Guidelines for annotation of givenness

The PROIEL Project

June 20, 2011

1 Introduction

1.1 General principles

The goal of the PROIEL givenness annotation is to encode *how the intended audience determined the reference of nominal phrases found in utterances in the New Testament*. By the *intended audience* of the narrative of the New Testament we mean for the most part Hellenized Jews of the 1st century AD, holding then-current religious beliefs and having basic knowledge of the geography of the area where the NT narrative takes place. However, in the many passages of direct speech, the intended audience is the addressee, who will have knowledge of referents in the concrete narrated situation, in addition to the general knowledge we assume for persons of the period. For purposes of IS annotation, we treat 'mixed speech' (where there are indications both of direct and indirect speech) as direct speech.

We assume that hearers consult different kinds of knowledge when trying to pin down the reference of a nominal expression. Following the tradition of Discourse Representation Theory (DRT) we can model these as different *contexts*, and these contexts are the basis of our givenness tags.

Second, and also following the DRT tradition, we assume that universes of discourse referents can be embedded inside each other. Discourse referents inside such embeddings are not generally accessible from outside the embedding. Typical embeddings are negation, mood, certain classes of verbs. These give rise to *non-specific* referents.

1.2 What we annotate

The general principle is to annotate

(1) referential noun phrases (including 'null anaphora'), except relative pronouns and appositions

We also annotate the *verb* of headless relative clauses, since in our dependency grammar analysis, these are equivalent to noun phrases. (I.e. they are restrictive modifiers of null pronouns, even though we do not represent these.)

There is one exception to appositions not being tagged, namely when they depende on interrogative pronouns in constructions such as 'Who has sinned, this man or his parents', where our syntactic analysis makes *man* and *parents* appositions on *who*. These are tagged nevertheless, as they are not really coreferent with their heads.

Also, note that although appositions do not get their own tag, they can influence the tag of their heads. For example, if an apposition is a dislocated generalizing relative clause ('Whoever does this, he...'), its head will get a QUANT tag.

Notice that not all NPs are referring expressions: idioms and idiom chunks are not annotated: examples are χρείαν 'need' in the expression χρείαν ἔχειν 'have need', i.e. 'need', or ὀνόματι 'by name' in the various naming constructions. Other expressions that we consider non-referring are 'right' in ἐκ δεξιῶν 'to the right of', 'day' in ἄχρι ἧς ἡμέρας 'until the day when', and many collocations with πρόσωπον, in particular πρὸ προσώπου + genitive.

In a complex NP there can be one or several referring expressions, see section 3. Although the annotation interface presents the tags on words, it is important to bear in mind that they really apply to phrases (dependency subgraphs). If in doubt about how the NP has been analysed syntactically, consult the syntactic annotation.

In addition to NPs we annotate zero anaphora (i.e. pro-dropped arguments), most often unexpressed subjects, but sometimes also objects or obliques.

From a semantic point of view, this means most of the referents we annotate are individuals in a large sense (including humans, animals, concepts, kinds etc.). However, some noun phrases will refer to entities which are often thought to be of a different semantic type, such as events, times and places. We do annotate these noun phrases as well, but such entities will only casually be annotated, since they are often referred to by verbs (for events) or adverbials (for times and places), which fall outside the scope of our annotation.

In such cases it is often the case that NPs are linked to other linguistic material which is not itself annotated. An example would be an adverbial meaning 'thereafter' which gets the tag ACC-inf and is linked to the preceding finite verb, since it refers to a time which follows the time of the preceding finite verb of the narrative. Similarly, a NP like 'in that place' could be old and linked to a preceding adverbial 'there', even though the latter is itself not tagged.

Relative pronouns are not tagged because their referent is always identical with the head of the relative clause. Possessive pronouns are not tagged (unless nominalized), but genitives of the personal pronouns are tagged.

In noun phrases with nominalizations (e.g. articles on prepositions, adverbs, or genitives with a non-genitive article), we tag the syntactic head (which in our dependency tree is the preposition/adverb/genitive) with the information status of the whole phrase. Notice that when the head is a genitive, as in τὰ Καίσαρος, the genitive also introduces its own referent, but we cannot capture this, so we tag Καίσαρος with the information status of *the things of Caesar*, not *Caesar* himself.

