Morpho-syntactic analyso-parsing schema discussion guide
a.k.a. "the worksheet"

Authors: Omer Goldman, Leonie Weissweiler, Reut Tsarfaty

Sec. 1: organizing principles

We would like to create an annotation schema that does not refer to the wordhood of an element, since the definition of wordhood is highly contested. Thus, the main effort in creating data for morpho-syntactic analyso-parsing is in the unification of treatment of some categories of words/morphemes that are currently treated differently based the wordhood of each category.

Here is the list of categories to unified (potentially not complete, see below):

1. Case morphemes – adpositions – coverbs[1]
2. TAM morphemes – particles – modal verbs
3. Non-verbal TAM morphemes – copulas
4. Gender morphemes – noun class morphemes – classifiers[2]
5. Agreement morphemes – clitics – pronominal arguments
6. Arguments – incorporated arguments

The first 4 items in this list contain function words/morphemes, and in constructing a morpho-syntactic tree they will all be annotated with morphology-style features. The last 2 items contain arguments, these are content words/morphemes that are to be represented in independent nodes in a morpho-syntactic tree to make the predicate-argument structure recoverable.

Q1: Are there any other categories defined by wordhood that are not included in the list above?

Q2: Are there any categories from the list above that are included under the same item but you think they should be treated differently? If so, why?
It may be better to revisit these questions again after completing the annotation examples.

Sec. 2: annotation examples

In the next pages there are example sentences. Each given with its UD annotation and a morpho-syntactic tree. The added or changed morpho-syntactic features are in red. Function words are kept in the MS tree for readability, but the features refer only to the content word.[3]

Q3: Go over the sentences and their trees. Do you find anything questionable or objectionable?
These trees represent a proposal, everything is up for discussion.

---

[1] https://en.wikipedia.org/wiki/Coverb
[2] https://en.wikipedia.org/wiki/Classifier_(linguistics)
[3] The authors of this document are definitely not native nor fluent in all languages included in the document. Please be considerate towards any grammatical mistakes in the examples and do contact us for correction.
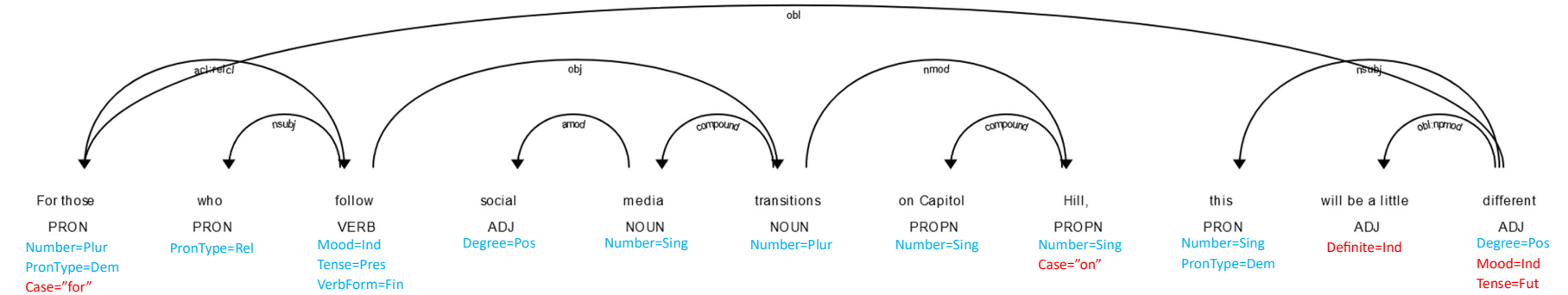
English: "For those who follow social media transition on Capitol Hill, this will be a little different."
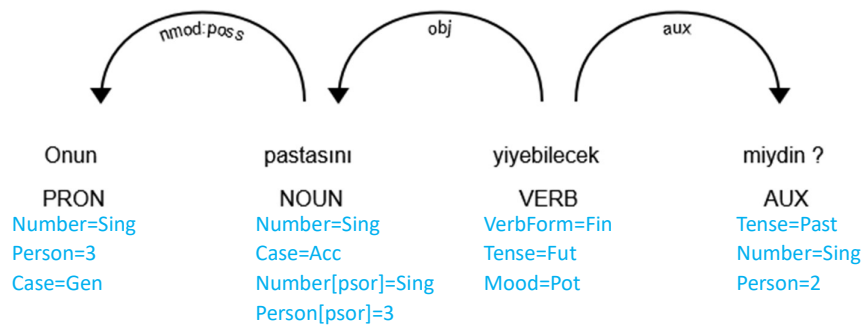
UD tree:
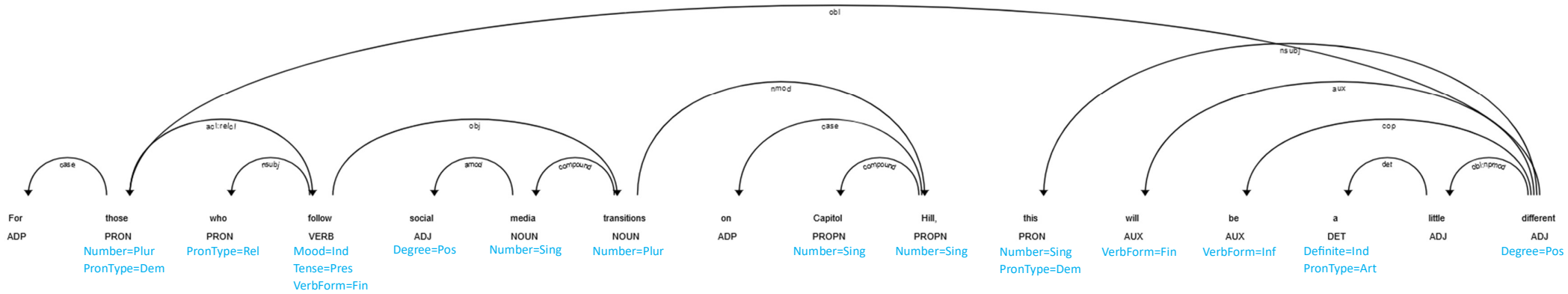
Morpho-syntactic tree:

