

Progress on modifying the ontology

Key:

Red typeface – work that is done

Dark green – work that is in progress

Dark blue – work postponed until later

~~Strikethrough~~ – disagreement with recommendations

Black – items that have not yet been addressed

1. Working Group 1

1.1 Separate form and function

1.1.1 More than one level of the form/function distinction:

1.1.1.1 Affix vs case

1.1.1.2 Case vs location

1.1.2 Need to be able to define:

1.1.2.1 Form w/out function: stem joiners

1.1.2.2 Function w/out form: zero morphemes

1.1.3 Same morpheme can be used for many functions.

1.1.4 There are functions represented by templates or processes:

1.1.4.1 Reduplication

1.1.4.2 CVC templates

1.1.4.3 Tone patterns

1.1.5 Mystery morphemes, e.g. things that get labelled SUF

1.1.6 Forms that sometimes have a function and sometimes do not

1.1.6.1 Athapaskan classifiers

1.1.7 Note: may be dependent on phonological shapes of bases.

1.1.8 Note: need to be able to enter the pattern

1.2 Flat structure

1.2.1 Hierarchy doesn't give you much (except for shrinking size of feature lists)

1.2.2 May cause you to make choices that shouldn't be made:

1.2.2.1 Open question whether lgs like Salish have a distinction between nouns and verbs. Ontology-based tool shouldn't force a choice in order to use a feature like 'transitive'.

1.2.3 Specific categories to be flattened:

1.2.3.1 Transitivity (not under Verb)

1.2.3.2 Definiteness (not under Determiner)

1.2.3.3 Possession (also not under Determiner)

1.3 'See also' relation: lateral connections

A list of relevant but not strictly related terms.

1.3.1 Salish feature 'control' involves notions of volition but isn't strictly a

VolitionFeature. But people who are interested in volition would certainly want to

find Salish control.

1.3.2. Noun incorporation: not always exactly a compound. (might use the term "incorporate").

1.4 Grammaticalization

1.4.1 A continuum from lexical to fully grammaticalized, sometimes in the same paradigm (e.g. directionals) sometimes in the same morpheme (dependent on context)

1.4.1.1 Body part prefixes

1.4.1.2 Come&go: auxiliaries/direction of motion

1.4.1.3 Classifiers

1.4.2 There are things that some people think are lexical and some people think are grammar and they're not always divided up in the same way.

1.5 Portmanteau morphemes & cross-cutting features

1.5.1 Person + voice

1.5.2 Person + tense/aspect

1.5.3 Tense + aspect

1.5.4 Agreement: Person X acting on Y

1.6 Alignment to hierarchies

1.6.1 Morphemes that indicate both subject (agent) and object.

1.6.2 Can't just add a feature "1>2" because searcher wants to know that 1 and 2 are involved and what the relation is.

1.6.3 A complex feature

1.6.4 Person hierarchies, animacy hierarchies

1.7 Gender of speech act participant

1.7.1 E.g. forms that are only used by women

1.7.2 These are specific morphemes that have to be glossed correctly

1.8 Default features

~~1.8.1 Sometimes you don't want to label something as A, B, not A, or not B, because sometimes the other thing may have a wider range that actually sometimes includes A, but it isn't correct to say that it's A.~~

~~1.8.2 Want to be able to say "B" without implying "not A"; or "not B" without implying "A"~~

1.9 Conflated features

1.9.1 If the tool has you give a list of features for some morpheme it may force you to imply OR rather than ALL AT ONCE.

1.9.2 Ex. Morpheme that means 'at' and 'in' in a sense in which these notions are conflated in the language.

1.9.3 Not the same as polysemy...

1.10 Inherent vs derived functions

1.10.1 E.g. transitivity

1.11 User interface issues

- 1.11.1 Expandable trees might prevent you from finding what you want (e.g. might not think to look for demonstratives under Determiner)
- 1.11.2 Commenting should be encouraged at every step of the data entry process. Try to get at the layer behind the label (?).
- 1.11.3 Need to be able to specify many things about a morpheme, at least:
 - 1.11.3.1 Form (e.g. infix, clitic, phrase)
 - 1.11.3.2 Function (e.g. nominalizer, aspect, distributive)
 - 1.11.3.3 'See also' list of features
- 1.11.4 A glossary of standard combinations:
 - 1.11.4.1 X form, y function: gets name X
 - 1.11.4.2 Y form, x function: gets name Y
- 1.11.5 UI shows canonical examples
- 1.11.6 Subset menus based on COPEs (If the goal is to limit what the user sees, this is another way to do it that doesn't force hierarchical relations among features)
- 1.11.7 If you type in a term like 'applicative', interface pops up a page asking you to pick which kind of applicative you're looking for.

