



university of
 groningen

faculty of arts

Tagging Syntactic Complexity:the XCSY tier

Authors:

Rasmus Steinkrauss (University of Groningen)

Cornelia Lahmann (University of Groningen)

Project:

L1 attrition and L2 acquisition: a perspective from *Kindertransport* survivors

Date: 01-12-2014

Content:

Introduction	3
Basics I: Clauses, AS-units, utterances	3
Basics II: Where to put the xcsy-tier	3
UTT, AS, Clause type.....	5
Word count	10
Subject	11
Verbs.....	13
Noun phrases	14
Embedded clauses.....	17
Passives.....	19
German Konjunktiv	20
German complex verb constructions.....	21
DC-RELS	22

Introduction

Utterances people produce show varying degrees of syntactic complexity. For example, the utterance “Here!” as an answer to “Where are you?” is considerably less complex than “Knowing that you’ll make me take the dog, who has been waiting for this for hours, for a walk, I won’t be telling you where I am”.

By measuring the syntactic complexity of a speaker’s utterances, we hope to capture an aspect of that speaker’s language proficiency. Therefore, it is useful to measure syntactic complexity in our data.

These guidelines explain how to annotate our data for syntactic complexity. This annotation happens not on the main tier line – i.e. the line containing the utterance – but on an extra line, the syntactic complexity tier line. It starts with %xcsy: for Complexity/Syntax.

Basics I: Clauses, AS-units, utterances

The annotation relies on a correct splitting up of the utterances into main tier lines. To recall: each main tier line contains one clause (generally speaking). If an utterance consists of several clauses, the main tier lines are to be connected using +/. and +, .

- [1] *ABC: I don’t understand +/. Clause 1 (main clause)
 *ABC: +, what he is saying ! Clause 2 (dependent clause)

The above utterance consists of one main clause with one dependent clause. Other utterances may consist of something that is not even a clause:

- [2] *ABC: in Minnesota@z:geo . Non-clause (as answer to *where was that?*)

Not only main clauses, also non-clauses may govern a dependent clause:

- [3] *ABC: in Minnesota@z:geo +/. Non-clause
 *ABC: +, where my husband comes from. Dependent clause

One main or non-clause together with all its dependent clauses forms an AS-unit (analysis of speech unit). An utterance may consist of several AS-units:

- [4] *ABC: I asked him +/. AS-unit 1, main clause
 *ABC: +, but he didn’t tell me +/. AS-unit 2, main clause
 *ABC: +, where he was going +/. AS-unit 2, dependent clause
 *ABC: +, so I don’t know. AS-unit 3, main clause

Any utterances that are split up incorrectly need to be changed for syntactic complexity to be annotated correctly!

Basics II: Where to put the xcsy-tier

The syntactic complexity of a clause is noted in the syntactic complexity tier (xcsy-tier). There is exactly one xcsy-tier for each clause (or non-clause). The xcsy-tier follows each clause directly on the next line.

- [5] *ABC: I don’t understand +/. Main tier for clause 1
 %xcsy: ... Complexity tier for clause 1

*ABC: +, what he is saying ! Main tier for clause 2
 %xcsy: ... Complexity tier for clause 2

If a clause is split over several main tier lines because another clause intervenes, the xcsy-tier follows the first main tier of that clause. Note that there still is only one xcsy-tier for every clause:

[6] *ABC: He is +/. Clause 1, Part 1
 %xcsy: ... Complexity tier for all of clause 1
 *ABC: +, as far as I know +/. Clause 2
 %xcsy: ... Complexity tier for clause 2
 *ABC: +, a successful painter . Clause 1, Part 2

UTT, AS, Clause type

UTT

The first xcsy-tier in an utterance starts with **UTT** to mark the beginning of an utterance.

[7] *ABC: I asked him .
 %xcsy: **UTT**|AS|MC:3|SUBJ:1|V:1:0|NP:2:2|

AS

The first clause in an AS-unit gets the tag **AS** to mark the beginning of a new AS-unit. Tags like UTT and AS are generally separated by | .

[8] *ABC: I asked him +/.
 %xcsy: **UTT|AS**|MC:3|SUBJ:1|V:1:0| NP:2:2| Start of utterance, start of AS-unit 1
 *ABC: +, but he didn't tell me .
 %xcsy: **AS**|MC:6|SUBJ:1|V:1:1| NP:2:2| Still same utterance, start of AS-unit 2

MC
DC
XC

Each clause is then tagged for clause type:

MC = main clause, **DC** = dependent clause, **XC** = non-clause

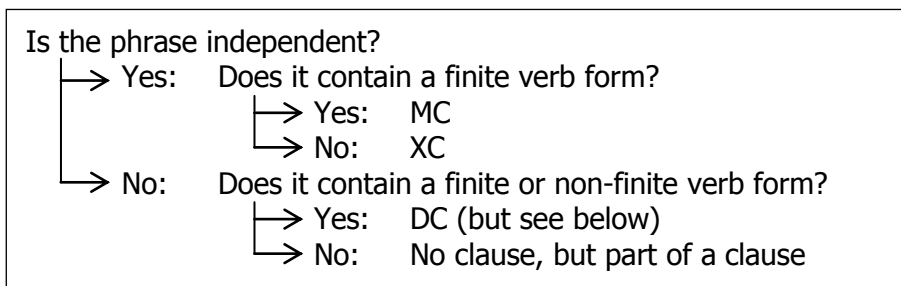
[9] *ABC: I asked him +/.
 %xcsy: **UTT|AS|MC**:3|SUBJ:1|V:1:0| NP:2:2| Start of utt. and AS-unit 1, main clause
 *ABC: +, but he didn't tell me +/.
 %xcsy: **AS|MC**:6|SUBJ:1|V:1:1| NP:2:2| Start of AS-unit 2, main clause
 *ABC: +, where he was going +/.
 %xcsy: **DC**:4|SUBJ:1|V:1:1| NP:1:1| Still AS-unit 2, dependent clause
 *ABC: +, so I don't know.
 %xcsy: **AS|MC**:5|SUBJ:1|V:1:1| NP:1:1| Start of AS-unit 3, main clause

