**Coding**

The committee agreed to my request that you explain your coding decisions for each of the coded words in the short English and Nuuchahnulth passages.

This has now been added to the ends of Sections 3.3.2: English & 3.3.3: Nuuchahnulth, respectively.

Since the reason for coding is often the same for many words, it would be simplest to give each reason for coding, and list the words where this reason applies. The reasons will presumably be straightforward for most cases, but require a couple of sentences for the harder cases.

I liked the idea of this approach, and adopted it in my explanation of the coding decisions.

Below I give my notes where I questioned or disagreed with your coding decision. These are probably the ones where you’ll need a couple of sentences or so to justify your coding decision.

I’ve made sure to address each of the issues on specific data points/decisions that you raised in the description of how I coded the sample passages. In cases where I adopted your proposed change, I’ve also updated my coding of similar data elsewhere in the corpus, and adjusted the resulting counts and statistics as appropriate. (Thankfully the particular cases you pointed out happened to be very amenable to bulk find/replace, so it was fairly quick to do.)

Since I address each of your comments about individual data points and decisions in the dissertation text itself now, I won’t reply to all your comments about them individually here, but please let me know if you feel any haven’t been sufficiently addressed.

EDIT: Actually, I’ve replied to each of your English comments here, as well as addressed them in the dissertation.

The changes to the coding that arose as a result of your comments affected the statistical results as follows: {{TODO: Point out which counts/statistics changed.}}

**English**

line 1, *working*MOD: I would exclude because it is overtly coded (*-ing*) and a secondary predicate (not talked about, but doesn’t fit in the three-way classification of propositional act functions)

Including this token in the analysis was unintentional, a vestige of an earlier stage where I was coding postnominal participles as modifiers. I later realized, as you state, that these cases simply fall outside the 3-way classification and so I decided to exclude them. I thought I caught all these cases, but I checked and found a couple hundred more that I had missed for unknown reasons. I’ve now removed all these cases from the analysis.

line 2, *get*PRED *up*: *get up* is a lexicalized phrasal verb, so I would code it as *get*\_*up*PRED (p. 102)

line 4, *got*PRED *pretty bad* : *get* is grammatical (inchoative) and excluded (p. 99); it is arguably an aspectually modulated copula form and overtly coded so *bad* would be excluded from coding, as it was

Agreed on both of these. Honestly I couldn’t tell you why I originally coded them this way, since I was explicitly excluding phrasal verbs, and I clearly considered *got* in that case to be a copula, as you said. In any case, they’ve been fixed now, as well as half a dozen other phrasal verb uses I found in the small corpus sample. (*get* was not examined in the 100-item sample, so no changes were needed there.)

line 6, *looked*PRED *like soldiers*REF: *look like* is grammatical (evidential copular expression); *soldiers* might be excluded due to overt coding if *look like* is treated as a evidentially modulated copula

I don’t think I’d call this a copular expression (I’d maybe treat it as a lexicalized phrasal verb meaning ‘have the appearance of’), but I can certainly agree that *look like* is a complex expression of some sort, and that therefore *looked* here should be excluded from the analysis (but *soldiers* retained). I’ve adjusted the analysis here, and 335 places in the 100-item corpus where this applies as well.

line 7, *street*MOD *clothes*REF: this is a binominal lexeme, and *street* is part of that lexeme, because it is typifying. We discussed this case at length; my preference would be to either code it as a complex nominal (*street*\_*clothes*REF) or, to deal with the data sparsity issue of treating complex lexical items as distinct forms that you raised, just code the head (*street* *clothes*REF)

I agree with you on this and other similar cases of binominal lexemes. However, changing the coding of this case wouldn’t be consistent with how I treated multi-word compounds generally. To be consistent, I either need to leave this case as is, or recode all similar cases. Unfortunately, recoding those cases would involve a great deal of work, because it would require manual reinspection of each token. I think for this study I have to accept that my operationalization of compounds is imperfect, with the result that the English data likely display more reference-modification polyfunctionality than the language actually has. So, I’ve acknowledged this limitation in the section where I discuss my treatment of compounds {{give section and page number}}, and stated its potential effect. I’ve also explicitly discussed *street clothes* at length when presenting my coding decisions for the sample passage.

line 8, *“Those are guerillas*REF*!”* : a predicate nominal with a copula, so excluded because it is overtly coded

Not sure how I missed this case (I think I intended the annotation to go on the next token of *guerillas*, which I’ve now done). I checked for other cases where an item coded as a referent or modifier is immediately preceded by a form of *be*, and found 32 other cases to be corrected. These all appear to be coding errors. There were thousands and thousands of instances of copula constructions however, so thankfully this constitutes a very small error rate. (I’m sure there are plenty of other one-off errors I’ve made too, given the size of the data set. If I had funding to do so, I’d love to hire multiple research assistants to help with data coding, to ensure consistency.)

