

Guide to Annotation of Hebrew Syntax for CHILDES

October 2011

Following is a dependency annotation scheme for Hebrew CHILDES utterances. The scheme is defined in terms of independent utterances (in other words, no inter-utterances relations are accounted for).

A dependency structure is a tree where relations are defined (and marked) between two words: a head and a dependent. Importantly, the current scheme is designed to handle spoken and not written language. As such, the issue of what is a Word needs to be carefully considered. For example, multi-lexemic expressions that are typically written as separate words are treated as a single item in our transcription and should be analyzed as such by the parser. Conversely, morphemes that are typically written as part of the orthographic word are treated as separate items (e.g., simplex prepositions). However, fused morphemes such as the definite preposition *ba-* ‘in-the’ may be treated as two items.

In our scheme we intend to treat fused morphemes of inflected prepositions (*'ba-*), prepositional possessives (*'shela-*) and ET marker (*'oto-*) as multiple tokens and split them accordingly. We also intend to do so in cases of possessive suffixes (*'axoti*). The reason for this decision is we should to avoid sparseness of data as we would like to consider all these inflected instances of a certain preposition as a single entity. We also recall that the transliteration is manual due to the fact that the data is of spoken language, and thus we allow ourselves to change and modify it to our needs.

However, we do not intend to split inflected verbs (*'axalti*) to two tokens where one represents the subject and the other the base form of the verb, as this split will create problems considering cases where the verb is inflected but the subject does explicitly appear. We wish to avoid inconsistency and thus we do not intend to follow up on such a split in this case.

Although dependencies are marked between two words, and not between phrasal constituents, the hierarchical nature of syntactic structure is reflected by two levels of relations: relations existing between the head (or root) of the clause and its constituents and the relations between the heads of these constituents and their dependents. The goal of the definitions is to provide a framework that allows for a consistent treatment of dependencies, on the one hand, and for differentiation between levels, on the other – as specified below. The goal is not to provide a correct theoretical framework but rather an efficient and informative one.

Dependency Relations

As noted, a dependency structure must be a tree, i.e., it must have a single root

[marked with a special relation named Root]. Hence, the following scheme assumes that in every utterance, one word is this root. The root depends on no other word (in the same utterance).

The tree is a directed tree which is comprised of dependency relations between couples of tokens, where each token represents a node in the tree and the edges between a node and its children nodes represent the relations between the tokens. In such a relation between a couple of tokens, the parent node is referred to as the head and the children nodes are referred to as the dependents. Each token may have a single head (i.e. one parent node) and several dependents (i.e. children nodes). The root of the tree is a special token called ROOT that is added to the sentence, on which the main part of the sentence (the one with no head within the sentence itself) depends on.

Typically, the root is a verb, the copula in copula constructions, or the predicating element in a copula construction with zero copula (see below). As such, it typically carries the tense marking in the clause. When an utterance is lacking any of the above, the root is the element on which other elements depend (such as the noun with respect to its modifiers). Single word utterances contain only the root.

(1) laševet
1|0|Root

Taxonomy of Dependents

We distinguish between two types of dependents, arguments [A] and modifiers [M]. Arguments are subcategorized dependents of the heads that they modify: typically, they are semantically required by the head, their properties are determined by the head, and they can occur at most once (often, exactly once). Modifiers, on the other hand, are non-subcategorized dependents: typically, they select the head that they depend on (in the sense that they specify the properties of the head they depend on rather than vice versa), and, consequently, they may occur zero or more times.

It is also important, however, to remember that we are dealing with spoken language – and with language spoken by children who are still acquiring their grammar. The study of spoken language entails dealing with situations where arguments and heads may be missing from the actual utterance as part of what we could see as processing constraints. We should decide whether we treat these missing elements as “dropped” (and, consequently, to mark them in the analysis) or not. In the GRASP scheme for English, the decision was to mark the missing element as elided and to relate to it in the analysis. Following the scheme for English, we intend to maintain a single scheme suited for child language and adult language, and thus referring these dropped elements as elided.

Examples Format

The majority of the examples given in the following sections are taken directly from the Hebrew CHILDES corpora. The examples are presented in the following format:

<Actual utterance (main tier)>
<Morphological analysis (mor tier)>
<Syntactic analysis (XGRA tier)>
<English gloss>

For example:

ʔat	mešaḳēret	!
% mor: pro:person gen:fem&num:sg	part gen:fem&num:sg	!
1 2 Aagr	2 0 Root	3 2 Punct
<i>you</i>	<i>lie</i>	<i>!</i>

If the actual utterance is split into more than one line, then the other tiers in the examples are also split and each line of analysis follows its corresponding part of utterance.

<Actual utterance – first part (main tier)>
<Morphological analysis (mor tier)>
<Syntactic analysis (XGRA tier)>
<English gloss>

<Actual utterance – second part (main tier)>
<Morphological analysis (mor tier)>
<Syntactic analysis (XGRA tier)>
<English gloss>

...

<Actual utterance –last part (main tier)>
<Morphological analysis (mor tier)>
<Syntactic analysis (XGRA tier)>
<English gloss>

The **actual utterance** is presented in the same transliteration as in the CHILDES corpora.

