, M., and F. Ojeda. 1978. Basic word order in Yucatec Maya. In N. England (ed.). *Papers in Mayan Linguistics*, 69-77. Columbia: University of Missouri.

- Galloway, B. D. 2009. *Dictionary of Upriver Halkomelem*, Volume I. Berkeley: University of California Press.
- Gil, D. 2005. Word order without syntactic categories: How Riau Indonesian does it. In A. Carnie, H. Harley, and S. Dooley (eds.). *Verb first: On the syntax of verb-initial languages*, 243-263. Amsterdam: John Benjamins.
- Gutiérrez-Bravo, R., and J. Monforte. 2008. La alternancia sujeto-inicial/verbo-inicial y la Teoría de Optimidad. In R. Gutiérrez-Bravo and E. Herrera (eds.). *Teoría de Optimidad: Estudios de Sintaxis y Fonología*, 61-99. Mexico City: El Colegio de México.
- Gutiérrez-Bravo, R., and J. Monforte. 2009. Focus, Agent Focus, and relative clauses in Yucatec Maya. In H. Avelino, J. Coon, and E. Norcliffe (eds.). *New Perspectives on Mayan Linguistics*, 83-95. Cambridge, MA: MIT Working Papers in Linguistics.
- Gutiérrez-Bravo, R., and J. Monforte. 2010. On the nature of word order in Yucatec Maya. In J. Camacho, R. Gutiérrez-Bravo and L. Sánchez (eds.). *Information structure in indigenous languages of the Americas*, 139-170. Berlin: Mouton de Gruyter.
- Haviland, J. 1994. "Te xa setel xulem"[The buzzards were circling]: Categories of verbal roots in (Zinacantec) Tzotzil. *Linguistics* 32, 691-741.
- Herring, S., and J. Paolillo. 1995. Focus position in SOV languages. In P. Downing and M. Noonan (eds.). *Word order in discourse*, 163-198. Amsterdam: John Benjamins.
- Hofling, C. 1984. On proto-Yucatecan word order. *Journal of Mayan Linguistics* 2, 35-64.
- Kaufman, D. 2009. Austronesian nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35, 1-49.
- Kayne, R. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kibrik, A. E., S.V. Kodzasov, V.A. Zvegincev, and I.F. Olovjannikova. 1977. *Arčinskij jazyk. Teksty i slovari*. Moscow: MGU.
- Lehmann, W. 1973. A structural principle of language and its implications. *Language* 49, 42-66.
- Lehmann, W. 1978. The great underlying ground-plans. In Winifred Lehmann (ed.). *Syntactic typology*, 3-55. Austin: University of Texas Press.
- Lois, X., and V. Vapnarsky. 2006. Root indeterminacy and polyvalence in Yucatecan Mayan languages. In X. Lois and V. Vapnarsky (eds.). *Lexical categories and root classes in Amerindian languages*, 69-116. Bern: Peter Lang.
- Long C., R. and S. Cruz. 1999. *Diccionario Zapoteco de San Bartolomé Zoogocho Oaxaca*. Coyoacán D.F.: Instituto Lingüístico de Verano.

Lupyan, G. and M.H. Christiansen. 2002. Case, word order and language learnability: Insights from connectionist modeling. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, 596-601. Mahwah, NJ: Lawrence Erlbaum.

Malvestitti, M. 2006. Polyvalence in Mapuzungun: Contributions from a Patagonian variety of language. In X. Lois and V. Vapnarsky (eds.). *Lexical categories and root classes in Amerindian languages*, 189-210. Bern: Peter Lang.

Miháltz M., C. Hatvani, J. Kuti, G. Szarvas, J. Csirik, G. Prószéky, T. Váradi. 2008. Methods and Results of the Hungarian WordNet Project. In A. Tanács, D. Csendes, V. Vincze, C. Fellbaum, and P. Vossen (eds.). *Proceedings of the Fourth Global WordNet Conference (GWC-2008)*.

Miller, G., R. Beckwith, C. D. Fellbaum, D. Gross, and K. Miller. 1990. WordNet: An online lexical database. *International Journal of Lexicography* 3, 235–244.

Minozzi, S. 2009. The Latin Wordnet project. *Proceedings of the 15th International Colloquium on Latin Linguistics (ICLL)*, Innsbrucker Beitrage zur Sprachwissenschaft.

Pala, K., D. Hlaváčková, and V. Němčík. 2008. Semi-automatic linking of new Czech synsets using Princeton WordNet. *Intelligent Information Systems 2008*, 369-374.

Pawley, A. 2006. Where have all the verbs gone? Remarks on the organisation of languages with small, closed verb classes. Paper presented at the 11th Binnennial Rice Univerity Linguistics Symposium, 16-18 March 2006.

Sasse, H.-J. 1993. Syntactic categories and subcategories. In J. Jacobs, A. von Stechow, W. Sternefeld, and T. Vennemann (eds.), *Syntax: An international handbook of contemporary research*, 646-685. Berlin: Mouton de Gruyter.

Schachter, P. 1985. Parts-of-speech systems. In T. Shopen (ed.), *Language typology and syntactic description, I: Clause structure*, 3-61. Cambridge: Cambridge University Press.

Seo, S.-K. 1998. Word frequency ranking for contemporary Korean words based on Yonsei corpora 1-9. *Technical report of Yonsei language & information studies*. Seoul: Yonsei University.

Siewierska, A. 1997. Variation in major constituent order: a global and an European perspective. In A. Siewierska (ed.). *Constituent order in the languages of Europe: Eurotyp 20-1*, 475-551. Berlin: New York: Mouton de Gruyter.

van Eberbroeck, E. 2006. Language type frequency and learnability from a connectionist perspective. *Linguistic Typology* 7, 1–50.

Vennemann, T. 1974. Analogy in generative grammar: the origin of word order. *Proceedings of the Eleventh International Congress of Linguists (1972)*, 79-83. Bologna: il Mulino.

Vennemann, T. 1976. Categorical grammar and the order of meaningful elements. In A. Juilland (ed.). *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*, 615-634. Saratoga, CA.: Anma Libri.

Vikner, S. 1994. Scandinavian object shift and West Germanic scrambling. In N. Corver and H. van Riemsdijk (eds.). *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena*, 487-517. Berlin: Mouton de Gruyter.

Xalilov, M. X. 1999. *Cezsko-russkij slovar'*. Moscow: Akademia.

Xu, R., Z. Gao, Y. Qu, and Z. Huang. 2008. An integrated approach for automatic construction of bilingual Chinese-English WordNet. *Lecture Notes in Computer Science* 5367, 302-314.

Williams, H. W. 1957. *A dictionary of the Maori language*. 6 ed. Wellington: Government Printer.

<http://nzetc.victoria.ac.nz/tm/scholarly/tei-WillDict-t1-front-d2.html>

Yoon, A., S. Hwang, E. Lee and H.-C. Kwon. 2009. Construction of Korean Wordnet 『KorLex 1.5』 Journal of the Korean Institute of Information Scientists and Engineers: Software and Applications 36 (1), 92-108.

Maria Polinsky
Department of Linguistics
Harvard University
Polinsky@fas.harvard.edu