1.12 New categories/features I

- 1.12.1 ClassifierFeature as a sister to GenderFeature
- 1.12.2 Alienability
- 1.12.3 Phrasal affixes (complementizer)
- 1.12.4 Derivational affixes (nominalizer)
- 1.12.5 Thematically rich affixes:
 - 1.12.5.1 Body part prefixes
 - 1.12.5.2 Incorporates
 - 1.12.5.2.1 Nouns
 - 1.12.5.2.2 Prepositions
 - 1.12.5.2.3 Adjectives
- 1.12.6 Move possession out from under Case
- 1.12.7 Resultative: under aspect
- 1.12.8 Distributive: under Quantifier
- 1.12.9 Move Transitivity out from under Verb
- 1.12.10 Cislocative, translocative: portmanteaus of participant + direction
- 1.12.11 Salish control vs. non-control: ~ control is volitional, non is nonvolitional/accidental. Control 'pour' = pour vs non-control 'pour' = 'spill'. Noncontrol can also mean 'manage to get it poured'. Noncontrol has implications of completion. Connections to categories like volition.

1.13 Final note

- 1.13.1 Even the most perfect ontology will never be a substitution for RFG: Reading the Flipping Grammar
- 1.13.2 Goal of ontology: to get you to the grammar.

1. Working Group 2

2.1. Parts of Speech (rename “lexical categories”?)

2.1.1 Our immediate reaction to the ontology covering parts of speech was the varied nature of the definitions used. Some definitions are semantic, some morphosyntactic, and some a combination of the two. **Semantic definitions are notoriously problematic as means to identify parts-of-speech categories, so we recommend not only that the definitions be regularized, but that the semantic properties not be presented as definitional.** The ontology should allow for each the lexical categories of each language to be expressed based on their morphosyntactic behavior, with no expectation that every language will have every class, even those as “basic” (from a Eurocentric point of view) as noun and verb.

2.1.2 A second point that came up is that whether or a particular part of speech in a given language is represented by free or bound morphemes can change over time and across closely related dialects. Some classes that we would think of as parts of speech, such as classifiers, are necessarily bound in some languages, while in other languages they may be free. Therefore, we recommend that the dimension of morphological boundedness be separated out from the notion of part of speech and treated as orthogonal. We recommend that GOLD developers look at Dixon and Aikhenvald’s (2002) work on wordhood and that the notions of grammatical word and phonological word might prove useful to linguists using the ontology.

2.1.3 In discussing classifiers, another issue related to word-hood came up, which is how to treat the word that results from the combination of a numeral, demonstrative or quantifier with a following classifier. The combination forms a grammatical and phonological word in a language like Mandarin, but it isn’t a “part of speech”, instead it is phrasal. This brings up the point, returned to below, that the ontology needs to develop to include phrasal units and constructions (i.e., find a way to discuss levels of syntactic constituency higher than the word).

2.1.4 The first specific observation we made at the level of the lexical category is that in some languages with classifiers the numeral ‘one’ plus the general classifier has grammaticalized into an indefinite article (Givón 1981). Interestingly, in such languages there is an indefinite article but no definite article. Thus we need the notion of article and of indefinite article in the absence of definite article. In writing this up now, I wonder whether the concept “article” is even necessary if indefinite and definite article can occur independently. This observation is in line with the idea that a flatter structure to the ontology might be preferable.

2.1.5 We also made some observations on coordinating conjunctions. Some languages have coordinating conjunctions that work at many levels of structure,

such as English *and*, which can join nouns, adjectives, NP's, clauses, etc. In other languages (such as Newar), conjunctions can conjoin nouns or noun phrases, but cannot join adjectives or adverbs, or work at higher levels of structure, such as the clause level. Therefore, you might want to consider sub-classifying coordinating conjunction by the lexical or syntactic units which they may conjoin. (Note languages without clausal coordinating conjunctions will often use serial verb constructions or converb/clause-chaining constructions to fulfill functions covered by conjunctions in languages like English.) See Genetti 2005, Bisang 1995.

2.1.6 The lexical category of demonstrative needs to be significantly sub-classified. Proximal and distal are the most basic divisions. Other languages sub-classify demonstratives based on the vertical dimension, reference to speaker and hearer, orientation to a river, etc. In addition, some languages have distinct demonstratives for locations (here, there, thither, yon), types, manners (like this, like that), size (this big, that big), and quantity (this much, that much) (Diessel 1999).