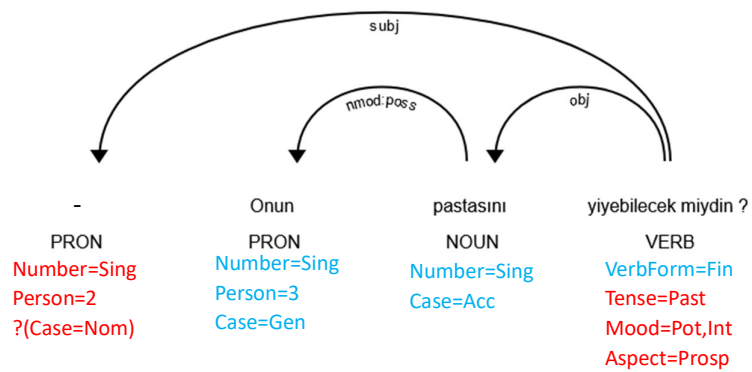Turkish: "Onun pastasını yiyebilecek miydin?"

On-un pasta-sı-nı yi-yebil-ecek mi-ydi-n?

He-GEN cake-POSS(3,SG)-ACC eat-can-FUT Qpart-PAST-NOM(2,SG)

Were you going to be able to eat his cake?

UD tree:

Onun	pastasını	yiyebilecek	miydin ?
PRON	NOUN	VERB	AUX
Number=Sing	Number=Sing	VerbForm=Fin	Tense=Past
Person=3	Case=Acc	Tense=Fut	Number=Sing
Case=Gen	Number[psor]=Sing	Mood=Pot	Person=2
	Person[psor]=3		

Morpho-syntactic tree:

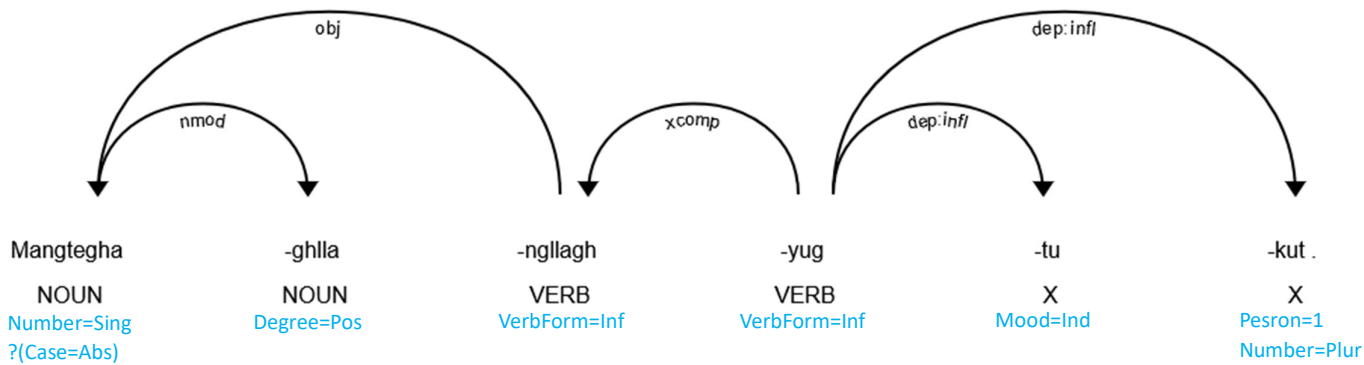| -	Onun	pastasını	yiyebilecek miydin ?
PRON	PRON	NOUN	VERB
Number=Sing	Number=Sing	Number=Sing	VerbForm=Fin
Person=2	Person=3	Case=Acc	Tense=Past
?(Case=Nom)	Case=Gen		Mood=Pot,Int
			Aspect=Prosp

Yupik: "Mangteghaghllangllaghyugtukut."