1.3 Contexts of knowledge

We assume the following knowledge contexts:

1. previous discourse
2. environment context
3. conditional knowledge
4. encyclopedic knowledge

The *previous discourse* contains a number of discourse referents that have been introduced. Definite NPs and pronouns typically refer to these referents and the reference is found through anaphoric resolution. The real-world referent need not be known to the reader.

The *environment context* is the surroundings common to the speaker and the adressee. In written narratives, there is typically no clear environment common to speaker and adressee, so this context is of little use in establishing reference; but it is important in dialogue and direct speech, elements of the environment contexts are often referred to by deictic expressions.

Next there are 'shared assumptions' of speaker and addressee. These are of two kinds, one context containing *conditional knowledge* in the form of legitimate inferences, which license such definite NPs as 'I saw a man. *The nose* was red.'

Some of which are time- and place-independent such as 'people have noses', 'crowds contain people'. Others are more tied to the specific culture of the New Testament such as 'people have sins'.

Notice that these inferences are not necessarily predictive: the great majority of people have noses, but not all of them have mothers-in-law, and in 1st century Galilee, most of them would not have a house. When a mother-in-law or a house turns up with a definite article, although not previously introduced in the text, we still assume that the addressee will scan the text for a known referent from whom the relationship can be inferred.

The other kind of shared knowledge is encyclopedic knowledge: direct knowledge of referents such as 'the sun', 'Jerusalem', 'Messiah' and 'Galilee'. These are entities that the speaker can count on the addressee, i.e. any Hellenized 1st century Jew, to know.

Finally, the speaker can of course introduce new referents to the discourse, which were not previously available to the addressee.

1.4 Specificity

Some discourse referents are established inside embeddings, and their life-span is limited to the span of that embedding. In the simplest cases, the embedding extends to (part of) a sentence: 'Bill didn't see a unicorn. *The unicorn had a gold mane.' or 'Every farmer who owns a donkey beats it. *It is sad.' In other cases, the embedding is extended over several sentences, often by repeating the

embedding item or synonymous expressions. 'John wants a car. It should be red. *It is outside his garage.'

Although natural language allows several layers of embeddings, our annotation scheme only makes a difference between referents that belong to the main discourse layer (and can be picked up at any time), and embedded referents, whether in nested embeddings or not. Embedded referents get one of our *non-specific* tags.

Specificity is sometimes a understood as epistemic specificity, as 'a student' in 'A student cheated on the exam yesterday.' followed by 'We are all trying to find out who it was'. Under this view referents can be non-specific because they are not known to the speaker, although they do establish a referent in the main discourse universe. We do not treat these as non-specific.

2 The tags

2.1 General principles

The general principle of the annotation is very simple: We annotate according to the context where the referent is available to the addressee:

tag	context
OLD	preceding discourse
ACC-SIT	available in the discourse situation
ACC-INF	inferrable from preceding discourse
ACC-GEN	encyclopedic ('world') knowledge
NEW	—

KIND is used for generic referents, such as 'the lion' in 'The lion has a mane'. Such referents are also generally available from encyclopedic knowledge, but since they are of a different type (kinds rather than individuals) they get a special tag.

Similar principles, with some exceptions, hold for the non-specific tags, as we will see in section 2.4.

2.2 Specific referents

2.2.1 Old

OLD means that there is an expression in the preceding discourse which refers to *the same referent* as the element which receives the OLD tag. 'Preceding discourse' is here taken to mean the preceding text of the same book (but not necessarily the same chapter) of the New Testament.

There is a special subtag OLD-INACT which should be used for elements which are present in the preceding discourse, but too far away to appear in the window the annotator is presented with. Currently the limit is set at 13 sentences, a measure that may be adjusted when we have experimented more with the data. In such cases, the annotator must decide whether the referent

is in fact introduced anew or presented as old information, as a continuing participant in the same strain of narrative. The latter is most often the case with the important characters of the NT, such as the main disciples, but it may also be the case with frame elements such as places. An example is Jerusalem in the Easter tale, where much of story revolves around Jesus entering, leaving and reentering Jerusalem. Another example is the recurring boat in which Jesus and the disciples travel around in Galilee.