DC-
REL

Dependent clauses that are relative clauses are tagged as **DC-REL** (also note the use of XC for non-clauses):

[10] *ABC: in Tuebingen@z:geo +/. Non-clause
 %xcsy: **UTT|AS|XC**:2|SUBJ:0|V:0:0| NP:1:7|
 *ABC: +, which is a really old city . Relative clause
 %xcsy: **DC-REL**:6|SUBJ:1|V:1:0| NP:2:5|

To decide whether a phrase is a MC, DC or XC, you can use the following flow sheet:



This leads to the following definition of clause types:

MC: Independent main clause with a finite verb

- This includes trailing off and incomplete utterances as long as there is a finite verb form. The subject of a MC is optional.
- A MC may contain more than just one finite verb if the subject stays the same and is not repeated:
 One MC: [We stopped by and had a coffee.]
 Two MC: [We stopped by] [and we had a coffee.]
[We stopped by] [and I had a coffee.]
- The amount of non-finite verbs is irrelevant:
 One MC, one finite, two non-finites: [You can get lost on the ship and go to places.]

XC: Independent non-clausal phrase without a finite verb

- This includes pragmatically appropriate as well as inappropriate / ungrammatical / incomplete phrases without a verb or only a non-finite verb. A XC does not normally contain a subject. Cf. [2], [10] and [16] for examples.

DC: Dependent clause with a finite or non-finite verb

- A dependent clause is a verb phrase that fulfills a syntactic role within a MC or XC. A DC is always governed by a MC/XC and can never occur on its own.
- Clauses that have the form of a dependent clause but occur on their own are tagged as MC:

[11] *ABC: because I forgot. Main clause (as answer to *why didn't you call?*)
 %xcsy: UTT | AS | MC : 3 | SUBJ : 1 | V : 1 : 0 | NP : 1 : 1 |

[12] *ABC: I didn't call +/. Main clause
 %xcsy: UTT | AS | MC : 4 | SUBJ : 1 | V : 1 : 1 | NP : 1 : 1 |
 *ABC: +, because I forgot. Dependent clause
 %xcsy: DC : 3 | SUBJ : 1 | V : 1 : 0 | NP : 1 : 1 |

- The role of the DC within the MC/XC is either that of an adverb, a complement (subject/object/predicative), or a relative clause. Anything else does not qualify as a DC. Relative clauses are tagged as DC-REL.

Adverbial DCs fulfill the role of an adverb. That means that they specify the circumstances under which the event in the MC took place (like time, place, condition etc.) or the way it took place. They are not strictly necessary to complete the MC.

[13] *ABC: He came over +/. Main clause
 %xcsy: UTT | AS | MC : 3 | SUBJ : 1 | V : 1 : 0 | NP : 1 : 1 |
 *ABC: +, while I sat there. Adverbial DC: specifies time of event in MC
 %xcsy: DC : 4 | SUBJ : 1 | V : 1 : 0 | NP : 1 : 1 |

Complement DCs fulfill the role of a subject, object or predicative in the MC. They are necessary to complete the MC. (A predicative is the part that usually follows a linking verb (copula) like *be*, *stay*, *become*, *be called* etc.; cf. [40], [41].)

[14] *ABC: She did +/. Main clause
 %xcsy: UTT|AS|MC:2|SUBJ:1|V:1:0|NP:1:1|
 *ABC: +, what I hated most. Complement DC: direct object to *did* in the MC
 %xcsy: DC:4|SUBJ:1|V:1:0|NP:2:2|

Relative DCs refer to a noun phrase in the MC and give extra information about it.

[15] *ABC: The guy +/. Main clause
 %xcsy: UTT|AS|MC:4|SUBJ:5|V:1:0|NP:1:6|
 *ABC: +, who wears a hat +/. Relative DC: modifies *guy* in the main clause
 %xcsy: DC-REL:4|SUBJ:1|V:1:0|NP:2:3|
 *ABC: +, looks tired . (Part of the main clause; no %xcsy-tier, cf. [6])

- Relative clauses normally start with a relative pronoun/adverb. In English, this may be omitted in some cases.

[16] *ABC: At the time +/. Non-clause
 %xcsy: UTT|AS|XC:3|SUBJ:0|V:0:0|NP:1:6|
 *ABC: +, when I was born . Relative DC with relative adverb: modifies *time*
 %xcsy: DC-REL:4|SUBJ:1|V:1:1|NP:1:1|PASS|

[17] *ABC: The lady +/. Main clause
 %xcsy: UTT|AS|MC:5|SUBJ:4|V:1:0|NP:1:5|
 *ABC: +, I talked to +/. Rel. DC without rel. pron./adv.: modifies *lady*
 %xcsy: DC-REL:3|SUBJ:1|V:1:0|NP:1:1|
 *ABC: +, was really nice .

- Some DC only have a non-finite verb, no finite verb (see also Verbs, p. 12).

[18] *ABC: Entering the store +/. Adverbial DC
 %xcsy: UTT|AS|DC:3|SUBJ:0|V:0:1|NP:1:2|
 *ABC: +, she ordered him +/. Main clause
 %xcsy: MC:3|SUBJ:1|VP:1:0|NP:2:2|
 *ABC: +, to hand him the money. Complement DC
 %xcsy: DC:5|SUBJ:0|VP:0:1|NP:2:3|

[19] *ABC: The girl +/. Main clause
 %xcsy: UTT|AS|MC:4|SUBJ:4|V:1:0|NP:1:5|
 *ABC: +, entering the store +/. Relative DC
 %xcsy: DC-REL:3|SUBJ:0|V:0:1|NP:1:2|
 *ABC: +, looked frightened. (Part of the main clause)

[20] *ABC: To get out of here +/. Complement DC (see remark for [27])
 %xcsy: UTT|AS|DC:5|SUBJ:0|V:0:1|NP:0:0|
 *ABC: +, was all +/. Main clause
 %xcsy: MC:2|SUBJ:3|V:1:0|NP:1:3|
 *ABC: +, I wanted. Relative DC (with finite verb)
 %xcsy: DC-REL:2|SUBJ:1|V:1:0|NP:1:1|

- In some cases, the object of the MC functions as the logical subject of a non-finite DC. Under such circumstances the logical subject is moved to the DC.