2.1.7 Regarding classifiers, it was noted that there are a number of syntactically distinct types of nominal classification systems, including numeral classifiers, noun classifiers, nominal classifiers, etc. (Aikhenvald 2003). The category of classifier needs to be expanded to include these. Currently classifiers are listed under nominal particles; they need to be moved from here and made their own category. It would probably be better to have "classifier" as a part of speech, which sub-classifies by morphosyntactic behavior.

2.1.8 Regarding the semantics of classifiers, if one compares across types of classification systems, one finds that there are few semantic commonalities. For example, while animacy and human/non-human may be marked by numeral classifier languages, they are not relevant to noun classifiers, such as those in Dyrbal. In addition, languages differ with respect to which categories are marked and which categories are elaborated. It was noted that some languages allow different choice of classifiers in order to allow different classifications of the same noun. (Becker (1975: 113) illustrates this nicely in Burmese, where a host of different classifiers can be used with the noun 'river', each emphasizing a different aspect of the way people interact with rivers (e.g. as a route of transportation, a means of sustenance, a curvy thing in its shape, etc.).

2.1.9 The category particle needs further refinement. It should be decided whether particles should be subcategorized based on semantic properties (e.g., evidential particle, question particle, negative particle, topic particle) or on position in the clause (NP-final, second-position, sentence-final). These seem to be independent dimensions, and the ideal would be for the ontology to allow both types of sub-classification and the ability to link between them. Then it would be interesting to track which particle types are most likely to occur where, allowing, of course, that other connections are possible.

2.1.10 There are multiple ways of categorizing verbs. Currently the ontology includes intransitive, transitive, and ditransitive. Copula should be added to this list. In addition, a separate category for auxiliary verb is needed. There are many ways in which verbs sub-classify semantically. GOLD developers need to decide on a principled basis on which to do this. Again, would multiple subclassifications be appropriate?

2.1.11 A few smaller observations on particular parts of speech are:

~~2.1.11.1 adverbs are of course problematic, and it is unlikely that an umbrella category “adverb” is appropriate;~~

2.1.11.2 add interjections;

2.1.11.3 add ideophones;

2.1.11.4 move “existential markers” to be a sub-class of verbs;

2.1.11.5 possibly add honorifics (they are not a separate part-of-speech in most S-T languages to our knowledge, but might be in other language families (in Tibetan, they sub-classify other parts of speech, e.g., there are honorific and non-honorific nouns, honorific, non-honorific, and humilific verbs, etc., so this could be a cross-cutting dimension there). Many languages have honorific pronouns, so this is another dimension to add there.

2.1.12 Moving to a larger issue, there are some lexical categories in languages that are unusual. One such case was described by Genetti and Hildebrandt (2004), the case of “verbal adjectives” or “adjectival verbs” in Manange. Manange is a Tibeto-Burman language spoken in the Annapurna region of Nepal. That language has well-defined categories of adjective and verb. In addition, there is a third class of lexemes that are intermediate between these two, having the morphological behavior of verbs but the syntactic behavior of adjectives. We wondered how the ontology would deal with this case. We considered the following possibilities:

2.1.12.1 allow multiple categorizations of one category, in this case the category would map to both verbs and adjectives, although it is important that it not be seen as a supercategory that combines properties of both;

2.1.12.2 define it by its relation to the two other classes, that is, have a relation “morphology of X” and “syntax of X” and allow the category to map to both, however, it is unclear how useful this will be outside of the Manange case;

2.1.12.3 allow an “other” category here and elsewhere in the ontology (throughout?) that allows language-particular definition of categories and makes explicit our knowledge that many languages have unique features;

2.1.12.4 allow for lexical categories that result from the intersection of two other categories (visualize a Venn diagram) and that allows explicit specification of the particular properties of each class that are shared; this is the most satisfying account descriptively.

2.1.13 In addition to the discussion of particular parts of speech, we began to consider the ontology of constructions, especially in relation to the notion of auxiliary verb (which one can argue is a separate part of speech) and serial verb (which can have similar functions as auxiliary verbs but is defined at the level of construction). We considered briefly the following construction types which result in the sequencing of verbs, verb phrases, or clauses: serial verbs, verb-auxiliary constructions; converbs/clause-chains, etc. These overlap substantially in their functional possibilities but can be differentiated on the basis of morphosyntactic properties. This argues for a clear separation of the form and function and an explicit means to link between them. We will return to this point below.

2.1.14 We spoke briefly about reduplication, which has phonological, morphological, syntactic, and semantic dimensions. Where to include this in the ontology was a matter of debate, with one member advocating a positioning in phonology, and another thinking that perhaps other areas would be as appropriate. It was noted that there are a number of semantic features which typically are linked to reduplication, such as plural, collectivity, iterativity, durativity, and intensification. One suggestion is to add “reduplicant” to the ontology, and then allow independent specification of its phonological properties (full, syllable-level, root, mora, etc.) and pre-association with particular semantic concepts. It was noted that the Open World assumption would allow other semantic links to be instituted as needed.