When a recurring referent is judged to be introduced anew, ACC-GEN is usually the best option. Recurrent characters must normally be considered generally known. Some characters do not participate in storylines of their own, but are instead invoked repeatedly, such as Moses, various prophets, and even God himself. For such characters, ACC-GEN, rather than OLD-INACT is the right choice.

OLD elements which are *not* inactive should receive an anaphoric link which points back to the previous mention of the referent. In cases where the preceding mention was a coordinated phrase, the link should point back to the nearest member of that phrase. If the anaphor summarizes a longer stretch of discourse, e.g. when a whole speech is referred to as 'his words' or 'this parable', we link to the last finite verb in that stretch of discourse.

There are some cases where referents are OLD but it is not really possible to link it to any previous material nevertheless. This can happen for example with temporal expressions like 'it came to pass in those days', where 'those days' are the ones being narrated in the context, but not necessarily referred to with a previous referring expression.

The personal pronouns 'you' and 'I' are always considered old, except when we have speech direct towards a general public, such as in commandments and moral advice. For the annotation in these cases, consult section **??**.

There may be doubt whether two expressions are really coreferential, as in cases where Jesus and the disciples are intermittently together and apart. In these cases the coreference analysis should be preferred.

2.2.2 Acc-gen

The ACC-GEN tag is used on elements which are not present in the preceding discourse but which are known and idenfiable to the adressee, e.g. 'the Pope'. Notice such referents must be unique in order to be identifiable.

What we consider generally known referents for the original public of the New Testament is of course dependent on our interpretation of their beliefs and knowledge. In particular, we consider that many religious concepts were known, so διαβολος is ACC-GEN. 'An angel' is not, however, since there are many angels. The cities and areas of Galilee, Judea and the surrounding regions are also considered generally known.

It is enough that the speaker can assume that the adressee is familiar with and able to identify the referent: in direct discourse, for example, knowledge that is common to the participants in the discourse will be ACC-GEN, even if they are not generally known to the intended readers.

Plural groups pose special problems in that they can be confused with kind-referring NPs. Consider the NP *the prophets*. This can refer to a known group of prophets (say the group of prophets that are acknowledged within a religious tradition, such as those who have their own book in the Old Testament). This is a uniquely identifiable group, which should get the ACC-GEN tag. If in doubt, annotators could ask themselves whether it would make sense to subsitute a list of referents for the plural NP. For kind-referring NPs this generally does not make sense.

Here is a list of concepts which we take as generally known:

- God
- (the) eternal life
- the devil
- the Evil
- Easter ...

2.2.3 Acc-inf

ACC-INF are elements which are inferrable from the preceding discourse via defeasible inferences like 'people have ears', 'grown men have a mothers-in-law'. Generally inferences from humans to body parts, clothing (including footwear!) and family members are valid, as well as inferences which build on religious beliefs such as 'people have sins'; 'demons are a general problem' and so on.

There are also inferences based on part-whole or whole-part relationships. If the text mentions Simon, and then uses a plural pronoun to refer to Simon and some (previously unmentioned people) the pronoun will be tagged as ACC-INF.

A particular case is inferences between unique elements, such as Paris and the Eiffel tower, Galilee and the Gennesareth see, Jerusalem and the temple and so on. Often both members belong to the encyclopedic knowledge of the public, but we rank the ACC-INF tag over the ACC-GEN tag and use it to signal the relationship.

Inferences should be textually licensed, ie. they should start from something which is present in the text: this is not necessarily a referent, and not always a single word, but still something present in the text. We do not recognize inferences based on narrative patterns, implicit situation types etc. In particular, the similes often assume a situation without presenting it properly in the text: 'The light does not come to be put under the bucket or under the bed'. We should not, in such cases, assume a situation with buckets, lights or beds available or inferrable, since no such situation has been presented.