In [21], *her* is the logical subject of the non-finite DC (she is doing the liking / travelling). Although *her* is not in nominative case, it is counted as the subject of the DC.

[21] *XYZ: as a child, she [///] (...) you know, I remember +/.
 %xcsy:UTT|AS|MC:6|SUBJ:1|V:1:0|NP:2:3|
 *XYZ: +, her liking the good life and travelling and so on.
 %xcsy:DC:10|SUBJ:1|V:0:2|NP:2:4|

- Nominalized verbs (gerunds in English) are not considered dependent clauses when they are introduced by a determiner:

In [22], *the swimming in the lake* is not a dependent clause because *swimming* is used with the article *the*. This is also why it is not considered a verb form in this case (cf. [49]; see also Verbs, p. 12). In [23], *swimming in the lake* does not have a determiner and is therefore considered a complement DC:

[22] *ABC: The swimming in the lake was fun . One main clause only
 %xcsy:UTT|AS|MC:7|SUBJ:4|V:1:0|NP:2:6|

[23] *ABC: Swimming in the lake +/. Main clause
 %xcsy:UTT|AS|DC:4|SUBJ:0|V:0:1|NP:1:2|
 *ABC: +, was fun . Dependent clause
 %xcsy:MC:2|SUBJ:4|V:1:0|NP:1:1|

- A DC always contains more than just the verb. If it doesn't, it is part of the MC/XC.

Although *to ask* in [24] is the complement to *(have a) chance*, it is not considered a DC because it only consists of the (non-finite) verb form *to ask* and nothing else. In [25] however, the complement consists of more than just the verb and is therefore considered a DC.

[24] *ABC: I had no chance to ask. Main clause
 %xcsy:UTT|AS|MC:6|SUBJ:1|V:1:1|NP:2:3|

[25] *ABC: I had no chance +/. Main clause
 %xcsy:UTT|AS|MC:4|SUBJ:1|V:1:0|NP:2:3|
 *ABC: +, to ask him . Complement DC
 %xcsy:DC:3|SUBJ:0|V:0:1|NP:1:1|

- Verb phrases governed by the following ((semi-)auxiliary and modal) verbs are never considered dependent clauses, irrespective of their length:

English: *will/would, can/could, shall/should, may/might, must
 want to, have to, ought / used to
 get, dare, need, have better
 be able / allowed / supposed / about / likely to, be going to,
 seem / tend / happen / mean / turn out to*

German: *können, müssen, sollen, wollen, mögen, dürfen
 haben, werden, lassen, brauchen, bleiben, helfen, gehen, kommen, fahren,
 sehen, hören, lernen, lehren, spüren, fühlen*

[26] *ABC: I wanted to leave quickly. Main clause.
 %xcsy: UTT|AS|MC:5|SUBJ:1|V:1:1|NP:1:1|
To leave quickly is no DC because it is governed by *want*.

[27] *ABC: I wished +/. Main clause
 %xcsy: UTT|AS|MC:2|SUBJ:1|V:1:0|NP:1:1|
 *ABC: +, to leave quickly. Complement DC, governed by *wish*
 %xcsy: DC:3|SUBJ:0|V:0:1|NP:0:0|

To get out of here in [20] is a DC because it is not governed by *I want* but by *was* – it is the predicative: [SUBJ All I wanted] [was] [PRED to get out of here] (cf. [40], [41]).

Word count

Every MC/XC/DC tag is followed by a colon : and the number of words in the clause.

[28] *ABC: I have no idea .
 %xcsy: UTT|AS|MC:4|SUBJ:1|V:1:0|NP:2:3|
 4 words: *I, have, no, idea*

- **Contracted forms** with apostrophe are counted as the number of words they consist of; **incomplete words** are counted as if they were complete.

[29] *ABC: I'm not sure about (th)em !
 %xcsy: UTT|AS|MC:6|SUBJ:1|V:1:0|NP:2:2|
 6 words: *I, am, not, sure, about, them*

- **Repeated words [//] and [x N]** are counted only once, **material linked with + or _** is counted as one word.

[30] *ABC: we lived in the [//] the Berliner_Strasse@z:geo .
 %xcsy: UTT|AS|MC:5|SUBJ:1|V:1:0|NP:2:3|
 5 words: *we, lived, in, the, Berliner_Strasse*

- The language of the words does not matter.
- Material marked with **@o, @z:... and as quotations [""]** is counted normally
- **English numbers** (cf. [61]) are counted as one word.

[31] *ABC: also ich hab(e) angefangen als farmhand@s in
 Ontario@z:geo in [*] nineteen@z:num@s fifty@z:num@s.
 %xcsy: UTT|AS|MC:10|SUBJ:1|V:1:1|NP:2:4|
 10 words: *also, ich, habe, angefangen, als, farmhand, in, Ontario, in, nineteen fifty*

- Don't count the following material:
retracings [//], reformulations [///]
xx / xxx / www
everything marked with @l / @fp / @i
material preceded by &
missing words marked with 0.

[32] *ABC: ah@fp <all of this was> [///] I_mean@z:ep we didn't &r
 really have a chance xxx.
 %xcsy: UTT|AS|MC:8|SUBJ:1|V:1:1|NP:2:3|
 8 words: *I_mean, we, did, not, really, have, a, chance*

[33] *ABC: oh@i it's spelled s@l t@l r@l a@l u@l s@l s@l.
 %xcsy: UTT|AS|MC:3|SUBJ:1|V:1:1|NP:1:1|PASS|
 3 words: *it, is, spelled*

[34] *ABC: <one@z:num month> [///] two@z:num months later we moved
 0to [*] New_York@z:geo.
 %xcsy: UTT|AS|MC:6|SUBJ:1|V:1:0|NP:3:4|
 6 words: *two, months, later, we moved, New_York*

Subject

SUBJ

Every MC/XC/DC is tagged for the occurrence or non-occurrence of a subject. Main and dependent clauses normally contain a subject; non-clauses normally don't.