2.2 Other considerations

2.2.1 Having concluded our discussion of parts of speech, we moved on in the ontology and discussed aspect. We recommend that you consider adding viewpoint and situational aspect to the list of semantic features, otherwise they look fine. We noted that like many other semantic features, the aspectual features are realized in language by a variety of forms, including affixes, particles, auxiliary verbs, and serial-verb constructions. Once again, we recommend explicitly incorporating into GOLD the independence of the formal and functional dimensions, and then allowing mechanisms to link the two. It would be helpful then to have a list of tendencies showing which formal categories tend to link to which semantic features.

2.2.2 Moving now to area-specific recommendations for the future development of GOLD, we suggest you consider building implicational universals into the ontology (e.g., if a language has a dual, it also has a plural). Also, with such a large inventory of features, we wondered whether GOLD should be constrained so that certain features do not co-occur (such as the relation 1>3 and inverse).

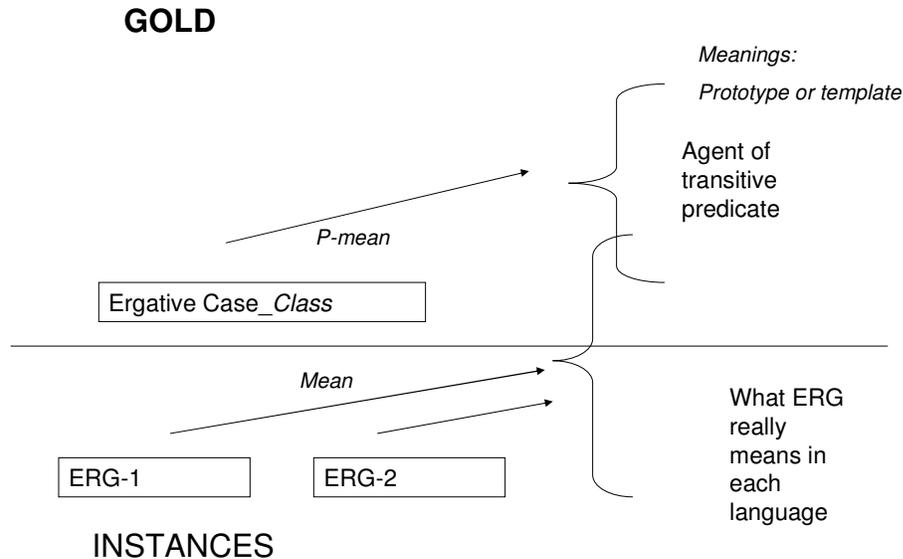
2.2.3 We had a very interesting discussion with regards to grammaticalization. Linguistic categories are often not discrete but range along a cline and over time they shade into one another through the gradual process of language change. This is difficult to incorporate in the current conceptualization of GOLD, which treats categories as a set of discrete elements meant to map to linguistic forms.

2.2.4 We wondered how GOLD could be modified to account for typical grammaticalization patterns, such as the reanalysis of the numeral ‘1’ plus a general classifier into an indefinite article (Givón 1981), the development of the lexical verb ‘stay’ into a marker of progressive aspect (Bybee et al 1994), or the development of third-person-singular pronouns into distal demonstratives. We would like to suggest that E-MELD organize a working group that would construct an ontology of common grammaticalization patterns (possibly in a COPE). Heine and Kuteva (2002) would be an obvious resource for this group. One question the group could address is what is inventory of relations between grammatical classes as exemplified in historical change. Of course, during the process of grammaticalization, the original and innovative uses of a form go through a period (sometimes unresolved) where multiple functions of a given form are possible (sometimes called “layering”) (Hopper 1991). As with other cases of polysemy, the function of a particular form in a particular instance will be determined by context (or even will be unresolved in certain contexts). Grace Wiersma provided a nice example of this from Cantonese, where there has been an extension of the bare Cantonese CLF to POSSESSIVE, but only with subclasses of nouns, e.g., body parts. So you have ONE + CLF + dog meaning ‘a dog’ but CLF + tongue meaning ‘his tongue’, i.e., ‘the dog’s (tongue)’ We wondered how one can derive a context-dependent assignment of a particular mapping to the ontology when multiple mappings are specified. In addition, is there a way to specify in the ontology that two meanings are likely to be realized by a single morpheme (e.g., ‘stay, sit, live’ and progressive aspect)? It also might be useful to know that certain grammaticalization patterns do not occur.