The **morphological analysis** presents pieces of information taken from the mor tier of the CHILDES corpora (the actual mor tier contains more data which may be less relevant to this guide). Each token is presented by its part of speech (POS) tag in the order of the appearance in the utterance itself. Some tokens are presented with more morphological data, separated from the POS tag with a '|' sign. Each piece of morphological information is in the form of <name:value>, the 'name' being the type

of information (e.g. 'gen' for gender, 'num' for number, 'root' for root) and the 'value' being the morphological analysis of this type for this token (e.g. 'fem' for feminine gender, 'sg' for singular number, 'twb' for the root). Pairs of <name:value> are separated between one another with an '&' sign.

In certain cases where a token was split into two morphemes in post-processing of the corpora, there may be one morphological analysis combined for the morphemes. This occurs most commonly in the split of inflected prepositions and possessive markers (e.g. 'ba-' is separated to 'be-' and 'ha-').

The **syntactic analysis** presents a suggested syntactic analysis for the utterance using dependency relations. Each token is represented with a triplet where the first element is the index of this token in the sentence (starting the count from 1), the second element is the index of the head of this token (the special ROOT marker is given the index 0) and the type of the relation. In some sections more than one syntactic analysis is appropriate so all possible analyses are presented (see further explanation below regarding copula constructions and accusative AT constructions).

The **English gloss** presents an English gloss for the utterance.

Types of arguments:

AgreementArgument [Aagr] = Identifies the dependent argument with which the predicate agrees. This argument cannot be a clause in itself. Typically, it is a nominal element (noun or pronoun).

(siv505b.mor, line 140)

(3)	ʔat	mešaḳēret	!
%mor:	pro:person gen:fem&num:sg	part gen:fem&num:sg	!
	1 2 Aagr	2 0 Root	3 2 Punct
	<i>you</i>	<i>lie</i>	<i>!</i>

NonFiniteArgument [Ainf] = Identifies the non-finite verbal argument of another clause or another verb or of a copula construction. It is dependent on the predicate of the matrix clause.

(siv505b.mor, line 36)

(4)	ʔaz	titēn	le-	ʔanī
%mor:	adv	ti#v	prep pers:1SG--i	
	1 2 Mcom	2 0 Root	3 2 Anonarg2	4 3 Aprep
	<i>so/then</i>	<i>give</i>	<i>to</i>	<i>I</i>
	laṣavōr	.		
	la#v	.		
	5 2 Ainf	6 2 Punct		

pass .

Non-agreementArgument [Anonagr] = Identifies any argument of a verb with which agreement does not hold. Typically, it is a nominal that is dependent on the verb. The dependent must be a non-clausal argument of the predicate.

(1eo300d.mor, line 477)

(5a) ?anī lo? rocē tipōt .
%mor: pro:person|num:sg neg
part|num:sg n
1|3|Agr 2|3|Mneg 3|0|Root 4|3|Anonagr 5|3|Punct
I no want drops .

(siv611a.mor, line 431)

(5b) hīne hu? macā? maqōm .
%mor: co pro:person|gen:masc&num:sg
v|gen:masc&num:sg
n
1|3|Mcom 2|3|Agr 3|0|Root 4|3|Anonagr 5|3|Punct
here he find place .

(siv611a.mor, line 440)

(5c) ken, hu? ?ohēv meqomōt ?im
%mor: co pro:person|gen:masc&num:sg
part|gen:masc&num:sg n
prep
1|3|Mcom 2|3|Agr 3|0|Root 4|3|Anonagr 5|4|Mpre
yes he love places with

liqlūk .
n .
6|5|Aprep 7|3|Punct
dirt .

The tag for non-agreeing arguments applies for what are typically termed indirect or oblique arguments. In these constructions, the nominal element is preceded by a preposition. The Anonarg dependency is marked on the prepositional element and the nominal element is marked as the argument of a preposition, Aprep:

(siv505b.mor, line 1961)

(5d) ?oy, pagāʕti be- ?at ?
%mor: co v prep|pers:2femSG--aḵ ?
1|2|Mcom 2|0|Root 3|2|Anonagr 4|3|Aprep 5|2|Punct
oh_no hurt in/at you ?

Note, that when prepositions are definite, some become phonologically merged with the article (be- ‘in’, le- ‘to’, ke- ‘like’, but not me- ‘from’ or other basic prepositions such as ‘al ‘on’). In the %MOR line, these elements are annotated as PREP&DET (indicating a portmanteau morpheme) and as stated earlier will be subsequently annotated separately on the %XGRA tier.

This separation will allow consistency with prepositions before an indefinite noun and avoids sparseness since 'ba-' and 'be-' are in essence the same prepositional argument or modifier to a verb or noun.

(leo300d.mor, line 465)

(5e)	yašānti	be-	ha-	bāyit
%mor: v		prepdet		n
	1 0 Root	2 1 Mpre	3 4Mdet	4 2 Aprep
	<i>sleep</i>	<i>in</i>	<i>the</i>	<i>house/home</i>

šel	ʔanī	.
prep pers:1SG-ī		.
5 4 Mposs	6 5 Aprep	7 1 Punct
<i>of</i>	<i>I</i>	.