A clause can contain maximally one subject; coordinated phases like *my mother and I* are one subject.

The subject tag is **SUBJ:x** with x = number of words that the subject is made of: 0 = no subject; 1, 2, 3... = number of words in the subject.

When calculating the number of words, don't count the definite or indefinite article of the main (= head) noun in the subject. Examples:

- 1 word: *I, we, boys, Dave, the guy, a girl*
- 2 words: *my dad, an old lady, we all, all the others*
- 3 words: *the girls at school, my high_school teacher, she and I*

In addition, the normal rules for word count apply.

Special cases:

- Relative clauses and appositions modifying the subject are included in the count (cf. also [17], [19] and [51] for rel. clauses and [50] for an apposition):

- [35] *ABC: Pete@z:per +/.
 %xcsy: UTT|AS|MC:5|**SUBJ:8**|V:1:0| 8-word subject in the main clause: *Pete...me*
 NP:3:10|
 *ABC: +, who wasn't much older than me +/.
 %xcsy: DC-REL:7|**SUBJ:1**|V:1:0| 1-word subject in the dep. clause: *who*
 NP:2:2|
 *ABC: +, basically showed me everything .
- [36] *ABC: and the [/] the officer and the other guy +/.
 %xcsy: UTT|AS|MC:12|**SUBJ:10**|V:1:0| 12-word subject in the main clause: *the...house*
 NP:2:12|
 *ABC: +, standing in_front_of our house +/.
 %xcsy: DC-REL:4|**SUBJ:0**|V:1:0| no overt subject in DC
 NP:1:2|
 *ABC: +, looked up to our window .

- If the subject NP is followed by a pronoun referring to the whole subject NP, the pronoun (including its post-modifiers) is counted as subject. For the NP count, the subject NP and the pronoun are counted as two separate noun phrases.

- [37] *ABC: my brothers and sisters +/.
 %xcsy: UTT|AS|MC:6|**SUBJ:2**|V:1:0|**NP:2:9**|
 *ABC: +, who stayed in Russia, +/.
 %xcsy: DC-REL:4|**SUBJ:1**|V:1:0|**NP:2:2**|
 *ABC: +, they all died.

Subject in the MC: Not *my brothers and sisters...*, but *they all* (2 words)
 Two NPs in the MC: *my brothers and sisters who stayed in Russia, they all* (9 words)

- Subjects containing *all*, *both* etc. are sometimes split by the verb. They should still be tagged as one subject:

[38] *ABC: they have all left the country.

%xcsy: UTT|AS|MC:6|**SUBJ:2**|V:1:1|NP:2:4| subject: *they all* (2 words)

- The subject of a MC may also consist of a complement DC.
(This means that complement DCs that are a subject are counted as a subject! This is different from how NPs are counted: a complement DC is never counted as a NP in the main clause.)

[39] *ABC: having gotten out of there +/.

%xcsy: UTT|AS|DC:5|**SUBJ:0**|V:0:2| no overt subject in DC
NP:0:0|

*ABC: +, was a relief .

%xcsy: MC:3|**SUBJ:5**|V:1:0| The 5-word DC is subject of the MC
NP:1:2| (but no NP in the MC) (cf. [20], [66])

- In copula constructions like [39], the subject may sometimes be hard to identify. Use common sense to decide what the subject should be.

E.g., the subject in [20] *to get out of here was all I wanted* may either be *to get out of here* or *all I wanted*. If *all I wanted* is considered the subject, the utterance would be tagged as in [40] (= [20]). If *to get out of here* is considered the subject, the tagging would be as in [41]:

[40] *ABC: To get out of here +/.

Complement DC: **predicative**

%xcsy: UTT|AS|DC:5|SUBJ:0|V:0:1|NP:0:0|

*ABC: +, was all +/.

Main clause; **subject: *all I wanted***

%xcsy: MC:2|**SUBJ:3**|V:1:0|NP:1:3|

*ABC: +, I wanted.

Relative DC modifying the **subject**

%xcsy: DC-REL:2|SUBJ:1|V:1:0|NP:1:1|

[41] *ABC: To get out of here +/.

Complement DC: **subject**

%xcsy: UTT|AS|DC:5|SUBJ:0|V:0:1|NP:0:0|

*ABC: +, was all +/.

Main clause; **subject: *to get out of here***

%xcsy: MC:2|**SUBJ:5**|V:1:0|NP:1:3|

*ABC: +, I wanted.

Relative DC modifying the **predicative**

%xcsy: DC-REL:2|SUBJ:1|V:1:0|NP:1:1|

Because *All I wanted was to get out of here* sounds less marked and subjects come first in unmarked sentences, *all I wanted* was ultimately analyzed as the subject of the sentence.

- In English existential constructions with *there* (e.g., *there were/is/are...*), count the logical subject as SUBJ. The *there* is considered the adverbial of the clause:

[42] *ABC: There were some other Jews. Subject: *some other Jews*

%xcsy: UTT|AS|MC:5|**SUBJ:3**|V:1:0|NP:1:3|

V

Verbs

The subject tag is followed by the number verb forms in the clause. The tag is **V:x:y** with x = number of finite verb forms; y = number of non-finite verb forms. Both numbers need to be filled in.