2.2.5 A number of larger theoretical questions came up during our final discussion. We wondered whether the goal of the ontology is to produce a description of what is possible in natural language or to provide interoperability between different language resources (both of which came up at various points during the conference) and whether there are tensions between these different goals. Related to this issue is whether GOLD is a metalanguage for doing linguistics or a metalanguage for finding language resources. The structure of GOLD should differ depending on the ultimate goal. We also wondered how theory-neutral GOLD really is, and whether implicit theoretical assumptions might be restrictive of the types of generalizations that the ontology is allowed to express. It will not come as a surprise that we didn’t resolve these questions, but we feel that it is important that GOLD developers consider such foundational issues frequently as they are in the process of developing GOLD.

2.2.6 Finally, it was noted that languages carve up functional space in different ways. For example, ergative case in one language will not occur on exactly the same set of expressions as it would in another language. We wondered how GOLD will incorporate this insight. We discussed various possibilities, including the idea that categories such as ergative have prototypical meanings – or possibly

templates, or core meanings – that are expressed at different level than the classes themselves. We constructed the following diagram to elucidate our thinking:



In this diagram, the GOLD ontology is represented above the horizontal line. It has a category of ergative case that is linked to a prototypical meaning or template by a relation “p-mean” (prototypical meaning). Particular instances of ergative case in particular languages are represented below the line. These are each linked to meanings by the relation “mean” (meaning, specific to that language). The main point is that each of the meanings overlaps to some extent the prototypical meaning of the ontology. One could also use Venn diagrams to show this. Perhaps the prototypical meaning would be the intersection of all the particular meanings.

2. Working Group 3

3.1 Overarching questions

3.1.1 Where does linguistics stop? Who is our audience?

3.1.1.1 e.g. Q2. If our audience includes descendants of speakers, should we be including kinship structures in our ontology since these have linguistic manifestations, e.g. kinship pronouns, kinship agreement markers

3.1.1.2 Suggest that working groups be formed with librarians for examples to create ontologies for these overlapping domains overall, and with anthropologists for a specific domain such as kinship.

3.1.2 Is our audience people at the start of data analysis, or at the archiving end of it, or both?

3.1.2.1 Yes to both.

3.1.2.2 We need an LINGS101 COPE – possibly a generic COPE=GOLD – which contains what a graduate student sitting down needs to understand. So they can look at their data and make a pass at analysing it. In fact we need them to start early so that they don't get wedded to weird understandings and labels.

3.1.2.3 We need people to consider GOLD when archiving their data. However, the focus is not on helping linguists but on plugging the results into the worldwide community – we want to buy interoperability. The hook for the individual is making work open to typology, and perhaps for gaining help as to what other people have called similar concepts.

3.1.3 How do people learn what the labels mean?

3.1.3.1 Same as Group 2: The definitions are inadequate, and it is hard to find the paper/online versions. We need to have attached to each label

3.1.3.1.1 a field guide to that label e.g. for perfective you'd go to Östen Dahl's Field guide to TMI,

3.1.3.1.2 lots of examples that you should see if you can translate into the language concerned using the morpheme concerned.

3.1.3.2 It is possible that the Field Guide should in fact be a TWIKI space where people discuss and add to information [but can't delete previous information...]. However, we may need to be even more prescriptive – requiring them to scroll down the description and examples before they can enter the term 'perfective'.

3.1.4 How can we present the information to users in a helpful way?

3.1.4.1 We probably can't separate GOLD as an interface from GOLD as the underlying knowledge. It is difficult to distinguish concept from term-set. But perhaps it would help people to approach it in a visually different way – opening out rather than drilling down. There are other types of display tools which are designed to display network graphs in effective ways, e.g. library catalogue lists displayed in network displays – dynamic and done in XML.

3.1.5 GOLD is an ontology but it has been presented to us as a hierarchy.

3.1.5.1 This has presented problems because we can't see the cross-cutting property relations, even though some of them have been built in. It is not a strict taxonomy, because the principles of differentiation do not stay the same not only across levels but even within levels.

3.1.5.2 This relates to a problem between universal and language – particular instantiations? For example, we might have definiteness as a feature in the morphosyntactic feature ontology. But in English we don't have a feature definiteness. We have definite articles and indefinite articles as parts of speech. in the GOLD ontology

3.1.6 If the ontology is to work universally how do we reconcile local differences of classification, and do we build them in?

3.1.6.1 We didn't solve this, but came up with several examples.

3.1.6.1.1 For example, the ontology includes 'definite article' and 'indefinite article' as parts of speech. But definiteness must be a morphosyntactic feature for those languages that take agreement in

definiteness. So why in the part of the speech class don't we have just 'article', and allow articles to be marked for the features +/- definite?