Similarly, parts of the narrative in the NT are rather predictable, e.g. when Jesus arrives at a new place we may 'infer' (in a non-technical sense) that there will be people in need of healing there. But we should not use the ACC-INF tag in such cases unless the inference is licensed by some element present in the text.

On the other hand, there are items that are predicable not from the narrative structure of the NT, but from 'world knowledge': in a city, there will be some dogs; in a region, there will be some synagogues; and so on. These *can*be ACC-INF

No element should be considered inferrable from its own dependent: an inference should always be licensed by something outside the phrase, and dependents which are referential expressions should receive their own tag. If an element is inferrable from its own dependent, but this dependent is previously mentioned in the text, we use a special kind of annotation, see section 3.1.

Since ACC-INF elements must be textually licensed, they should receive an anaphoric link pointing back to the element which licences the inference. This is the only exception from the rule that links are interpreted as coreference links.

2.2.4 Acc-sit

ACC-SIT elements are elements which are available in the real world context surrounding the discourse. In a narrative there is no such context, since narrator and reader are in different places. But in direct speech we often find (implicitly or explicitly) deictic elements, which are tagged as ACC-SIT.

Sometimes, the tracking of a referent involves inference from an element which is present in the situation, ie. a combination of deixis and inference. An example would be someone pointing to a car and saying 'The battery is dead.' We count such elements as acc-sit as well, since this is the smaller category.

Cataphoric expressions in direct speech are treated as 'mental deixis' and given the ACC-SIT tag, in examples such as: 'Do you not know *this*, what David did when he was hungry?'

2.2.5 New

New elements are elements which are not present in the preceding discourse, nor inferrable from the preceding discourse, pointed out through deixis or generally known.

The label covers two different phenomena: Firstly, elements which are explicitly presented as new via various constructions. These typically do not take the article. Secondly, the label covers elements which are 'imposed' on the reader via accommodation. These are typically long definite expressions, descriptions which make it possible for the addressee to 'construct' a corresponding referent.

2.3 Generic referents

2.3.1 Kind

Generic expressions are annotated as KIND. These are phrases such as 'prophets' in 'Prophets wear sandals'. It is not always easy to tell whether an NP is kind-referring, or refers to a unique group ('the prophet of old times') or is a non-specific plural. A kind-referring NP can generally not be replaced by a list of referents: 'Prophets wear sandals' means something different from 'Jesaiah,

Eliah, ... wear sandals.', as it is a generalization which applies to all 'potential prophets'.

For the boundary between kinds and non-specific (quantified) NPs, see section 2.4.1.

KIND referents which are picked up get the OLD tag. (This may change later.)

KIND is also used for abstract concepts such as 'death', 'illness', names of specific illnesses and so on, when the reference is to the general concept. However, these concepts are typically 'instantiated' in specific situations, e.g. when 'someone meets his death'. These uses are tagged with specific tags.

It is sometimes difficult to draw the borderline between such concepts and unique concepts, which are always identical and generally known, so that they get the ACC-GEN tag. Examples here are 'eternal life', 'festival of easter' (implying that it is in some sense 'the same easter' which arrives every year).

NPs containing 'such' (τοιοῦτος) will in most cases be KIND.

Kind-referring NPs in cases of contained bridging are treated as other cases of contained bridging, i.e. NEW + anchoring. TODO: check in Sermont of the Mount

2.4 Nonspecific referents

Nonspecific referents are those which only exist inside certain embeddings, such as negation, modality, quantification etc. They cannot be referred to from outside this embedding, which is therefore called an opaque context. An example if found in Luke 8:16

(2) No one lights a lamp and hides it.
Οὐδεὶς δὲ λύχνον ἅψας καλύπτει αὐτὸν

The lamp can be referred to by a pronoun inside the scope of the negation. But outside of negation, e.g. in the next sentence, it does not make sense to use a pronoun *it* to refer to the lamp which no one lights. So, *a lamp*/λύχνον is non-specific and cannot be picked up in the following discourse. So it gets the NONSPEC tag. However, it can be picked up inside its own embedding, in which case we use the tag NONSPEC-OLD. In a similar way, there can be inferences from a non-specific referent inside an embedding:

(3) "No one sews a patch of unshrunk cloth on an old garment. If he does, the new piece will pull away from the old, making the tear worse.