The same dependency relations are noted for cases where the non-agreeing argument is preceded by the accusative marker (typically termed the direct object). There are two options to annotate these sentences:

- 1) The accusative marker is marked as argument of the verb, and the nominal element is the argument of the preposition Aprep:

(leo300d.mor, line 486)

(5f)	loʔ	rocē	ʔet	ha-	ṭipōt	!
%mor: neg		part	acc	det	n	!
	1 2 Mneg	2 0 Root	3 2 Anonagr	4 5 Mdet	5 3 Aprep	6 2 Punct
	<i>no</i>	<i>want</i>	<i>AT</i>	<i>the</i>	<i>drops</i>	<i>!</i>

Or

- 2) The nominal element is the argument of the preposition Aprep, and the accusative marker is its dependent:

	loʔ	rocē	ʔet	ha-
%mor: neg		part	acc	det
	1 2 Mneg	2 0 Root	3 2 Anonagr	4 5 Mdet
	<i>no</i>	<i>want</i>	<i>AT</i>	<i>the</i>
	ṭipōt	!		
	<i>n</i>	<i>!</i>		

5|3|Aprep 6|2|Punct
drops !

In instances where more than one non-agreeing argument occurs in the construction (typically termed a ditransitive construction), the dependent is marked as 2:

(leo300d.mor, line 24)

(6) hi? natnā le- ?anī māšehu
 %mor: pro:person|gen:fem&num:sg v|gen:fem&num:sg
 prep|pers:1SG—i pro:indef
 1|2|Aagr 2|0|Root 3|2|Anonagr2 4|3|Aprep 5|2|Anonagr
she *give* *to* *I* *something*

be- ha- ?ōzen .
prepdet *n* .
 6|2|Mpre 7|8|Mdet 8|6|Aprep 9|2|Punct
in *the* *ear* .

Non-agreement arguments can also occur as finite clausal dependents on the root. As in the case for nominal arguments, the whole construction depends on the main verb of the matrix clause. The relation Anonagr is marked between the complementizer introducing the clause and the predicate of the main clause, on the one hand, and the relation RootSub is marked between the predicate of the clausal argument and the complementizer, on the other:

(leo300d.mor, line 160)

(7a) ?atā rocē še- ?anī ?esarēq
 %mor: pro:person|gen:masc&num:sg part|gen:masc&num:sg
 conj:subor pro:person|num:sg v|num:sg
 1|2|Aagr 2|0|Root 3|2|Anonagr 4|5|Aagr 5|3|RootSub
you *want* *that* *I* *comb*

?et ?atā ?
 acc|pers:2mascSG-ḵā ?
 6|5|Anonagr 7|6|Aprep 8|2|Punct
AT *you* ?

In the case of direct speech complementation –

(hag204c.mor, line 126)

(7b) bo? Benc, qarā? ?āriq,
 %mor:v n:prop v n:prop
 1|6|Mcom 2|6|Mvoc 3|0|Root 4|3|Aagr
come *Benc* *read* *Ariq*

Navxān n:prop 5 6 Aagr <i>Navxan</i>	loqēax part 6 3 Asub <i>take</i>	ʔet acc 7 6 Anonagr <i>AT</i>	Šmanmān n:prop 8 7 Aprep <i>Shmanman</i>
le prep 9 6 Mpre <i>to</i>	ṭiyūl n 10 9 Aprep <i>trip</i>	.	.

ArgumentOfCopula [Acop] = Identifies the argument in a copula construction (nominal and adjectival), or the arguments of verbs such as nehefax and nihiya (‘become’). There are typically two options in annotating such utterances: either the copula is the head and the nominal or adjectival predicate is its dependent, or vice versa. A more elaborate discussion with regards to the syntactic tagging of copular utterances is given below.

(leo300d.mor, line 210)

(8a) ken,	ʔavāl	hala@c [:hayā] [*]	le-	ʔanī
% mor: co	conj:coord	v gen:masc&num:sg		
prep pers:1SG--i				
1 3 Mcom	2 3 Mcom	3 0 Root	4 3 Anonagr	5 4 Aprep
<i>yes</i>	<i>but</i>	<i>be/exist</i>	<i>to</i>	<i>I</i>
qcat	qar	.		
qn	adj gen:masc&num:sg	.		
6 7 Mquant	7 3 Acop	8 3 Punct		
<i>little</i>	<i>cold</i>	.		

Or

ken,	ʔavāl	hala@c [:hayā] [*]	le-	ʔanī
% mor: co	conj:coord	v gen:masc&num:sg		
prep pers:1SG--i				
1 3 Mcom	2 3 Mcom	3 7 Acop	4 3 Anonagr	5 4 Aprep
<i>yes</i>	<i>but</i>	<i>be/exist</i>	<i>to</i>	<i>I</i>
qcat	qar	.		
qn	adj gen:masc&num:sg	.		
6 7 Mquant	7 0 Root	8 3 Punct		
<i>little</i>	<i>cold</i>	.		

In cases where the copula is zero, the nominal predicate in any case becomes the root of the utterances.

(leo300d.mor, line 336)

(8b) ʔanī ʕayēf .
%mor:pro:person|num:sg adj|num:sg .
1|2|Agr 2|0|Root 3|2|Punct
I tired .

Finite clausal dependents on the root can also occur in copula constructions. As in the case of arguments dependent on a verb, the second root is marked as a root that is an argument:

(8c) ha- be'aya hi še- `ani ʕacuva
%mor:det n pro:person
conj:subor pro:person|num:sg adj|num:sg
1|2|Mdet 2|3|Agr 3|0|Root 4|3|Acop 5|6|Agr 6|4|RootSub
the problem she that I sad

Types of modifiers:

Mdet – specifies a relation between a determiner and a noun. Most commonly the determiner is specified by 'ha-'. The noun is the head and the determiner is the dependent.

