Outline

Tuesday, October 15, 2019 8:51 PM

1. **Intro: What part of speech is *friend*?**
    a. Twitter Data
        i. Verb
            1) *What's your user? I would love to friend you and look at it when finished!*
            2) *If we don't have mutual friends we can't get friended.*
            3) *I accidentally downloaded Facebook and created a profile and friended a bunch of people*
        ii. Noun
            1) *I got a spooky box from my best friends.*
            2) *Secrets don't make friends, Luke.*
            3) *Just think I saw an old college friend on TV meeting Hilary Clinton*
        iii. Adjective
            1) *the guy became the national symbol of friend zone in just a day*
            2) *Facebook just put me in the damn friend zone with my wife*
            3) *can someone help me with some friend drama?*
    b. Dictionaries
        i. Dictionary.com - verb and noun, but not adjective
        ii. Merriam-Webster.com - verb and noun, but not adjective
    c. Why not adjectives? Should they be treated like compounds?
        i. *friend zone* vs. *friendzone*
            1) *friendzone* does not appear in the Google Books corpus
            2) Incidentally, Oxford Dictionaries reports that this can be a verb too!
        ii. *health care vs. healthcare*
            1) If compounds undergo univerbation, shouldn't *friend zone* as well?
            2) *just ate two slices of veggie pizza for lunch so basically I'm all healthed up for at least a month*
        iii. Conclusion: We simply don't traditionally analyze these uses as adjectives because historically they weren't, and they're morphologically unmarked.
            1) Linguists have been pretty selective about which criteria they use to determine parts of speech, and tend to cherry pick to accommodate either tradition or their particular theoretical approach (Croft; chapter from *Lexical polycategoriality* which critiques POS analyses).
    d. Problem for other words in English:
        i. *able*
            1) N: *that feeling of abling to run 22 miles a week*
            2) V: *always abling and abetting the horses*
            3) A: *an able mind overcomes challenges*
        ii. *time*
            1) N: *still one of my favorite series of all time*
            2) V: *I'm so bored in this class that I'm timing how long I can hold my breath*
            3) A: *2 years ago today (or yesterday depending on your time zone)*

iii. English is sometimes said / implied to have clear-cut parts of speech (Schachter & Shopen) and sometimes said / implied to have very flexible, fuzzy parts of speech (*Flexible word classes* volume?)

e. Problem for other languages than English

- [x] i. Nuuhchahnulth examples
- [x] ii. Iñuit examples
- [x] iii. Riau Indonesian examples
- [x] iv. Mundari examples
- [x] v. Chitimacha example
- [x] vi. Mohawk example
- [x] vii. I have yet to find a language where flexibility hasn't been at some point discussed explicitly in the literature, or discussed in a grammatical description (but went otherwise unnoticed)

2. Problem(s)

a. **Non-Problem I:** How should linguists analyze these cases? 2 approaches:

i. *conversion / zero-derivation* - traditional approach; often favored by generativist approaches (though note Distributed Morphology)

ii. *underspecification* - newer approach, gradually gaining attention

1) Hengeveld, Rijkhoff, Gil, other well-known cases

b. **Non-Problem II:** Determining lexemehood is tricky

i. Cognitive literature suggests a connection between different functional uses of the same form

ii. Categories are prototypal (Rosch; Lakoff; Taylor)

iii. we're not even sure when two *senses* of a word stop being related

1) Dictionary.com lists 148 senses of *run*, some nouns, some verbs (but again no adjectives)
2) fast pedestrian motion [1]: *I run every day*
3) conduct a political campaign [9]: *he ran a fair campaign*
4) come undone [17]: *these stockings run easily*
5) operate or function [29]: *does it run well?*
6) get or become [34]: *the well ran dry*

iv. Linguistic data itself suggests *grammatical* distinctions between the words:

1) categories of conversion match that of overt derivation
2) conversion involves item-specific knowledge - new meaning is unpredictable
   a) BUT, some languages do have predictably flexible lexemes
   b) BUT, *all* constructions have item-specific meanings when combined (e.g. singular vs. plural - plural forms have some meanings not available to singular forms)
      i) (historically) *brother* vs. *brethren* (now separate lexemes)
      ii) (historically) *cloth* vs. *clothes* (now separate lexemes)
      iii) (historically) *new* vs. *news*
      iv) *blind* (hunting) vs. *blinds* (window)
      v) *custom* vs. *customs*
      vi) *arm* vs. *arms* (military sense is not available in the singular)
      vii) *brain* vs. *brains*
      viii) *wood* vs. *woods* (don't refer to the same type of thing: one is a material, the other is a group of things that are made of the material)

c) BUT, at *some* point somebody *did* use a word flexibly, or do an instant conversion, or whatever, that allowed it to jump the POS boundary

v. Almost anything can be lexicalized (made into a new word)

