

Replicating Macaulay & Brice (1997)

Systematization Notes

Hadas Kotek

PROGRESS group, 12/11/2018

1. Language

- Two specifications:

E	English
O	Other

- Annotating the language for each example doubles as a way to keep track of which rows have been looked at, for those rows that have no other specification

2. Argument

- An argument is the nominal elements that is part of the argument structure of a predicate.
- Every row in the file corresponds to an argument in the examples.
- The column includes full DPs, pronouns, PRO and Op (the latter for arguments in tough movement).
- Arguments appearing between brackets ([...]) are reconstructed covert arguments. For example, for *She finds and helps women*, the arguments of *find* are *she* and *[women]*, and the arguments of *help* are *[she]* and *women*.
- This column is intended to have no typos even if the original example has one.

3. Predicate

- The predicate can be a verb, adjective, adverb, etc.
- There is no special notation for reconstructed predicates due to ellipsis. For example, for *Kim could run a marathon, and so could Lee*, there is a row for *Kim* and its predicate is *run*, and there is another row for *Lee* and its predicate is *run*.
- Full prepositional phrases appear when the lexical verb is a copula.

4. Grammatical function

- Three specifications:

S	Subject of a full clause or a small clause
O	Direct object of the verb
Q	Oblique: indirect object, object of a preposition, etc.

- *Talk, lie, go, come* take obliques (indirect objects) with *to*.
- Oblique indirect objects of ditransitives may be marked by different prepositions (such as *to* for *give*, *in* for *put*, *with* for *spray*, *from* for *differentiate*, etc.) or not be marked by any preposition (as in double object constructions).

5. θ -role

- Four specifications with acronym PEAR (definitions from M&B):

P	Patient	an entity that undergoes an action or is described
E	Experiencer	the individual who feels or perceives the event
A	Agent	the initiator, doer of action; must be capable of volition or deliberate action
R	Recipient	a person who receives something (concrete or abstract)

- Roles have been assigned according to the nature of the predicate. For example, the external argument of *take* is an agent in *take a book*, but an experiencer in *take a class*; for *arrive*, people arriving are agents, buy things arriving are patients; for *leave* meaning 'depart', the direct object is a source, but when meaning 'abandon, not bring', the direct object is a patient.
- The indirect object of ditransitives may be marked as recipient

6. Gender

- Four specifications:

F	Feminine
M	Masculine
A	Not gendered or ambiguous
U	Unfamiliar (e.g. a name for which you don't know it's gender, if any)

- *Andrea, Chris, Dr. Kelly, Guojing, Jean, Kim, Lee, Mel, Morgan*, are considered not gendered or ambiguous proper names except when a pronoun or other element suggests feminine or masculine.
- Arguments that are not names, are not gendered, and have no other element like a pronoun suggesting gender, such as *the teacher* and *four students*, are considered not gendered or ambiguous

7. Number

- Three specifications:

S	Singular
P	Plural (including dual)
U	Unknown/ambiguous

- Singular is the default number. It applies to *what* and *who* in questions, *it* when refers to a clause, mass nouns (*wine, porridge...*), etc.
- Relative uses of *who* are marked with the number of the antecedent (*The person who...*)

8. Proper name

- Two specifications:

Y	Yes, the argument is a proper name
N	No, the argument is not a proper name

- Names of places, teams, bands, institutions, languages, holidays, as well as book titles are proper names.

9. Pronoun

- Two specifications:

Y	Yes, the argument is a pronoun
N	No, the argument is not a pronoun

- Pronouns also include interrogative or relative pronouns such as *Who made these cookies?* and *Robin is the student who has taken every linguistics class.*

10. Content

- Eight specifications:

V	Violence
A	Appearance (physical)
O	Occupation
R	Romantic/sexual/physical intimacy

B	Books (reading, writing, book-related)
E	Emotion (positive, negative, etc.)
K	Kinship term (mother, father, sister, grandparent, etc.)

- This category codes several content areas of interest. An argument can be coded for more than one content area.
- Content is at the level of the predicate. In the sentence *Tenzin hit Kelsang*, both the arguments of *Tenzin* and *Kelsang* would be marked with **V** for violence.
- **A** (Appearance) codes anything involving a person's physical appearance, such as describing or assigning a value judgment to it.
- **O** (Occupation) describes any situation where an argument has an occupation or profession, such as *Alice is a carpenter* or *The taxi driver didn't know what time it was*.
- Examples where an argument is experiencing an emotion are coded as **E** (emotion).
- Any example featuring widely held negative or oversimplified representations, such as those based on gender is coded as **S** (stereotype).
- Some arguments are kinship terms, such as *the mother* or *John's grandchildren*; these are coded as **K**.

11. Comments

- This column can include any notes you think will be useful later. It shouldn't be used that often

Practical notes:

- Ignore really fragmentary things that look like they may have not been extracted properly (because we used automated extraction)
- If there is light syntactic markup, but you can still read/infer the intended surface sentence, then code the sentence as normal.
 - o Example: indices, brackets
- If the example is an interlinear gloss/is in a non-English language with English translation, code the English translation
- "Pipes" — the character “ | ” — indicate line breaks in the original PDF. If something is in a language other than English, or seems jumbled, jump to end of the example and there may be a gloss/translation in quotation marks. This is the sentence you should use.
- If you find sub-examples b, c, d, etc. under one entry, and they are also coherent sentences, split them out into new entries. (You can also append the letter to the example number if there are any, so "35" > "35a", "35b", "35c", etc.)
- If the example sentence is there, but there are is additional noise/nonsense that was appended by accident, then code the sentence as normal and ignore/delete the extra stuff.
- If the sentence is there but followed by many piped lines that appear to be the next paragraph in the article (or start of it), code as normal and delete the extra stuff.
- Ignore sentences that are not true example sentences (illustrations of some linguistic/syntactic concept), but rather descriptions of syntactic theory, etc. We weren't able to filter all of those out automatically.
 - o Example:
(56) Def can attach as sister to A, N, and maximal NP.