(leo300d.mor, line 53)

(10a) ha- ʔōzen .
%mor: det n .
1|2|Mdet 2|0|Root 3|2|Punct
the ear .

Madj – specifies a relation between an adjective and a noun. The noun is the head and the adjective is the dependent.

(leo300d.mor, line 112)

(10b) be- masrēq ʔaxēr tistarqī .
%mor: prep n|gen:masc&num:sg
adj|gen:masc&num:sg ti#v .
1|4|Mpre 2|1|Aprep 3|2|Madj 4|0|Root 5|4|Punct
in comb different comb_onself .

Mnoun – specifies a relation between a noun modifier and a noun in a construct state. The noun is the head and the noun modifier is the dependent.

(hag204c.mor, line 297)

(10c) holqīm, ʔāriq, Benc, Navxān we-
%mor: part|num:pl n:prop n:prop n:prop conj
1|0|Root 2|5|Coord 3|5|Coord 4|5|Coord 5|1|Agr

<i>walk</i>	<i>Ariq</i>	<i>Benc</i>	<i>Navxan</i>	<i>and</i>
Šmanmān, n:prop	lanūax la#v	be- prep	cel n	ha- det
6 5 Coord	7 1 Ainf	8 7 Mpre	9 8 Aprep	10 11 Mdet
<i>Shmanman</i>	<i>rest</i>	<i>in/at</i>	<i>shadow</i>	<i>the</i>
Şec n
11 9 Mnoun	12 1 Punct			
<i>tree</i>	.			

Madv – specifies a relation between an adverbial modifier and a verb or noun. The verb or noun is the head and the adverb is the dependent.

(hag204c.mor, line 287)

(10d)	ma	Şoşīm	Şakšāyw	?
%mor:que		part	adv	?
	1 2 Anonagr	2 0 Root	3 2 Madv	4 2 Punct
	<i>what</i>	<i>do</i>	<i>now</i>	?

Mvoc – specifies a vocative. We relate a vocative to the entire sentence, so the main predicate (the root) of the sentence is the head of the vocative and the vocative is the dependent. This follows the definition of this relation in the English scheme.

(siv505b.mor, line 16)

(10e)	Asaf	tedabēr	.
%mor:n:prop		te#v	.
	1 2 Mvoc	2 0 Root	3 2 Punct
	<i>Asaf</i>	<i>speak</i>	.

Mcom – specifies a communicator. Communicators include phrases such as 'ʔavāl', 'ʔaz', 'kāḡa', 'ken' and others. Similar to Mvoc, the main predicate (the root) of the sentence is the head of the relation and the communicator is the dependent.

Mposs – specifies a possessive marker, most commonly 'šel'. The noun is the head and the possessive marker is the dependent. Note that as stated above a separation is made between the possessive marker and its inflected suffix into two tokens, and the relation between the possessive marker 'šel' and its inflection is marked with a 'Aprep' relation.

(siv505b.mor, line 55)

(10f)	ʔavāl	ze	ha-	cad
%mor:conj:coord		pro:dem gen:masc&num:sg		

det	n gen:masc&num:sg		
1 4 Mcom	2 4 Agr	3 4 Mdet	4 0 Root
<i>but</i>	<i>this</i>	<i>the</i>	<i>side</i>

šel	ʔanī	.
prep pers:1SG-ī		.
5 4 Mposs	6 5 Aprep	7 4 Punct
<i>of</i>	<i>I</i>	.

Mneg – specifies a negation of a verb or a noun. Negation phrases include 'lo?', 'ʔal' and others. The noun or the verb is the head and the negation phrase is the dependent.

(siv505b.mor, line 50)

(10g) Asaf,	ʔal	targīz	ʔet
%mor:n:prop	neg	ta#v	acc
1 3 Mvoc	2 3 Mneg	3 0 Root	4 3 Anonagr
<i>Asaf</i>	<i>don't</i>	<i>annoy</i>	<i>AT</i>

Siwān	.
n:prop	.
5 4 Aprep	6 3 Punct
<i>Siwān</i>	.

Mquant – specifies a relation between most commonly a noun and a nominal quantifier. The noun is the head and the nominal quantifier is the dependent.

(leo300d.mor, line 210)

(10h) ken,	ʔavāl	hala@c [: hayā] [*]	le-	ʔanī
%mor:co	conj:coord	v gen:masc&num:sg		
prep pers:1SG--i				
1 3 Mcom	2 3 Mcom	3 0 Root	4 3 Anonagr	5 4 Aprep
<i>yes</i>	<i>but</i>	<i>be/exist</i>	<i>to</i>	

qcat	qar	.
qn	adj gen:masc&num:sg	.
6 7 Mquant	7 3 Acop	8 3 Punct
<i>little</i>	<i>cold</i>	.

Msub – specifies a relative clause. A relative clause can also serve as a nominal modifier; the noun is the head and the relativizer is the dependent. Note that the verb in the second clause is marked as the Root of the modifier.

(leo300d.mor, line 561)

(10i) balōn	še-	hitpocēc	.
%mor: n gen:masc&num:sg	conj:subor		

v|gen:masc&num:sg .
 1|0|Root 2|1|Msub 3|2|RootSub 4|1|Punct
balloon that explode .