1) [Interest in lexicalization as contrasted with grammaticalization. What can be lexicalized?]
2) cases of conversion / zero-derivation (already discussed)
3) cases of fully-inflected words (already discussed)
4) cases of entire phrases (already discussed with *friendzone*, but also *comeback*)

c. **Real Problem:** How common is flexibility? Does it vary within a language? Across languages?

i. Sidestep the issues of how to analyze these forms, and whether distinct uses are related.
ii. The question is, "How common is it for a word to be used for non-prototypical functions without any overt marking?"
iii. An immediate follow-up question is, what's special about the flexible words? Could it be semantic domain?
iv. Croft's markedness theory of parts of speech
v. We're interested in the cases which made Croft add the qualification "at *least* as marked" rather than simply state "*more* marked". Why aren't cases of derivation always marked? What determines when derivation is marked vs. unmarked? These are my long-term research questions, and this research is a very very very small step towards answering those, and only a first step as well.

3. **Solution** (what I'm doing)

a. Answering the first part: How flexible are words of English, and English generally?

i. For a given word, count how often that word is used in referring, predicating, and referring constructions in a large corpus (database) of English
ii. Can assign that word an overall flexibility score (0-1)
iii. For a given language, describe how flexible words are generally by using this metric

b. Answering the second part: Does flexibility correlate with semantic domain?
c. Laying the groundwork for:

a) more data (following the same methods)
b) additional correlations

4. Methods

a. *not* Twitter
b. Spoken portion of the Open American National Corpus (OANC) (3.5 million words)
c. Randomly selected wordforms from 100 different frequency bins

i. frequency vs. corpus dispersion

d. Created a list of every instance of those 100 lexemes (~380,000 total)
e. Annotated each token for its function: reference, predication, or modification

i. distinction between *token*, *wordform*, and *lexeme*
ii. problems [with examples]
iii. guidelines
iv. examples of how I coded data

f. Data annotation still in progress (about 1/3 done for English)

i. As a result, this talk is somewhat theoretical and methodological, with few concrete results

5. Results

a. tokens for specific lexemes
b. flexibility diagram for specific lexemes
c. flexibility score for specific lexemes
d. all flexibility diagrams for lexemes of English in one chart
e. flexibility diagram for all of English (so far)
f. flexibility score for English
    i. Can sum up my entire dissertation in 1 statistic
g. (potential) correlates between semantic domain [define] and flexibility
    i. [do statistics on body part terms vs. non-body part terms, as a dumb initial look at correlation]
    ii. So far *back* is the only lexeme which had a fairly even spread across functions
        1) Interestingly, I predicted this (just because of the great number of spatial and instrumental and other metaphors derived from body part terms), and it seemed to hold true

6. Conclusions

a. Most words of English do not exhibit much flexibility.
    i. They're a little boring! But that's okay! It says something about linguists' perception of English as a flexible language - that perception seems to be based on the very notable cases that stand out, rather than actual empirical data
    ii. Would the empirical data look any different for a language like, say, Nuuchahnulth?
b. All lexical words of English exhibit *some* flexibility (combination of inflectional possibilities for verbs + omnipredicativity for other categories ensures this)
    i. most words can predicate - almost feels like cheating!
c. Some semantic domains (appear to, preliminarily) exhibit more flexibility than others: body part terms

7. Next Steps

a. Nuuchahnulth
b. Other languages
c. More words from all languages
d. Diachronic spread of word uses across POS boundaries
    i. Notice that the verbal examples of *friend* are sense-specific - they're all about friending someone on social media. Can you use *friend* as a verb for in-person connections? Maybe.
e. [List some of your other near-term research questions]

8. Background / Context of Research

a. how I became interested in the problem
    i. POS tagging for RS
    ii. lexical categories course @ LSA Institute 2011
b. I don't typically work with English.