3.1.6.1.2 Do we interpret the English auxiliary WILL as future and ED as past – or do we interpret WILL as part of auxiliaries along with modals?

3.1.7 Should we build in implicational universals such as those produced by typologists?

3.1.7.1 We probably should not include putative universals, since they can be falsified by investigating more languages. Ontologists don't make claims about the world, and if you take that seriously you should not have putative universals in systems. You should have only concepts that build them. But the concepts we need should be those manifested in the world – we don't want to have 700 cases that don't exist anywhere we build on our knowledge of the world.

3.1.8 What do we do if one morpheme has two distinct meanings? E.g. English 'with' can be glossed as 'comitative', or 'instrumental'.

3.1.8.1 There are several suggestions

3.1.8.1.1 Gloss it with an overarching category, CASE

3.1.8.1.2 What happens if you have two categories with little in common, e.g. NOMINATIVE and PAST TENSE? In that case, have two homonymous morphemes 'with=INST', 'with=COM'. Use the lexicons, not GOLD to search for answers to questions such as "what languages have a single morpheme for 'comitative' and 'instrumental'.

3.1.8.1.3 Maybe you should say something like benefactive has as its primary or basic use the benefactive situation. TDS has to deal with it – and has taken a prototype stance – daughter nodes may or may not have the same set of features as the prototype.

3.1.8.2 Polyfunctionality/polysemy/homonymy is a serious problem. In GOLD you can't have an underspecified class – you can have an overspecified class. You can have INST+COM, this maps to both verb and noun but not to an underspecified thing, which maps to both. But what sense does it make to say it is both an adjective and a noun? Many people wouldn't buy it. But perhaps they will if they realise that they would want both to be brought up by a search.

3.1.8.3 We need instructions which say that something can belong to more than one category simultaneously.

3.1.9 What are the links between 2-3-5: (2) Concepts from Morphological Unit (from Core); (3) Concepts from Syntactic Unit; (5) Concepts from Part Of Speech [derived from Syntactic Unit]

3.1.9.1 Is the structure which has Morphological unit and Syntactic unit as sisters and 'part of speech' correct?

3.1.9.2 The ontology as presented to us is a taxonomy, and does not include property relations. 'Part of speech' is a grammatical unit, which may be realised as a morphological unit or as a syntactic unit.

3.2 Concepts from Morphological Unit (from Core)

3.2.1 Is it appropriate to have a taxonomy which includes form (clitic, bound root) and meaning/function (inflectional and derivational units)?

3.2.1.1 This is a hierarchy, not a taxonomy. We should consider whether some of these ideas, such as derivation and inflection are crosscutting properties.

3.2.2 Is it appropriate to have a taxonomy which includes proclitic and enclitic, but not prefix, infix and suffix?

3.2.2.1 Either mark the position (pro/en, pre/in/suf), or don't mark it.

3.2.3 If we include prefix, suffix, should we also include ablaut and reduplication and the patterns from Semitic roots?

3.2.3.1 There is a tension here between creating a terminological database and an ontology, and between using terms, which the user expects, and not using them, or using terms they don't expect.

3.2.4 Should we consider Semitic templatic roots simply as bound roots?

3.2.4.1 Question noted.

3.3 Concepts from PartOfSpeech

3.3.1 What is Part of Speech a class of?

3.3.1.1 A class of types, not tokens.

3.3.2 How do I say that there is no distinct class of Adjectives, but that the concepts expressed by Adjectives are expressed by Nouns?

3.3.2.1 We say that the word belongs to both classes. This creates a problem with precatatorial roots. The system requires classification of such forms as, say, both Noun and Verb.

3.3.3 Transitivity

3.3.3.1 Classifying verbs as intransitive, transitive and ditransitive raises several concerns. If these subclasses are to be kept, they need expanding to include semi-transitive and ambi-transitive, and perhaps stative etc. If they are to be kept, then we need to consider how to include the complement-taking possibilities of nouns, adjectives and adpositions.

3.3.3.2 What we have to have here is not a list of choices, but a set of relationships – to build a complex category expressing the subcategorization possibilities. This means distinguishing the number of arguments a predicate has from the way they are expressed by case, or as parts of speech or phrase, NPs, AdjPs, clauses, VPs etc, Adps.