The second sentence continuous the embedding created by the negation in the first one, and *the tear* clearly belongs to the mentioned *old garment*, so it should get the NONSPEC-INF.

Notice that interrogative pronouns and negative pronouns will always get a nonspecific tag.

2.4.1 Quantification and delimitation of non-specific referents

We distinguish two basic types of non-specific referents, the quantifier restrictions and 'the rest'.

Quantifier restrictions are the words which determine the range of the quantifier, such as *students* in

(4) All students love a professor

Here, *students* is tagged as a QUANT, whereas professor is NONSPEC on the reading where it varies with the student which professor they love; or new, if professor outscopes the quantifier and all students love the same professor.

There are clear formal criteria for what we consider quantifier restrictions: the noun must appear in either a quantifying construction or be modified by a quantifier. The following constructions can be (and mostly are) considered quantifying:

- headless relative clauses
- substantivized participles with a definite article (and without specific reference)

The following words are consider quantifiers:

- πολύς
- πᾶς
- τοσοῦτος
- πόσος

In cases where the restrictor is just implicit, the quantifier itself will be the syntactic head of the phrase and therefore get the QUANT tag. Notice that it is possible to link from quantified expressions back to antecedents which refer to the same referent, e.g. *the people... and all the men*. Such links do not necessarily express coreferentiality, but can also be inference links, e.g. in *the house... all the inhabitants*.

If there are further mentions of the NP which serve as the quantifier restrictor, they should be linked back to the previous *specific* instance, skipping the QUANT-tagged NP, e.g. *the crowd* is linked to *the people* in *the people... and all the men ... the crowd*. If, on the other hand, there are new quantifications over the same domain, the second quantified NP also gets a quant tag, but is linked back to the previous quantified phrase.

However, more often than not in such cases, there is no mention of the referent before the QUANT-tagged one: instead, a referent is introduced with a quantified expression, and the text then goes on to talk about this referent in a specific manner. In such cases, the NON-SPEC OLD tag is used for all further mentions in the same sentence; if there are mentions in later sentences, a specific referent is introduced via inference from the last mentioned quantified phrase.

We now turn to the other non-specific tags. The most difficult problem is the treatment bare plurals and bare singulars of mass noun. In Greek, these are ambiguous between a specific and a non-specific reading. As a test, the annotator can ask whether it makes sense to add 'certain' to the translation (or for mass nouns, 'a certain amount of'). If so, the NP should get a specific tag. This test will not always yield conclusive answers. When it doesn't, the annotator should simply check if the NP gets picked up in the following discourse (i.e. in a new sentence, or in the same sentence, but with a different syntactic function). If yes, it should get a specific tag; if no, a non-specific tag. NPs that have a cardinal numeral are treated like bare plurals.

In other cases it is uncertain whether a (often singular) NP is kind-referring or non-specific. Consider the following pair:

(5) If man lands on the moon, it will be a great step forwards.

(6) If a man lands on the moon, it will be a great step forwards.

man in 5 is kind-referring, but *a man* in 6 is non-specific. The semantic difference here is subtle, since mankind cannot land on the moon without some invidual man landing on the moon. However, the article resolves the issue, since indefinites are not normally kind-referring in English (except on a taxonomic reading). In Greek, on the other hand, there is no indefinite article, and much less is known about the formal expression of generic NPs. In such cases, the annotator should ask whether the NP is in anyway referentially dependent on another element of the sentence. If not, it should get the kind tag; else, a non-specific tag.

The same reasoning can be used for plural NPs:

(7) John the baptist wandered in deserted places

The 'deserted places' is here referentially dependent on 'John the baptist' in the sense that it denotes those deserted places in which he wandered. In other words, the sentence can be paraphrased:

(8) It is generally true of John the Baptist that he wandered in deserted places

and not

(9) It is generally true of deserted places that John the Baptist wandered in them

which would be the reading where 'deserted places' were kind-referring.