An adverbial clause can also serve as a verbal modifier:

(leo300d.mor, line 231)

(10j)	šāmta	māyim	xamīm	ʕal	ha-	guf
%mor: v	prep	det	n	adj gen:masc&num:pl		
	1 0 Root	2 1 Anonagr	3 2 Madj	4 1 Mpre	5 6 Mdet	6 4 Aprep
	<i>put</i>	<i>water</i>	<i>hot</i>	<i>on</i>	<i>the</i>	<i>body</i>
	še-	loʔ	yihiyē			
	conj:subor	neg	yi#v gen:masc&num:sg			
	7 1 Msub	8 9 Mneg	9 7 RootSub			
	<i>that</i>	<i>no</i>	<i>be/exist</i>			
	le-	ʔatā	qar	.		
	prep pers:2mascSG--eḵā	adj gen:masc&num:sg	.			
	10 9 Anonagr	11 10 Aprep	12 9 Acop	13 1 Punct		
	<i>to</i>	<i>you</i>	<i>cold</i>	.		

Coordination [Coord] = specifies a coordination relation between coordinated items and coordinators, most commonly 'we-'. Note that normally the coordinators are attached to the following word orthographically, but in our transliteration they are separated and considered a separate token.

There are various ways of dealing with coordination and coordinated elements. We intend to follow the solution proposed in the English CHILDES scheme as described below (see also a more elaborated discussion on coordinated elements and the options to deal with them below):

The English scheme expresses the approach where the head of the coordination is the coordinator, and the dependents are the coordinated items where the relation between the coordinator and the coordinated items is marked with the name COORD.

In case there are two or more elements of coordination with multiple coordinators, the coordinators of the coordinated elements are linked from left to right (the right-most coordinator is the head of the others) between one another in a COORD relation.

In the absence of the coordinator the right-most coordinated item is the head of the relation.

(siv505b.mor, line 212)

(11a)	yeš	le-	ʔanī	zikaṛōn	ṭov
%mor: co		prep pers:1SG—i		n gen:masc&num:sg	

adj gen:masc&num:sg				
1 0 Root	2 1 Anonagr		3 2 Aprep	4 1 Aagr
<i>be/exist</i>	<i>to</i>		<i>I</i>	<i>memory</i>
				5 6 Coord
				<i>good</i>

ʔavāl	loʔ	ʔov	meʔōd	.
conj:coord	neg	adj gen:masc&num:sg	adv	.
6 4 Madj	7 8 Mneg	8 6 Coord	9 8 Madj	10 1 Punct
<i>but</i>	<i>no</i>	<i>good</i>	<i>very</i>	.

(lio300a.mor, line 159)

(11b) tagīdi ze ze hayā kše- hayīt

%mor: ta#v	pro:dem gen:masc&num:sg			v gen:masc&num:sg	
	pro:dem gen:masc&num:sg			v gen:masc&num:sg	
	conj:subor			v gen:fem&num:sg	
1 8 Mcom	2 4 Aagr	3 4 Aagr	4 8 Coord	5 4 Msub	6 5 RootSub
<i>say</i>	<i>it/this</i>	<i>it/this</i>	<i>be/exist</i>	<i>when</i>	<i>be/exist</i>

qʔanā	we-	ʕakšāyw	ʔat
adj gen:fem&num:sg	conj		
adv	pro:person gen:fem&num:sg		
7 6 Acop	8 0 Root	9 11 Madv	10 11 Aagr
<i>small</i>	<i>and</i>	<i>now</i>	<i>you</i>

gdolā	?
adj gen:fem&num:sg	?
11 8 Coord	12 8 Punct
<i>big</i>	?

(leo300d.mor, line 68)

(11c) le- ʔat kaʔāv ha- garōn

%mor: prep pers:2femSG—aḵ		v gen:masc&num:sg		
det	n gen:masc&num:sg			
1 3 Anonagr	2 1 Aprep	3 6 Coord	4 5 Mdet	5 3 Aagr
<i>to</i>	<i>you</i>	<i>hurt</i>	<i>the</i>	<i>throat</i>

we-	le-	ʔanī	kaʔāv	ha-
conj	prep pers:1SG—i		v gen:masc&num:sg	
det				
6 0 Root	7 9 Anonagr	8 7 Aprep	9 6 Coord	10 11 Mdet
<i>and</i>	<i>to</i>	<i>I</i>	<i>hurt</i>	<i>the</i>

ʔōzen	.
n gen:fem&num:sg	.
11 9 Agr	12 6 Punct
<i>ear</i>	.

(hag204c.mor, line 175)

(11d)	huʔ	rac	we-	hitxabēʔ	meʔaxorēy
%mor:	pro:person gen:masc&num:sg			v gen:masc&num:sg	
	conj	v gen:masc&num:sg			
	prep				
	1 3 Aagr	2 3 Coord	3 0 Root	4 3 Coord	5 4 Mpre
	<i>he</i>	<i>run</i>	<i>and</i>	<i>hide</i>	<i>behind</i>
	ha-	ʕec	.		
	det	n	.		
	6 7 Mdet	7 5 Aprep	8 3 Punct		
	<i>the</i>	<i>tree</i>	.		