3.3.4 Do we allow complex features, e.g. having names for commonly used collocations of features e.g. imperfect =imperfective+past –

3.3.4.1 Question noted

3.3.5 Concrete suggestion to Parts Of Speech add:

3.3.5.1 a subclass of cardinal directions, directionals

3.3.5.2 a subclass of conversational noises (turn-taking, back-channeling, floor-holding)

3.3.5.3 a subclass of ideophones (mimetic words)

3.4 TAM and polarity

3.4.1 Tense, mood, aspect, evidentiality, polarity – a continuum

3.4.1.1 There aren't really any discrete categories in the area of tense, mood, aspect etc.

3.4.1.2 An ideal ontology should reflect the non-discrete character of the categories

3.4.2 Difference between Force and Mood?

3.4.2.1 Force is on a different level (speech act theory)

3.4.2.2 Distinguish speech act type and sentence/clause type

3.4.2.3 Hard to understand differences between hierarchies (diagrams and numbering)

3.4.2.4 Some of the labels are hard to understand

3.4.3 Relate ontology to IGT!

3.4.3.1 IGTs don't have to be just one tier and

3.4.3.2 ...ontology should match up with the tiers (David Nathan)...

3.4.3.3 ...or at least, the tools should be compatible with the ontology

3.4.4. Do not list combinations with a compositional semantics!

3.4.4.1 You do not really need things like FutureInFuture, PastIn Present etc. which are really combinations of relative and absolute tenses

3.4.4.2 Difficult cases: the Pluperfect – a combination of Past and Perfect or an independent entity?

3.4.5 Don't hide the perfect!

3.4.5.1 Perfect and progressive are hidden under "phasal aspect" together with "inceptive", whatever that is

3.4.6 Mood is grammatical

3.4.6.1 Moods like Subjunctive, Portative etc. typically show up in certain types of subordinate clauses

3.4.6.2 ...are best described as parts of constructions

3.4.6.3 ...and should probably be united under the heading "Non-indicative moods"

3.4.7 Should irrealis be in?

3.4.7.1 Yes, say the Australianists and Austronesianists (Austrists?)

3.4.7.2 Maybe, say typologists – it is questionable whether irrealis has sufficient cross-linguistic coherence

3.4.8 Where should question particles go?

3.4.8.1 The position of a question particle may reflect the focus structure of a sentence

3.4.8.2 How are significant word order choices represented in an IGT?

3.4.9 Down-to-earth vs. Up-in-the-clouds categories

3.4.9.1 Should categories in the ontology be based on nitty-gritty empirical data or on lofty philosophical speculation?

3.4.10 Different kinds of futures: mixtures of tense, aspect, and modality

3.4.10.1 It'll rain vs. It's going to rain

3.4.10.2 I'll open the window vs. I'm going to open the window

3.4.10.3 Jag öppnar fönstret vs. Jag tänker öppna fönstret

3.4.10.4 Ja otkroju okno vs. Ja budu otkryvat' okno

3.4.11 An afterthought by Ö.D.:

3.4.11.1 In some sections of the ontology (notably Aspect and Voice) definitions have been culled from several different works with rather disparate terminologies and theoretical frameworks. The definitions have not been integrated with each other stylistically or content-wise, and terms are often left undefined. In other cases, definitions are from general encyclopedias or the like, while references to standard works in the particular area are lacking. Thus, Corbett's books on Gender and Number do not seem to have been consulted.

3. Working Group 4

4.1 General suggestions

4.1.1 Cross refs need to be supported

4.1.1.1 form/function, morphology/syntax...

4.1.1.2 featural distinctions can be more useful than a hierarchy

4.1.1.2.1 e.g. Case

4.1.2 Include prototypical lgs

4.1.3 Higher-level domains or multiple intersecting trees?

4.1.4 Historical Linguistics

4.1.4.1 Problem: the use of labels based on historical criteria rather than modern

4.1.4.2 Finnish: "Essive" is historical essive, but not used in the way that current linguists (and GOLD) define an essive

4.1.4.3 Georgian: historically had an ergative case, today this case is not ergative, but much more complex and essentially unlabelable.

4.1.5 Language variation

4.1.5.1 Dialectology

4.1.5.2 Sociolinguistics

4.1.5.2.1 Honorifics - Muna (Austronesian), Japanese, Uyghur (Turkic), Maithili (I-E) (Recommend making a separate class, cf. Gender, which can also show up on Vs, Adjs etc)

4.1.6 Pragmatics/Discourse (NOT discourse analysis)

4.1.6.1 Discourse Form (cf. 1.1.3) text unit

4.1.6.1.1 Cleft (not mentioned)

4.1.6.1.2 Prototypical languages: Japanese, Austronesian, NE Caucasian

4.1.6.1.3 Particles (x-ref 5.10 Particles)

4.1.6.1.4 Affixes

4.1.6.1.5 ...(form x, y, z,...) - leave these possible forms open

4.1.6.2 Discourse Function

4.1.6.2.1 Topic

4.1.6.2.1.1 Discourse Topic - need concept of topic to anchor references to topic in ontology 6.14.9, 6.14.18, 6.14.26, 6.14.28

4.1.6.2.1.2 Topic hierarchy

4.1.6.2.2 Focus (currently under 6.14.9, 6.14, 11) Focus
Antipassive Voice

4.2 Foundation Concepts from GOLD

4.2.1 [1.1.2] Syntactic Unit: {Metacomment: constitutes a hook on which one can hang cross-refs with the morphology}

4.2.2 [2.3.3.1] Endoclititics: A clitic that goes inside a word.