2.4.2 Verbs that create opaque contexts

Here is a tentative list:

- λέγω (say)

- ποιέω (make)
- ὁράω (see)
- οἶδα (know)
- λαλέω (speak)
- ἀποκρίνομαι (answer)
- λαμβάνω (take)
- πιστεύω (believe)
- θέλω (desire)
- γινώσκω (know)
- εὑρίσκω (find)
- ἐσθίω (eat)
- γράφω (write)
- ἵστημι (stand)
- ἀφίημι (send away)
- ἀποστέλλω (send away)
- ζητέω (seek)
- ἀγαπάω (love)
- αἴρω (lift up)
- κρίνω (select)
- τίθημι (put)
- φοβέω (be afraid)
- γεννάω (become the parent)
- πέμπω (send)
- ἐρωτάω (ask)
- αἰτέω (ask)
- κηρύσσω (announce)
- μαρτυρέω (be a witness)
- δοκέω (think)

- ἐπερωτάω (ask)
- φέρω (bring)
- θεωρέω (see)
- δοξάζω (praise)
- προσκυνέω (worship)
- πείθω (convince)
- διώκω (harass)
- θαυμάζω (be impressed)

3 Complex NPs

By complex NPs we mean NPs that contain one or more other NPs, such as *the brother of John the Baptist*. In most cases, each of these NPs get their own tag, but not always.

In many cases, the dependent NP is in some way known to the hearer and thus serves to identify the referent of the head (*the man's ears*). These are treated in section 3.1. In yet other cases it merely serves to specify a type. These can often be translated with possessive compounds in English. Examples are σπήλαιον λῃστῶν 'a thieves' den' and τρυμαλιᾶς τῆς ῥαφίδος 'a needle's eye'. Notice that, unlike in English, the depedent *can* take a definite article and is thus a full NP and not an N. Nevertheless, such NPs seem non-referential and should not get a discourse tag.

The distinction between dependent genitive NPs that help identify their head and those which specify the type of the head is not always clear in Greek. To simplify matters, we have decided to treat all NPs that would otherwise be kind-referring as type specifiers. They therefore do not get a tag, even in cases where they could seem refential, such as θεὸς νεκρῶν 'God of the dead'. Also cases like χώραν Γερασηνῶν 'the country of the Gerasenes' only get one tag, since the genitive is kind-referring. This principle holds for genitives that are ATR or PART dependents of other nouns. Although they are not themselves tagged, these dependents can still influence the choice of tag for their heads.

Notice the special treatment of NPs that have the same reference as their heads and precede their heads, such as ἑαυτοῦ in ἀφ' ἑαυτοῦ λαλῶν. It would be counterintuive to tag this as new (and its head as old and anaphorically pointing to it). To avoid this problem we simply refrain from tagging such NPs.

Finally, there are some special genitive constructions where the genitive does not get a tag, see section 3.2.

3.1 Contained bridging

If a referent which has not previously been mentioned is expressed by a complex NP which contains info-taggable dependents (excluding those inside appostions), special rules apply. We need to capture the fact that the NP itself contains the material necessary for the identification of the referent, so we use the NEW tag - even if the referent is also inferrable from something in the context. The relevant dependents are then tagged with their appropriate tags. Consider the following example:

(10) καὶ ἀπολαβόμενος αὐτὸν ἀπὸ τοῦ ὄχλου κατ' ἰδίαν ἔβαλεν τοὺς δακτύλους εἰς τὰ ὦτα αὐτοῦ
'He (= Jesus) took him aside from the crowd and put his (= Jesus') fingers in his (= the man's) ears'

To establish the referent of τοὺς δακτύλους we need to link it to a previously mentioned referent, so this is an ACC-INF. In τὰ ὦτα αὐτοῦ, however, this linking has been done for us (apart from the fact that we need to establish the referent of αὐτοῦ first, but that is another matter), so the NP provides the information we need and is considered NEW.