Interesting or problematic sentences:

(siv505b.mor, line 217)

(11e)	ʔābaʔ	teḳabē	huʔ	loʔ	rocē
%mor:	n gen:masc&num:sg	te#v gen:masc&num:sg			
	pro:person gen:masc&num:sg	neg	part gen:masc&num:sg		
	1 2 Mvoc	2 0 Root	3 5 Aagr	4 5 Mneg	5 2 Coord
	<i>father</i>	<i>turn_off</i>	<i>he</i>	<i>no</i>	<i>want</i>
	leṣaxēq	yafē	.		
	le#v	adv	.		
	6 5 Ainf	7 6 Madv	8 2 Punct		
	<i>play</i>	<i>okay</i>	.		

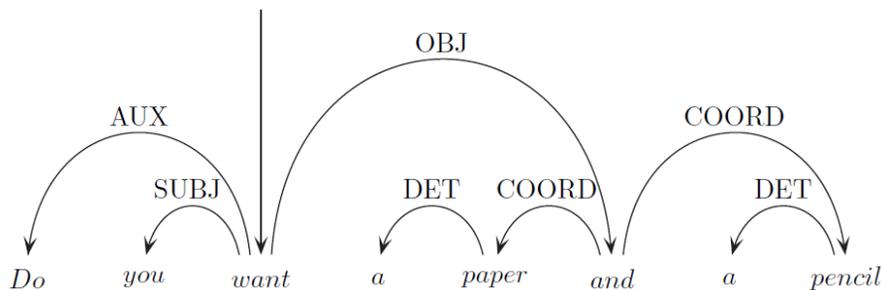
(lio300a.mor, line 355)

(11f)	loʔ	ʔat	loʔ	makirā	loʔ
%mor:	neg	pro:person gen:fem&num:sg			
	neg	part gen:fem&num:sg			
	neg				
	1 4 Mcom	2 4Aagr	3 4 Mneg	4 0 Root	5 6 Mneg
	<i>no</i>	<i>you</i>	<i>no</i>	<i>recognize</i>	<i>no</i>
	ʔet	ʔarye	we-	loʔ	ʔet
	acc	n:prop	conj	neg	acc
	6 8 Coord	7 6 Aprep	8 4 Anonagr	9 10 Mneg	10 8 Coord
	<i>AT</i>	<i>Arye</i>	<i>and</i>	<i>no</i>	<i>AT</i>
	ʔelijāhu	.			
	n:prop	.			
	11 10 Aprep	12 4 Punct			
	<i>Eliyahū</i>	.			

Possible solutions for the Coord relation

- 1) Original **English** CHILDES scheme: The coordinator is the head and the coordinated items are the dependents. In case there are multiple coordinators they are connected between one another from left to right. In case where the multiple coordinated items and one coordinator, the single coordinator is the head of all of the coordinated items. In case there is no coordinator, the right most coordinated item is selected as the head.

This way of representation allows the coordinator to be the head of a structure that takes all the dependents and the head outside of the structure that normally each of the coordinated items would have taken. For example:



In this sentence, the coordinator 'and' is the head of 'paper' and 'pencil' in the coordinated structure. It is also the dependent of 'want' since 'want' selects 'paper' and 'pencil' as its objects in this sentence.

KENJI SAGAE, ERIC DAVIS, ALON LAVIE, BRIAN MACWHINNEY and SHULY WINTNER (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language*, 37, pp 705-729 doi:10.1017/S0305000909990407

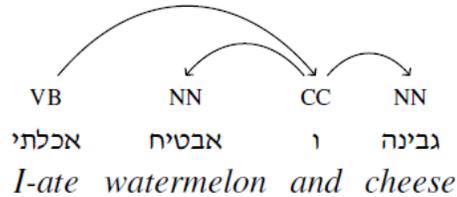
- 2) Stanford Parser **English** scheme: The first coordinated element is the head of the other coordinated elements in the Conj relation, and also the head of the coordinator in the CC relation. For example:

"Bill is big and honest"	<i>conj</i> (big, honest)
"Bill is big and honest"	<i>cc</i> (big, and)

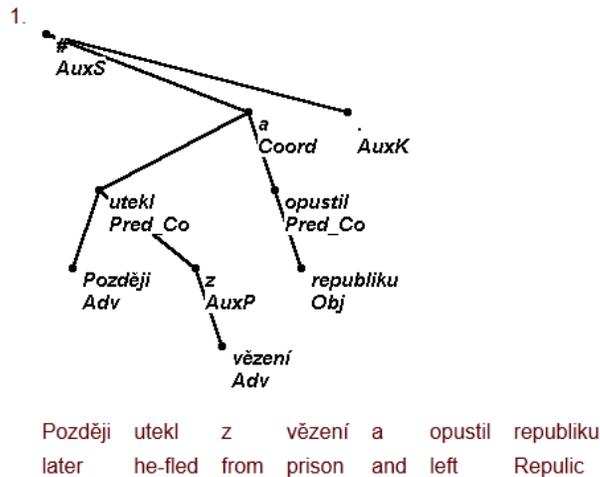
Relations are presented in the format type (head, dependent).