- [43] *ABC: they were accused of +/.
 %xcsy: UTT|AS|MC:4|SUBJ:1|**V:1:0**| 1 fin. V: *accused*
 NP:1:1|PASS|
- *ABC: +, filling the teeth of the horses +/.
 %xcsy: DC:6|SUBJ:0|**V:0:1**| 1 non-fin. V: *filing*
 NP:1:5|
- *ABC: +, to make them appear younger .
 %xcsy: DC:5|SUBJ:0|**V:0:2**| 2 non-fin. V: *make, appear*
 NP:1:1|

- Retraced and reformulated verbs are not counted; repeated verbs are just counted once (cf. [32]):

- [44] *ABC: <I had> [//] I_mean@z:ep I was [/] was not sure.
 %xcsy: UTT|AS|MC:5|SUBJ:1|**V:1:0**| only 1 fin. V: *was*
 NP:1:1|

- Phrasal verbs / trennbare Verben like *look after* and *anrufen* are counted as one form even though they consist of two words: *he looked after the child* and *er rief mich an* both contain one finite verb form (cf. [13], [36], [39]).
- Nominalized verbs (gerunds in English) are counted as (non-finite) verb forms only if they occur without a determiner (this is the normal case; cf. [22] and [49] for counter-examples).
- Participles that are used like adjectives are not counted as (non-finite) verb forms (cf. [19]). Some examples:
the scared man ran away 1 finite verb: *ran*
the show was amazing 1 finite verb: *was*
we lived in a town called Esch@z:geo 1 finite verb: *lived*

Sometimes, it is less clear if something is a verb or not. Possible examples:

he looked scared
he was scared

Rely on your own feeling in these cases, but be consistent. If you tag a construction like *he was scared* as V:1:1, remember to also tag it as a passive (see p. 19).

Obvious English passives should be tagged as verb forms (and passives):

he was scared by the noise 1 finite verb: *was*, 1 non-finite verb: *scared*
the town was called Esch@z:geo 1 finite verb: *was*, 1 non-finite verb: *called*

Also forms of *to be born* should be tagged as verbs (and passives).

German Zustandspassiv participles should not be counted as verbs:

die Schule war geschlossen 1 finite verb: *war*

While Vorgangspassiv participles are verbs (and passives):

die Schule wurde geschlossen 1 finite verb: *wurde*, 1 non-finite verb: *geschlossen*

Noun phrases

NP

Each clause is tagged for the number and total length of the noun phrases it contains. The tag follows the V tag and is **NP:x:y** with x = number of noun phrases; y = number of words in all noun phrases.

Both numbers need to be filled in. The tag has to be closed with a |.

- All general rules for counting words apply. Unlike for the SUBJ tag, the article of the main noun in each NP is included in the count.

[45] *ABC: finally the guy left the others alone .

%xcsy: UTT|AS|MC:7|SUBJ:1|V:1:0|**NP:2:4**|

2 NPs with 4 words: *the guy, the others*. Note that SUBJ only has 1 word: *guy*

[46] *ABC: her name was Carla_Wenke@z:per .

%xcsy: UTT|AS|MC:4|SUBJ:2|V:1:0|**NP:2:3**|

2 NPs with 3 words: *her name, Carla_Wenke*

- Often, noun phrases don't contain a real noun but consist of or are headed by a pronoun, a determiner or a nominalized adjective.

[47] *ABC: he didn't like this at_all .

%xcsy: UTT|AS|MC:6|SUBJ:1|V:1:1|**NP:2:2**|

2 NPs with 2 words: *he, this* (pronouns)

[48] *ABC: much was taken from the Jewish .

%xcsy: UTT|AS|MC:6|SUBJ:1|V:1:1|**NP:2:3**|PASS|

2 NPs with 3 words: *much* (determiner), *the Jewish* (article and nominalized adjective)

- A nominalized verb (gerund in English) is only considered a head of a noun phrase if it is used with a determiner.

[49] *ABC: the suffering was terrible .

%xcsy: UTT|AS|MC:4|SUBJ:1|V:1:1|**NP:1:2**|

1 NP with 2 words: *the suffering* (article and gerund); cf. [22].

- Appositions and relative clauses are counted as part of the NP, together with the noun they modify.

[50] *ABC: our neighbor, rabbi Schwarz@z:per, died .

%xcsy: UTT|AS|MC:5|SUBJ:4|V:1:0|**NP:1:4**|

1 NP with 4 words: *our neighbor rabbi Schwarz*

[51] *ABC: the brother of my father +/.

%xcsy: UTT|AS|MC:9|SUBJ:8|V:1:0|**NP:3:13**|

*ABC: +, who already lived in New_York@z:geo +/.

%xcsy: DC-REL:5|SUBJ:1|V:1:0|**NP:2:2**|

*ABC: +, got us an affidavit .

MC: 3 NPs with 13 words: *the brother...lived in New_York, us, an affidavit*

DC: 2 NPs with 2 words: *who, New_York*

- However, complement DCs are not counted as noun phrases of the main clause. (This differs from how the subject length is counted: a complement DC that is a subject is counted as a subject while no complement DC is ever counted as a noun phrase; see p. 12)

[52] *ABC: I simply didn't know +/. Main clause
 %xcsy: UTT | AS | MC : 5 | SUBJ : 1 | V : 1 : 1 | NP : 1 : 1 |
 *ABC: +, what to do ! Complement DC: direct object to *know* in the MC
 %xcsy: DC : 3 | SUBJ : 0 | V : 0 : 1 | NP : 1 : 1 |
 MC: only 1 NP with 1 word: *I*
 DC: 1 NP with 1 word: *what*

- Coordinated noun phrases are counted as one long noun phrase. E.g., *me and my brother* is one NP with 4 words (cf. [36]).
- For counting noun phrases in cases where the subject NP is followed by a pronoun referring to the whole subject NP (cases like *all the kids, they were really happy*) cf. [37].
- Numbers used without a noun are not counted as NPs: *I was born in nineteen@z:num twenty@z:num* contains one NP with one word: *I* (cf. [61]). In contrast, numbers used as determiners or as post-modifiers (as in dates) are part of an NP:

five@z:num years NP:1:2 (cf. [34])
may twenty@z:num third@z:num nineteen@z:num thirty@z:num NP:1:3

- German has many verbs requiring a reflexive pronoun. Some examples are: *sich erinnern / anhören / weigern / melden / beschweren...*
 Don't count reflexive pronouns governed by these verbs as NPs.