Three more points are worth making in this connection:

1. We only use this style of tagging if there is a licit inference from the dependent's referent to that of the head. F.ex. σκιὰν αὐτοῦ is OK since things have shadows, θυγάτριόν μου since many people have daughters, τὸ θυγάτριόν μου since many people have only one daughter. By contrast, consider

   (11) ἄφες πρῶτον χορτασθῆναι τὰ τέκνα· οὐ γάρ ἐστιν καλὸν λαβεῖν τὸν ἄρτον τῶν τέκνων καὶ τοῖς κυναρίοις βαλεῖν.
   "First let the children eat all they want," he told her, "for it is not right to take the children's bread and toss it to their dogs."

   There is no valid inference from children to bread: we need an inference from the verb 'eat' to establish the reference. So 'bread' is tagged as ACC-INF.

2. The inference should be from the dependent's *referent* to that of the head, irrespective of how that referent is realized: βόσκοντες is inferrable from αὐτοὺς in βόσκοντες αὐτοὺς in a context where αὐτοὺς refers to a herd of pigs, since these often have a herdsman; βασιλείας μου is also OK in a context where μου refers to a king - the relevant inference is "Kings have kingdoms", not "People (or 'I') have kingdoms"

3. Appositions (including non-restrictive relative clauses) by definition do not contribute to the establishment of reference, but only offer an alternative, coreferential expression. Therefore, the rules described in this section do not apply to elements which would be inferrable from an element in an

apposition. Consider τὸν Πέτρον καὶ Ἰάκωβον καὶ Ἰωάννην τὸν ἀδελφὸν Ἰακώβου. Peter, James and John are inferrable from a previous mention of οἱ μαθηταὶ αὐτοῦ and are all treated as ACC-INF. The fact that the OLD referent James appears inside an apposition to John does not change its status.

3.2 Special genitives

3.2.1 Complex nouns

Some genitives are not referential, but only part of complex nouns. Examples are *the son of man*, the title Jesus uses to refer to himself, and *the mount of Olives*, a hill outside Jerusalem. *man* and *olives* are not tagged in these constellations. Most but not all of these are proper names.

We use non-compositionality as the criterion to decide whether a gentive is part of complex noun which has a single referent. Phrases which function as proper names, but are compositinal (e.g. ἡ βασιλεία τοῦ θεοῦ 'the kingdom of God') are not treated as complex proper names. In this case the dependent *is* tagged

When the head is a proper noun, we also consider dependent genitives parts of this name. So Ἰάκωβος τὸν τοῦ Ζεβεδαίου 'John son of Zebedaeus' is treated as meaning 'John Zebebaieson'.

3.2.2 Descriptive genitives

The Greek of the New Testament sometimes uses the genitive of an abstract noun in cases where we would expect an adjective in 'normal' Greek. Examples are 'son of peace' (meaning 'peaceful man'), 'judge of injustice' (for 'unjust judge') and so on. These genitives are not tagged.

3.2.3 Definitive genitives

Some genitives, the definitive genitives of Greek and Latin school grammar, are appositional in character. An example is 'the country of Norway', 'arbor fici', τὰ ὅρια Τύρου 'the country of Tyros'. Generally a genitive can be considered definitive if it is the case that a sentence of the form 'DEPENDENT is a HEAD' (e.g. 'Norway is a country') is true. Just like appositions, these genitives are not tagged.

4 Some special constructions

4.1 Subjects of deontics and 'one' (= German *man*)

There are a lot of deontic statements in the New Testament. The subject of these can be expressed with a 2. person verb with a pronoun or a prodrop; or a 3. person/infinite verb with a *pro*. These are all tagged as KIND.

TODO: fix up the Sermont of the Mount in Luke

4.2 Comparisons

In comparisons like *You walk like a duck*, indefinite NPs are typically treated as non-specific. (And not kind, which would often be an option, since comparisons typically refer to generic properties.

4.3 Old Testament quotations

The NT contains many quotations from the OT, and these are often devoid of the context necessary to interpret the reference of NPs, so we can find *a voice crying in the wilderness]* without any previous mention of wilderness. Such referents are tagged as ACC-GEN, although references to speaker and addresee as well as pronouns which would find their reference in the original context are still tagged as old (without a link).