De Marneffe, M.-C., & Manning, C. D. (2010). Stanford typed dependencies manual. 20090110 <http://nlp.stanford.edu/software/dependencies.pdf>, (September 2008), 1-22. Citeseer. Retrieved from http://nlp.stanford.edu/downloads/dependencies_manual.pdf

- 3) Yoav Goldberg's **Hebrew** scheme in his thesis: The coordinator is the head and the coordinated elements are the dependents. In case there is more than one coordinating element, the last one is selected to be the head of the others. For example:



- 4) Prague Dependency Treebank Czech scheme: The coordinator is marked with a Coord relation and the coordinated elements are marked with their regular function with a suffix *_Co* appended to it. For example:



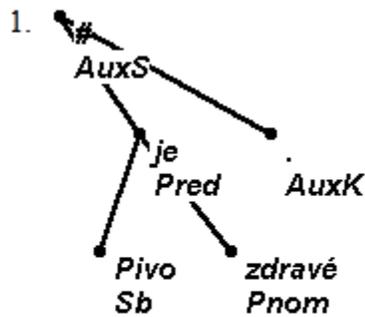
<http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/ch03s04.html#s2-coord>

Jan Hajič, et al.
 2006
Prague Dependency Treebank 2.0
 Linguistic Data Consortium, Philadelphia

Possible solutions for copula constructions and the Acop relation

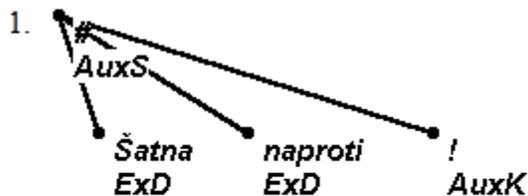
- 1) In **Czech** nominal sentences: In the Prague Czech Dependency Treebank the copula refers also to the past and the future (as well as the present) and is

marked with the Pred function and it is the head of the other parts in the sentence. For example:



Pivo je zdravé
 beer is healthy

When there is no copula and it is elided, then usually the parts who were dependent on the copula are now dependent on the token that the copula used to depend on with the function ExD whose role is to mark the ellipsis. For example:



šatna naproti!
 cloak-room opposite

http://ufal.mff.cuni.cz/pdt/Corpora/PDT_1.0/Doc/aman-en/ch03s02.html#prsljmst

Jan Hajič, et al.

2006

Prague Dependency Treebank 2.0

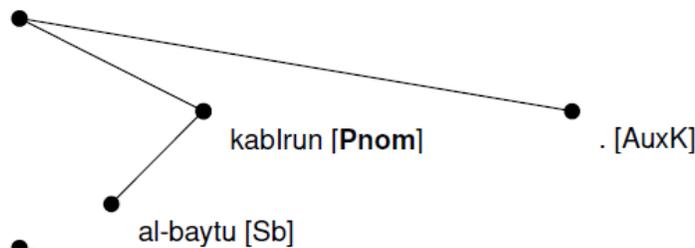
Linguistic Data Consortium, Philadelphia

- 2) In **Hebrew** nominal sentences: In Yoav Goldberg's thesis the past and future forms are described as an auxiliary (e.g. the word *hayita* in the sentence *hamenora hayita semel mashmauty*) and they are the head of these sentences.

In a nominal sentence in the present form without a copula the nominal predicate is the head (e.g. the word *xakam* in the sentence *hayed xakam*).

In a nominal sentence in the present form with a copula the nominal predicate is still the head and the copula is dependent on it (e.g. the word *הוא hu* is dependent on the word *חכם xakam* in the sentence *חכם הוא היילד hayeled hu xakam*).

- 3) In **Arabic** nominal sentences: In the Prague Arabic Dependency Treebank the sentences that don't have a copula are analyzed in a way the nominal predicate is the head and it is marked with the function Pnom. For example:



http://ufal.mff.cuni.cz/padt/PADT_1.0/docs/guides/PADT_Analytical.pdf

Jan Hajic, et al.
2004
Prague Arabic Dependency Treebank 1.0
Linguistic Data Consortium, Philadelphia

- 4) In **English** nominal sentences: The original scheme of CHILDES in English refers to the copula as the head and the nominal predicate as the dependent. For example:

Examples:

Mary is a student.

PRED(student-4, is-2)

KENJI SAGAE, ERIC DAVIS, ALON LAVIE, BRIAN MACWHINNEY and SHULY WINTNER (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language*, 37, pp 705-729
doi:10.1017/S0305000909990407

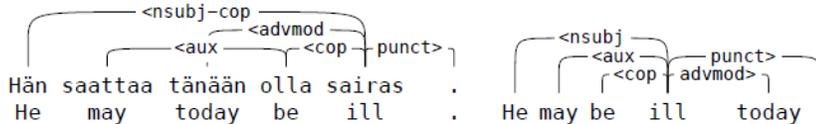
- 5) In **English** nominal sentences: The Stanford Parser scheme refers to the nominal predicate as the head and the copula as the dependent. This move was made from motives coming from the desire to create a scheme which is adaptable to other languages in which there is not necessarily a representation for a copula (for example Chinese). For example:

“Bill is big”

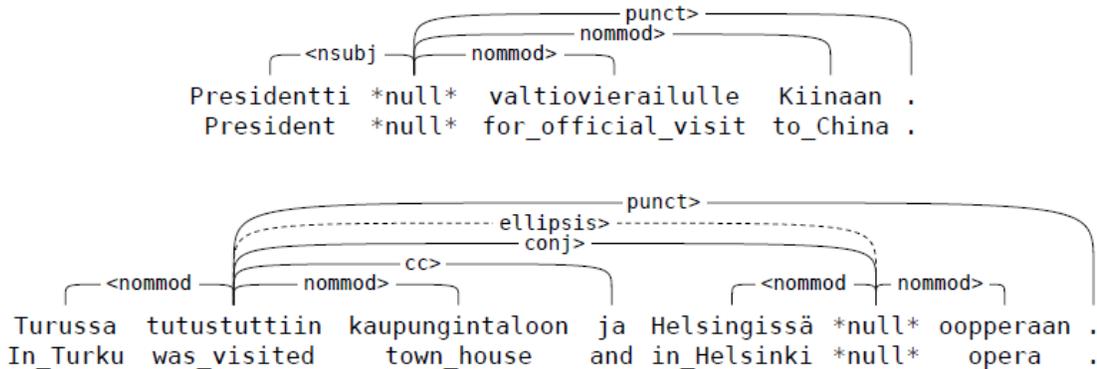
cop(big, is)