[53] *XYZ: man konnte sich an niemanden wenden .
 %xcsy: UTT | AS | MC : 6 | SUBJ : 1 | V : 1 : 1 | NP : 2 : 2 |
 2 NPs with 2 words: *man, niemanden*. *Sich* is required by *wenden* and not counted as NP.

[54] *XYZ: ich kann mich gar nicht erinnern .
 %xcsy: UTT | AS | MC : 6 | SUBJ : 1 | V : 1 : 1 | NP : 1 : 1 |
 1 NP with 1 word: *ich*. *Mich* is required by *erinnern* and not counted as NP.
 (*Sich erinnern an* is a dialectal variant of *etwas erinnern* and requires a reflexive. It is distinct from the standard German *jemanden erinnern an* where *erinnern* governs a true accusative object, cf. [56])

Reflexive pronouns that are true objects of the verb (i.e. that are not required by the verb and could be exchanged for a non-reflexive pronoun) are counted as NPs, just as usual:

[55] *XYZ: wir haben uns in New_York@z:geo getroffen .
 %xcsy: UTT | AS | MC : 6 | SUBJ : 1 | V : 1 : 1 | NP : 3 : 3 |
 3 NPs with 3 words: *wir, uns, New_York*. *Uns* is a normal object of *treffen* (it could be exchanged for, e.g., *sie*).

[56] *XYZ: er erinnert mich an meine Verabredung .
 %xcsy: UTT | AS | MC : 6 | SUBJ : 1 | V : 1 : 0 | NP : 3 : 4 |
 3 NPs with 4 words: *er, mich, meine Verabredung*.

Reflexive pronouns not governed by a verb are counted as NPs, just as usual:

[57] *XYZ: er hat das Geschaeftliche unter sich gehabt +/.
 %xcsy: UTT|AS|MC:7|SUBJ:1|V:1:1|NP:3:4|
 3 NPs with 4 words: *er, das Geschaeftliche, sich*. *Sich* is governed by *unter*

- Noun phrases contained in other noun phrases are not counted twice.

E.g., the phrase *the brother of my father who already lived in New York* in [51] is a noun phrase containing four other noun phrases:

the brother of my father who already lived in New York
 _____ 1
 _____ 2
 _____ 3
 _____ 4

For the MC, the whole NP is counted once (10 words). The four NPs it contains (number 1-4) are not added to the count. Together with the other two NPs *us* and *an affidavit* this results in a total of NP:3:13 for the MC.

For the DC, only the two NPs occurring within the DC, number 3 and 4, are counted.

- Sometimes it is unclear whether something is one or more noun phrases. Just use common sense or see if the noun phrases can be rearranged.

[58] *ABC: It never bothered the boys at school .
 %xcsy: UTT|AS|MC:7|SUBJ:1|V:1:0|NP:2:5|

In [58], *the boys at school* is probably one NP because it refers to the boys who were at the school - the utterance does not mean *at school, it never bothered the boys*. In contrast, in *it never bothered the boys in summer, the boys* and *in summer* would be two distinct NPs. Cf. [31] where *farmhand in Ontario* has been considered as one NP although two NPs would be equally justifiable.

Embedded clauses

EMB

Sometimes, speakers interrupt their current clause, insert another clause, and then carry on with the 1st clause. In such a case, the speaker has to keep in mind the start of the 1st clause during the production of the inserted clause in order to pick up the 1st clause again. This is mentally taxing and therefore a sign of complexity.

Such inserted ("embedded") clauses are marked with **EMB**. This tag only occurs in the tag of a clause that is embedded. Otherwise, it is simply left off. The tag has to be closed with a |.

[59] *ABC: meine Mutter is(t) in Polen +/. 1st MC starts
 %xcsy: UTT|AS|MC:6|SUBJ:2|V:1:1|NP:2:3
 *ABC: +, Gorme@z:geo hiess die Stadt +/. 2nd MC is inserted: EMB
 %xcsy: AS|MC:4|SUBJ:1|V:1:0|NP:2:3|**EMB**|
 *ABC: +, geblieben. 1st MC continues

What qualifies as an embedded clause?

1: EMB clauses only occur in between two parts of another clause.

- A clause that is split by an EMB clause can be recognized by the fact that it is split over several main tier lines, and that the 2nd part of the clause therefore does not have a xcsy-tier. Cf. [6], repeated below as [60], for an example:

[60] *ABC: He is +/. MC: part 1
 %xcsy: ...
 *ABC: +, as far as I know +/. DC: embedded
 %xcsy: ...
 *ABC: +, a successful painter . MC: part 2

The DC in [60] is an example of an embedded clause. In contrast, the XC and DC in [61] are not embedded clauses because the first MC is not interrupted: the last main tier line contains a new MC and therefore also has a xcsy-tier of its own.

[61] *ABC: so you can imagine +/. MC 1
 %xcsy: UTT|**AS|MC:4|SUBJ:1|V:1:1|NP:1:1|**
 *ABC: +, in thirty@z:num four@z:num +/. XC: not embedded
 %xcsy: **AS|XC:2|SUBJ:0|V:0:0|NP:0:0|**
 *XYS: +, when it all came to an end +/. DC: not embedded
 %xcsy: **DC-REL:7|SUBJ:2|V:1:0|NP:2:4|**
 *ABC: +, that was really the disaster
 of her life (..) . MC 2
 %xcsy: **AS|MC:8|SUBJ:1|V:1:0|NP:2:6|**

- (While EMB clauses are DCs as well as MCs (see below), the other clause that is split by the EMB clause is mostly a MC).
- A clause that is interrupting another clause so that the interrupted clause is not continued (i.e. self-interruption +//.) is not an EMB clause. EMB clauses are always linked to the other clauses by +/. and +, .

2: EMB clauses can be left off without the split clause becoming ungrammatical.

- The split clause remains complete even when the embedded clause is removed.
(This means that complement DCs are never embedded.)

[62] *ABC: and she +/. MC: part 1
 %xcsy: UTT|AS|MC:10|SUBJ:1|V:1:0|NP:2:6|
 *ABC: +, I am happy to say +/. DC: embedded
 %xcsy: DC:5|SUBJ:1|V:1:1|NP:1:1|**EMB**|
 *ABC: +, walked away with the first MC: part 2
 and third prize .

