UD QA session guidelines 2-5-2020

https://padlite.spline.de/p/clingdingud

Questions:

1. Is Obj used for central arguments in terms of subcategorization frames? For example, 'put' requires a prepositional phrase location, would this be an obj or obl?
   - Essentially no, @obj is used for unmarked/core dependents of predicates, it corresponds to "second core argument" or "most patient-like argument"
   - https://universaldependencies.org/u/dep/all.html#al-u-dep/obj
   - iobj for Bantu languages with applicative extension is okay even though it expresses non-core arguments like benfeciaries and instrumentals as this is indicated by the verb's morphology (this example is specifically called out in the UD documentation)
   - https://universaldependencies.org/u/dep/all.html#al-u-dep/iobj

2. What should you do with things that are not really full sentences? (e.g. newspaper headlines or photo captions)
   - annotate them as if annotating fragments
   - try to go to the highest level of structure possible

3. Can you have multiple case arcs leaving a noun? "The ball rolled from under the chair" . Would that be a compound?
   - Look it up in the English treebank and see
     - Looks like the English example does case to the closest prep and then dep from that preposition to the next preposition
     - english GUM and english lines have examples "from over" / "from under"
   - Probably going to be flat with two case arcs
   - Add that to the UD github issues page

4. In case of polypersonal agreement, the basque treebank used Number[nom], Number[dat] etc for different cases. This seems to be a case driven approach but what if you have a language with no case system?
   - Number[obj] / Number[subj]

5. The distinction between fixed and compound seems fuzzy. Is it basically that compound is used for matching pos tags?
    - If the syntactic relationship between two words is unclear then using fixed is likely a good solution
    - compound is almost always only used for noun noun compounds

Xibe

1). How to calculate the annotate agreement between annotators?

- annotate the same sentences

2. Auxilaries: ombi (to become), sembi (to call), bimbi(to have) . The current annotation: no matter what words are in front of those auxilaries, we all annotate them as AUX.

ex. terei tacin tesei banse de, uju waka oci geli jai ombi. His study their class DAT, first is-not AUX also two AUX. (root of this sentence is 'jai', and 'ombi' depends on 'jai')/

```
   Do we need to annotate them differently?   a. when there is another VERB before these
 auxilaries, we annotate them as AUX.  b. when there are ADJ, NOUN before the auxilaries, we
 annotate them as VERB.
```

3). pospositions

Since the case markers are annotated as ADP, there are lots of pospositions, they usually need to collocate with certain case to convey a meaning. Ex: aimaka inenggi šun i adali eldešembi. like day light ADP ADP shine.

4).

mini gebu be Mutešan sembi. my name ACC Mutešan call. My name is Mutešan.

after 'my name' there is ACC marker, so 'obj' should be object of 'sembi', what is 'Mutešan' then? nsubj?

5). We have several words, the POS in both the dictionary and grammar book are not persuasive. Can we decide by our own linguistic knowledge? For example;

ilanofi ('ilan-nofi', three people, three things), it looks like a noun, in the grammar book, it is a NUM.

akv (is not + ADJ), waka(is not + NOUN), in dictionary, it is NOUN, but now we annoate them as VERB, and has a relation of 'cop' with the words in front of it.

Revision #4

Created Wed, Feb 5, 2020 7:04 PM by kenneth

Updated Thu, Feb 6, 2020 2:11 AM by kenneth