De Marneffe, M.-C., & Manning, C. D. (2010). Stanford typed dependencies manual. 20090110 *httpnlp stanford edusoftwaredependencies pdf*, (September 2008), 1-22. Citeseer. Retrieved from http://nlp.stanford.edu/downloads/dependencies_manual.pdf

- 6) In **Finnish** nominal sentences: The Finnish nominal scheme is based on the one of the Stanford Parser. The scheme refers to the nominal predicate as the head and the copula as the dependent. In addition the relation between the nominal predicate and the subject in a nominal sentence is defined as *nsubj-cop* (instead of *nsubj* which marks a regular subject). For example (on the right – English, and on the left – Finnish):



In addition in the Finnish scheme there is a marker for ellipsis for complex sentences in which in one of the parts the governing token (e.g. the verb) is elided. Instead of the elided token there is token called *null*. If the elided part appears in the first segment of the sentence then an *ellipsis* relation is defined between the first segment and the *null* token. For ellipsis of copula no *null* token is defined nor an *ellipsis* relation. For example:



Katri Haverinen, Filip Ginter, Veronika Laippala, Timo Viljanen, and Tapio Salakoski. Dependency annotation of Wikipedia: First steps towards a Finnish treebank. In Proceedings of The Eighth International Workshop on Treebanks and Linguistic Theories (TLT8), pages 95–105, 2009.

Examples from the CHILDES files in Hebrew:

leo300d.mor

*CHI: ?anī ʕayēf .

%mor: pro:person|?anī&num:sg=1

adj|ʕayēf&root:ʕyf&ptn:qatel&gen:masc&num:sg&src:deverb=tired .

%XGRA: 1|2|Aagr 2|0|Root 3|2|Punct

*INV: ?avāl ?atā racīta lanūax, ?amārta še- ?atā qcat ʕayēf .

%mor: conj:coord|?avāl=but pro:person|?atā&gen:masc&num:sg=you

v|racā&root:rcy&ptn:qal&tense:past&pers:2&gen:masc&num:sg-ta=want

la#v|nax&root:nwx&ptn:qal&form:inf=rest

v|?amār&root:?mr&ptn:qal&tense:past&pers:2&gen:masc&num:sg-ta=say

conj:subor|še=that^conj:subor|še=that

pro:person|?atā&gen:masc&num:sg=you

qn|qcat=little^adv|qcat=a_little

adj|ʕayēf&root:ʕyf&ptn:qatel&gen:masc&num:sg&src:deverb=tired .

*XGRA: 1|0|Root 2|3|Aagr 3|1|RootSub 4|3|Ainf 5|1|Coord 6|5|Asub 7|9|Aagr

8|9|Mquant 9|5|RootSub 10|1|Punct

CHI: lo? naḵōn, nlo@c [: ?anī lo?] [] ʕayēf .

%mor: neg|lo?=no

adv|naḵōn=true^adj|naḵōn&root:nkn&ptn:qatol&gen:masc&num:sg=correct

pro:person|?anī&num:sg=1 neg|lo?=no

adj|ʕayēf&root:ʕyf&ptn:qatel&gen:masc&num:sg&src:deverb=tired .

%XGRA: 1|2|Mneg 2|5|Mcom 3|5|Aagr 4|5|Mneg 5|0|Root 6|5|Punct

*INV: lo? naḵōn, ?īma? ?eynēna po, ?īma? ?ariēla yešnā po ?

%mor: neg|lo?=no

adv|naḵōn=true^adj|naḵōn&root:nkn&ptn:qatol&gen:masc&num:sg=correct

n|?īma?&gen:masc&num:sg&stat:free=mother ?|?eynēna adv|po=here

n|?īma?&gen:masc&num:sg&stat:free=mother ?|?ariēla ?|yešnā

adv|po=here ?

%XGRA: 1|2|Mneg 2|4|Mcom 3|4|Aagr 4|0|Root 5|4|Acop 6|8|Agr 7|6|Mnoun

9|4|Coord 10|9|Acop 11|4|Punct

siv611a.mor

*CHI: naḵōn še- ha- balonīm ha- ?ēle yafīm ?

%mor: adv|naḵōn=true^adj|naḵōn&root:nkn&ptn:qatol&gen:masc&num:sg=correct

conj:subor|še=that^conj:subor|še=that det|ha=the

n|balōn&gen:masc&num:pl&pl:masc:match&stat:free-īm=balloon

det|ha=the pro:dem|?ēle&num:pl=these

adj|yafē&root:yfy&ptn:qatel&gen:masc&num:pl&src:deverb-īm=pretty ?

%XGRA: 1|0|Root 2|1|??? 3|4|Mdet 4|7|Aagr 5|6|Mdet 6|4|Madj 7|2|RootSub

8|1|Punct