In [62], the MC is still grammatical without the DC: *And she walked away with the first and third prize*. In [63] however, the split DC is not grammatical when MC 2 is left out: **...wie wir von unser Haus nach der gegangen sind*. MC2 is therefore not embedded:

[63] *ABC: ich erinnere(e) mich +/. MC1
 %xcsy: UTT|AS|MC:3|SUBJ:1|V:1:0|NP:3:11|
 ABC: +, ah@fp wie wir von unser []
 Haus nach [*] der +/. DC: part 1
 %xcsy: DC-REL:9|SUBJ:1|V:1:1|NP:3:4|
 *ABC: +, ich_glaube@z:ep es war die
 Karolinenstrasse@z:geo +/. MC 2: not embedded
 %xcsy: AS|MC:5|SUBJ:1|V:1:0|NP:2:3|
 *ABC: +, gegangen sind. DC: part 2

3: REL-DCs are never EMB.

- Relative clauses are never considered embedded clauses.
(Together with the above rule this means that besides MCs, only adverbial DCs may be embedded.)

[64] *ABC: for_instance my uncle +/. MC 1: part 1
 %xcsy: UTT|AS|MC:6|SUBJ:6|V:1:1|NP:1:6|PASS|
 *ABC: +, who had no children +/. DC-REL: not embedded
 %xcsy: **DC-REL**:4|SUBJ:1|V:1:0|NP:2:3|
 *ABC: +, was n(o)t interned (.) . MC 2: part 2

Is the EMB clause a MC or a DC?

The embedded clause is tagged as a DC

- when it cannot be used as an independent clause, i.e. when it cannot stand on its own (cf. *as far as I know* in [60])
- or when it clearly functions as an adverbial for the split clause (cf. *I am happy to say* in [62] which might stand on its own (however, the object of *say* would be missing) but which clearly is an adverbial for the MC, more specifically a disjunct/Satzadverbiale expressing the speakers attitude towards the MC).

The embedded clause is tagged as a MC when it is an independent clause. This means that it doesn't form part of the syntactic structure of the split clause and that it can be used as a clause of its own (see *Gorme hiess die Stadt* in [59]). Typical instances are word- or fact-searching questions: *We needed an – what was it called again? – affidavit where what was it called again? would be an embedded MC*

Passives

PASS

Using a passive construction in spoken speech can be considered advanced language use. Every clause containing a passive is therefore tagged with **PASS**. If there is no passive in a construction, the tag is simply left out.

The tag follows the NP tag or, if a clause is embedded, the EMB tag. It has to be closed with a |.

[65] *ABC: they were taken outside .
%xcsy: MC:4|SUBJ:1|V:1:1|NP:1:1|**PASS**|

English passives

- consist of a finite form of *be* or *get* (or a finite auxiliary verb + *be/get*) and a past participle:
was seen, is considered, will get promoted, was being shown, got admitted, can be said
- A *by*-phrase can usually be added to specify the agent:
they have been told (by the teacher)

German passives

- consist of a finite form of *werden* (or a finite auxiliary verb + *werden*) and a Partizip Perfekt:
wurde gesehen, war verboten worden. wird erzählt, konnte durchgeführt werden
- A *von*-phrase can usually be added to specify the agent:
es wurde ihnen (vom Beamten) zugesagt

If a verb form does not fulfill these criteria, it is not considered a passive form.

- German Zustandspassiv are formed with *sein* instead of *werden* and thus not considered passives: *war verboten, ist geöffnet, ist bewohnt gewesen*
- There needs to be a finite form of *be, get or werden* (or a finite auxiliary + *be/get/werden*). An infinite form is insufficient:

[66] *ABC: to get issued a visa +/.
%xcsy: UTT|AS|DC:5|SUBJ:0|**V:0:2**|NP:1:2|
*ABC: +, was really difficult .
%xcsy: MC:4|SUBJ:5|V:1:0|NP:0:0|

Sometimes, it can be hard to decide whether something is a passive or not (see p. 13). However, if forms such as *he was scared* are tagged as verb forms (V:1:1), they also need to be tagged as passives. Vice versa, participles not tagged as verb forms cannot be tagged as passive (therefore, German Zustandspassiv are never tagged as passives).

For examples of PASS clauses see [16], [33], [43], [48] and [64].

German Konjunktiv

KONJ

In German, using a subjunctive mood construction (Konjunktiv) in spoken speech can be considered advanced language use. Every German clause containing a Konjunktiv is therefore tagged with **KONJ**. English subjunctives are not tagged. If there is no Konjunktiv in a clause, the tag is simply left out.

The tag follows the NP tag, the EMB tag, or, if there is a passive construction, the PASS tag. It has to be closed with a |.

```
[67] *XYZ: wenn er in ein Konzentrationslager gekommen waere, +/.
%xcsy: UTT|AS|DC:7|SUBJ:1|V:1:0|NP:2:3|KONJ|
*XYZ: +, dann haetten wir ihn nie wiedergesehen .
%xcsy: MC:7|SUBJ:1|V:1:1|NP:2:2|KONJ|
```

- Failed attempts at using a Konjunktiv are not tagged with KONJ.

In [68], the Konjunktiv *würde* would be correct, but it is realized as *wurde*. The speaker can thus not be credited with a use of Konjunktiv:

```
[68] *XYZ: meine Freunde wurde [: wuerde] [* p:w] ich verlieren .
%xcsy: UTT|AS|MC:5|SUBJ:1|V:1:1|NP:2:3|
```

- Likewise, incorrect uses of a Konjunktiv are not tagged as KONJ.