Mangtegha-ghlla-ngllagh-yug-tu-kut

House-big-make-want-IND-ABS(1,PL)

We want to make a big house.

UD tree:

Morpho-syntactic tree:
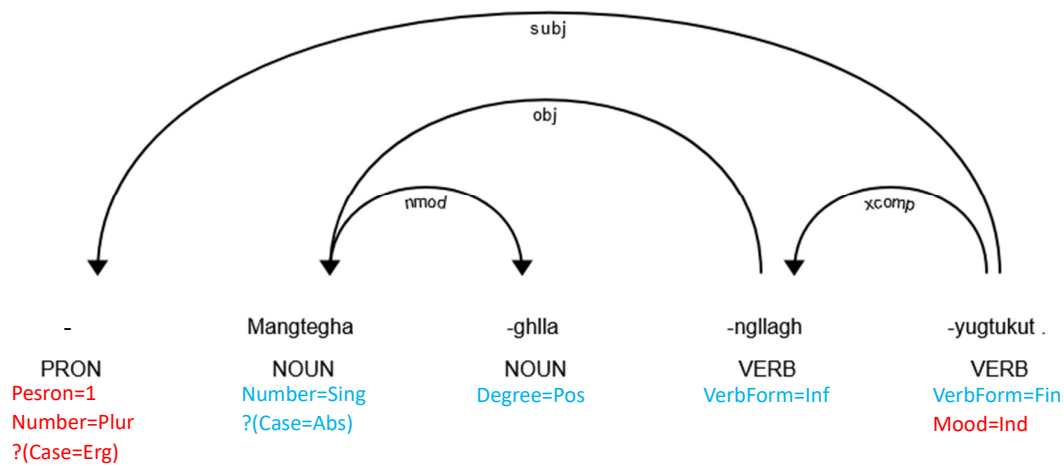
Sec. 3: New morpho-syntactic trees

Q4: Choose a few of the following sentences and draw their morph-syntactic trees (it may be easier to start with drawing their UD trees for reference).[3]

1. *Agreeing relativizer*
    German: Ich habe den Hund gestreichelt, der gestern gestorben ist.
    I have the-ACC-MASC dog petted, that-NOM-MASC yesterday died have.
    I have petted the dog that died yesterday.

2. *Coverbs + classifiers*
    Mandarin: 我给你收买三只猫.

    Wǒ gěi nǐ shōumǎi sān zhī māo.
    I for/give you buy three CL(animals) cat.
    I buy three cats for you.

3. *Prepositional predicates*
    Hebrew: הוא בבית.
    hu babayit.
    hu b-a-bayit.
    he in-the-house.
    he is in the house.

4. *Multiple copula types*
    Spanish: Sé que está loco.
    Know,PRES,IND,NOM(1,SG) that is,PRES,IND,NOM(3,SG) crazy.
    I know that he is crazy. (i.e., currently out of his mind)
    (c.f. "Sé que es un loco" = I know he is a crazy person)

5. *Expletives*
    English: It is my birthday today.
    (c.f. "today is my birthday")

6. *Case-role mismatch*
    Icelandic: Drengina vantar mat.
    Boy,PL,ACC,DEF lack,PRES,3,SG food,Sing,ACC,INDEF.
    The boys lack food.

7. *Consecutive verbs*
    Turkish: su içip bir muz yiyeceksin.
    su iç-ip bir muz yi-yecek-sin.
    water drink-and a banana eat-FUT-NOM(2,SG).
    You will drink water and eat a banana.

8. *conjugations vs. adpositions*
    Swahili: Nitaliwa na jini na simba.
    Ni-ta-li-wa na jini na simba.
    NOM(1,SG)-FUT-eat-PASS with genie with lion.
    I will be eaten by a genie and a lion.

9. *Cases vs. adpositions*
   Finnish: Asun mäen pohjalla.
   Asun mäen pohja-lla.
   Live,IND,PRES,NOM(1,SG) hill,GEN bottom-ADE/on.
   I live on the bottom of the hill.

10. *Non-grammaticalized determiners?*
    English: some men eat ice-cream.

Q5: Are there any additional sentences that exhibit an interesting phenomenon that we didn't take into account? Maybe something rare that happens to occur in a language you know, or a combination of phenomena that creates a possibly problematic case?
We encourage you to devote extra time to this question. This is where we'd like to see holes poked in the schema, to make sure we are being as inclusive as possible.

For any inquiries regarding the document, and for registration to the discussion group's mailing list please contact Omer, at omer.goldman@gmail.com, or Leonie, at weissweiler@cis.lmu.de.