4.4 Coordination

NP coordination most often involve two NPs that each create a referent, and then those two referents are 'summed' to a complex referent. In these cases, each NP gets the tag it should have based on normal criteria.

In some cases, though, the two NPs are coreferential and just offer different descriptions of the same referent, or – in the most frequent – case, coordinated restrictions on a quantifier, as in πᾶς ὁ θεωρῶν τὸν υἱὸν καὶ πιστεύων. In these cases, the second conjunct should not be linked to the first one, instead, each conjunct is given an independently motivated tag, which in almost all cases will be the same for both conjuncts.

4.5 Temporal expressions

We do not tag expressions of temporal extension, e.g. NPs such as "for three weeks" and so on.

4.6 Noun phrases containing co-referent NPs

Sometimes an NP has a dependent which is coreferent with the whole NP, e.g. ἑαυτοῦ in ὁ ἀφ' ἑαυτοῦ λαλῶν. If such a coreferent dependent follows the head, it can simply given the OLD tag and an anaphoric link to the head. However, if it precedes the head, we would get an unnatural annotation where the dependent 'introduces' the new referent, and the head is OLD and coreferent. To avoid this, we simply do not tag these dependents.

4.7 Light verb constructions

Insert list here.

5 Prodrop elements

The annotation tool allows for insertion of prodrop dependents on any token with verbal morphology: prodrop subjects, objects and obliques are all possible. However, we do not add prodrop elements to all verbs. In particular we do not insert prodrop tokens which are already present in the syntactic annotation via an XSUB-slash. We do not attempt to model the difference between VP coordination and sentence coordination, so we add prodrop tokens even if they are already indicated via 'shared argument' slashes. The only exception is relative clauses with coordinated verbs: we do not add prodrop tokens corresponding to the shared function of the relative pronoun (but we do for any other shared arguments).

We do not add prodrop subjects to adnominal, attributive participles. We do, however, add prodrop subjects to infinitives without XSUB slashes. This would typically be the case for articular infinitives. Imperatives do get prodrop subjects, but not when they are tagged as VOC, i.e. when the syntactic analysis suggests they are no longer verbs, as is sometimes the case with *ide* (in the meaning of *idou*).

Prodrop elements are typically OLD, but in a number of cases other tags are warranted. This is the case with second person prodrops, which are sometimes NON-SPEC. Also, NON-SPEC is used in cases like *Some guys came and circumcised him* - although in such cases, even NEW is warranted if the referent is picked up in the further discourse. (This is the same criterion as for bare plurals.)

6 Anaphoric links

Anaphoric links have been covered in the sections on OLD and ACC-INF referents. A brief summary of the main principles:

- Anaphoric links are set to have a maximum length of 13 sentences
- There are two kinds of links: coreference links and inference links. The latter only occur on elements which are tagged as ACC-inf – all other links are interpreted as coreference links, even when they occur on material tagged as NEW, e.g. in direct discourse
- Links to coordinated referents point to the last coordinate
- Anaphors that summarize a stretch of discourse or a complex utterance should have a link pointing to the last *finite* verb of that utterance.

6.0.1 Split antecedents

In cases where a plural pronoun has a split antecedent, it should be annotated as old, and the anaphoric link should point back to the last referent which is part of the group denoted by the pronoun. Notice that there can be cases where a plural pronoun does not have a split antecedent - but only one singular subset

antecedent, such as when the discourse moves from 'I' to 'we'. In these cases, ACC-INF is used instead of OLD.

6.1 Direct speech

Regarding the use of info tags, direct speech is treated as a separate universe which does not have access to the surrounding narrative, although the narrative has access to the direct speech. But anaphoric links mark coreference even across the boundaries between direct speech and narrative, so there can be coreference links out to the preceding narrative. However, remember that these should always be coreference links: this means that if a direct speech is introduced in a non-specific context ('Whoever says'), the 'I' of the direct discourse should *not* be linked back to 'Whoever', since it receives a specific tag.