The MC in [69] is a case where the Präteritum *konnte* is intended, but it is realized as the Konjunktiv *könnte*. Again, the speaker does not show mastery of the Konjunktiv and the clause is not tagged with KONJ.

```
[69] *XYZ: als ich nach Amerika kam +/.
%xcsy: UTT|AS|DC:5|SUBJ:1|V:1:0|NP:2:2|
*XYZ: +, koennte [*] ich nicht mal hello@s sagen .
%xcsy: MC:6|SUBJ:1|V:1:1|NP:1:1|
```

KV

German complex verb constructions

Just as passives and German Konjunktiv, certain combinations of finite and non-finite verbs in the same verb phrase can be rated as advanced language use. If such combinations occur, they are tagged as complex verbs, **KV**. The tag is only used for German.

If the KV tag occurs, it is the last tag in the tier. It is only used once, even if there are more than one complex verb constructions in the clause. It closes with a |.

What qualifies as a complex verb construction?

All instances where one finite verb form is used with two non-finite verb forms within the same verb phrase are complex verb constructions.

[70] *XYZ: vielleicht werden sie dich gehen lassen .

%xcsy: UTT|AS|MC:6|SUBJ:1|V:1:2|NP:2:2|**KV**|

Finite: *werden*; non-finite: *gehen lassen*

[71] *XYZ: er muss irgendwie Verbindungen gehabt haben .

%xcsy: UTT|AS|MC:6|SUBJ:1|V:1:2|NP:2:2|**KV**|

Finite: *muss*; non-finite: *gehabt haben*

- Often, these constructions are passive constructions.

[72] *XYZ: er ist nach zwei Wochen entlassen worden

%xcsy: UTT|MC:7|SUBJ:1|V:1:2|NP:2:3|**PASS**|**KV**|

Finite: *ist*; non-finite: *entlassen worden*

- Instances with 3 non-finite verb forms are also tagged as KV.

[73] *XYZ: sie hat ihn nicht gehen lassen wollen .

%xcsy: UTT|MC:7|SUBJ:1|V:1:3|NP:2:2|**KV**|

Finite: *hat*; non-finite: *gehen lassen wollen*

- The occurrence of other, additional verb forms has no effect on the KV tag.

[74] *XYZ: er blieb und wollte uns nachkommen lassen .

%xcsy: UTT|AS|MC:7|SUBJ:1|V:2:2|NP:2:2|**KV**|

Finite: *wollte*; non-finite: *nachkommen lassen*. *Blieb* does not affect the KV.

- If the second non-finite verb form in the clause belongs to a different verb phrase than the finite verb form, this is not considered a KV.

[75] *XYZ: wir konnten nur noch kriechen , nicht mehr laufen .

%xcsy: UTT|AS|MC:10|SUBJ:1|V:1:2|NP:2:2|

Finite: *konnten*; non-finite: *kriechen* | *laufen*, but they don't belong to the same verb phrase.

- Instances of *einkaufen/spazieren/schwimmen...gehen* and *zu tun haben* are an exception: they are never considered KV.

[76] *XYZ: da bin ich jeden Tag spazieren gegangen .

%xcsy: UTT|AS|MC:7|SUBJ:1|V:1:2|NP:2:3|

[77] *XYZ: die haben viel zu tun gehabt .

%xcsy: UTT|AS|MC:6|SUBJ:1|V:1:2|NP:2:2|

DC-RELS

DC-REL
A
B

Relative clauses in spoken English occur significantly less frequently than in written language (see e.g., Weinert, 2004). In a productive sentence combination task (Izumi, 2003, p.303) it was demonstrated that the accuracy for constructing subject relatives is significantly higher than for object relatives. Thus, we would like to capture the occurrence of relative clauses and make the distinction between subject and object relative clauses. Subject relative clauses are relative clauses in which the relative pronoun is the subject of the relative clause. They are tagged with A|. Object relative clauses, where the relative pronoun is the object in the relative clause, are tagged with B| at the very end of the %xcsy tier.

[78] *XYZ: +, and he worked in his uncle's firm +/.
 %xcsy: AS|MC:7|SUBJ:1|V:1:0|NP:2:12|
 *XYZ: +, which was the biggest leather (.) firm in North_Germany .
 %xcsy: DC-REL:8|SUBJ:1|V:1:0|NP:2:7|**A|**

- This is an example of an English subject WH relative clause. These type of relative clauses are very rare in all registers of English, but even more so in spoken English (see e.g., Weinert, 2004, p.10)

[79] *XYZ: mein Bruder Martin@z:per, +/.
 %xcsy: UTT|AS|MC:6|SUBJ:3|V:1:0|NP:1:9|
 *XYZ: +, der (.) damals bei der Epa@z:name arbeitete +/.
 %xcsy: DC-REL:6|SUBJ:1|V:1:0|NP:2:3|**A|**
 *XYZ: +, wurde dann arbeitslos .

- This is an example of a German subject "d"-relative clause (see König & Gast, 2009, p.198). These are more frequently used than the more bookish sounding German 'w-relatives, i.e. *welche/r/s*.

[80] *XYZ: and the other thing +/.
 %xcsy: UTT|AS|MC:7|SUBJ:2|V:1:1|NP:1:5|
 *XYZ: +, he did +/.
 %xcsy: DC-REL:2|SUBJ:1|V:1:0|NP:1:1|**B|**
 *XYZ: +, was to: bake.

- This is an example of an English object relative clause where the relative pronoun is in fact omitted. A study looking at the comprehension of relative clauses by Race & MacDonald (2003) shows that producers insert e.g., "that" to also facilitate production.

[81] *XYZ: sehr schwer +/.
 %xcsy: UTT|AS|XC:2|SUBJ:0|V:0:0|NP:0:0|
 *XYZ: +, als Frage zu beantworten +/.
 %xcsy: DC:4|SUBJ:0|V:0:1|NP:1:7|
 *XYZ: +, auf (.) die es keine Antwort gibt .
 %xcsy: DC-REL:6|SUBJ:1|V:1:0|NP:3:4|**B|**

- This is an example of a German object relative clause introduced by a clause-initial prepositional phrase which includes the 'd-relative'.

UTT | AS | MC : # | SUBJ : # | V : # : # | NP : # : # | EMB | PASS | KONJ | KV | DC-REL |