**Nuuchahnulth**

This is more challenging because there are some general issues of grammatical analysis that arose when I read Nakayama (2001), and they recur in the cases where I questioned or disagreed with your coding. I list the general issues first; we only discussed the first one in the defense. Because of the difficulty in dealing with complex predicates, I didn’t do a coding count. But I had no problems with 7/14 of the lines of the Nuuchahnulth sample, nor with many of the words in the remaining 7 lines.

*Verb serialization (we discussed this in the defense):* Nakayama describes serialization as conceptualization as a single state of affairs (Nakayama 2001:98; cf. p. 107), which is also a single complex predication (Nakayama 2001:99-100). Thus, a serial string should be treated as a single complex predicate, not a series of single separate preds. Nakayama calls each serialized element a “clause”, but warns the reader against considering this “clause” to be comparable to “clause” in other languages (i.e. as a comparative concept; Nakayama 2001:76). So this “clause” should not deter us from analyzed the serialized construction as a complex predicate. This is my preferred choice; you made a different choice, so please give the reason for your choice.

*Nominalizers:* Nakayama describes the Definite suffix as a nominalizer (2001:41), and Relative and Indefinite Relative suffixes as nominalizers (2001:36-37). Hence it appears they should be treated as overt coding of the reference function and any root/stem combined with them should be excluded from coding.

A last general question: why are ‘ “adverbials” (that is, predicate modifiers)’ (p. 111) coded in Nuuchahnulth (as pred), but excluded in English (p. 103)? Please add a sentence explaining your decision.

*Nuuchahnulth sample (comments below mostly based on general comments above)*

*p. 114*

*(i)* [time-done-rel.3] ‘when.it.occured’ : overtly coded referring phrase, so exclude

*(ii)* [sailing-into.a.bay sailing-cont there.mom-into.a.bay] ‘sailed into [the] bay of’ looks like serialization, so just one pred?

*(iii)* OK

*(iv)* [when-past-indef.3] ‘whenever.it.was’ : overtly coded referring phrase, so exclude

*(v)* looks like serialization, so just one pred?

*p. 115*

*(i)* OK

*(ii)* OK

*(iii)* not - grammatical element, exclude (we discussed this at length; please briefly give your justification for including it in Nuuchahnulth but not in English);

[large not-done large] - looks like serialization, so just one pred?

*(iv)* OK

*(v)* OK

*p. 116*

*(i)* [??-done-rel3-pl] ‘the way they spoke’ : overtly coded referring phrase, exclude;

why is [from] a pred?;

why isn’t [Chinook-belonging.to] either a mod or part of a complex ref?

*(ii)* OK

*(iii)* OK

*(iv)* [who-first-indef.3 hear-fin] : looks like serialization, so one unit; overtly coded referring phrase, so exclude;

[who-belonging.to-rel.3] : overtly coded referring phrase, so exclude

The committee recognizes that technically, one should also justify the choice not to code the remaining words. It is not clear how easy that would be, that is, if there are just a few reasons you would need to give, and list the noncoded words after each relevant reason. However, it is easy enough that I’ve pretty much done it here for you:

–For Nuuchahnulth, hardly any words were not coded, and there are only two reasons for noncoded words: they are names, or they were uninterpretable by Nakayama (p. 116, line (i)).

–For English, you could pretty easily explain almost all the words as grammatical function words (article, preposition, connective, personal pronoun) or as discourse markers (as you say on p. 99); plus some words that are overtly coded in their discourse function.

**Other changes**

The first two are really essential in my opinion. The third is highly desirable if possible in the short time available.

**(1) Tables 4.6 and 4.7**

‘Why is it that similar property words are so rigid in English yet so polyfunctional in Nuuchahnulth?’ (p. 162)

In part, it is because only 3 of the 13 words in Table 4.7 are properties, unlike Table 4.6; and those three words (‘big’, ‘aged’, ‘old’) are 3 of the 4 lowest-ranked words in that list (i.e. are less functionally diverse/polyfunctional). Please either note that only these three words are property concept words, or remove the non-property concept words from Table 4.7 and adjust the discussion in the text accordingly.

**(2) Typological markedness theory**

‘The marked member of a category may be literally marked with an affix or other overt morphological indicator, but this is just one of the ways an item can be a marked member of a category. The marked member of a category may also be less frequent, or have a smaller range of inflectional / distributional possibilities, *or show a semantic shift in the direction of the word’s prototypical function*’ (p. 72)

This is NOT part of typological markedness theory. Only how concepts are encoded formally is part of typological markedness theory. Concept identity is fixed in this approach. The generalization about semantic shift/extension in parts of speech in my 1991 book that you cite is an empirical observation about patterns of semantic shift, not definitional of typological markedness. Please remove the italicized part of this paragraph.

**(3) Triangle plots**

It would be helpful to have 3D versions of the triangle plots, in order to disentangle (near-)monofunctionality separately for reference, predication and modification. I don’t know how easy this would be to do quickly.

That would be really neat! I rummaged around StackExchange and other forums for an R package that would allow me to do this, but I couldn’t find one. I wasn’t entirely sure what you were requesting here though, so I emailed you about it, and will see if there’s another way to present the information you want.