4.2.2.1 Reference: Harris, Alice. 2002. Where in the Word is the Udi Clitic? *Language*.

4.2.3 [3.1.1] Clause: (cf. also 5.4. for the morphology)

4.2.4 [3.1.1.1] Subordinate Clause:

4.2.4.1 Complement subordinate

4.2.4.2 Adjunct subordinate

4.2.4.3 Relative

4.2.4.3.1 Refs: R+R grammar: Coord, Subord, Cosubordination - See Foley and VanValin; Haspelmath, Martin and Ekkehard König, eds. 1995. *Converbs in a Cross-Linguistic Perspective: Structure and Meaning of Adverbial Verb forms*. Berlin: Mouton de Gruyter)

4.2.5 [5] Concepts from Part of Speech : Separate out the notion of connectivity from PoS. Affixes may mark connectivity (FORM/FUNCTION)

4.2.6 [5.1] Adjective:

4.2.6.1 [5.1.1] Plain

4.2.6.2 [5.1.2] Comparative

4.2.6.3 [5.1.3] Superlative

4.2.6.4 [5.1.4] Elative (comparative and superlative)

4.2.6.5 [5.1.5] Relative adjectives

4.2.6.5.1 Refs: Dixon 2004, Wetzler 1996, *The Typology of Adjectival Predication*. Berlin: Mouton.

4.2.6.5.2 (Cross-References: 5.1.5 rel adjectives to a higher level concept Relativity)

4.2.7 [5.4] Connective: (see 3.1 for the clause level)

4.2.8 [5.13.4.4] Could be put somewhere else but this is where we think

4.2.8.1 Sound symbolism

4.2.8.2 Ideophones

4.2.8.2.1 Refs: Angela Bartens, 2003. *Ideophones and Sound Symbolism in Atlantic Creoles*. Suomal... (Finnish Acad.scis)

4.2.8.2.2 Erhard Foeltz, Kilian Hatz Ideophones, Benjamins.

4.2.8.3 Onomatopoeia

4.2.9 [5.14] Verb: Need good superordinate categories.

4.2.9.1 Refs. Kittilä, Seppo. "Towards a typology of transitivity". PhD. U of Turku 2003; Beth Levin, Malka Rappaport-Hovav 1995. *Unaccusativity at the Syntactic-Lexical Semantics Interface*. (Linguistic Inquiry Monographs 26). Cambridge MA.: MIT Press.

4.2.10 [5.14.1] Transitivity

4.2.10.1 [5.14.1.1] Ditransitive Verb:

4.2.11 [5.14.2] IntransitiveVerb: Cognate objects would have to be dealt with by the individual linguist

4.2.11.1 An intransitive verb is a verb that does not [at the moment!] (cf. 'The children play' vs. 'The children play ball') take a direct object

4.2.11.2 [5.14.2.1] Unergative and Active intransitive

4.2.11.3 [5.14.2.2] Unaccusative (incl. stative and inchoative, but also 'float', 'melt' etc.) and Active intransitive

4.2.12 [5.14.4] Monotransitive

4.2.13 [5.14.2] ??Semantic Arguments

4.2.13.1 [5.14.2.1] Sense verbs,

4.2.13.2 [5.14.2.2] Action verbs....

4.2.13.3 Def: Arguments realized based on verbal arguments rather than syntactic configurations; semantically not syntactically driven.

4.2.13.4 Ref: Autotype, Hidden Syntax

4.2.14 [6.1.3] FrequentiveAspect: (xref with pluractional)

4.2.15 [6.2] CaseFeature:

4.2.15.1 We suggest using a *featural* system rather than the hierarchical one you have here

4.2.15.2 Many lgs, like Avar, shows that you sometimes have to combine the features of location and motion, which you currently have separate.

4.2.15.3 Liquid locative in Ingush (anything that can surround you, e.g. trees)

4.2.15.4 cf. Avar, Sami

	in (illative)	on (allative)	under (sub)	at, beside	in
X	-in	-on	-under	-at	-in
to	-in-to			
from	-in-from			etc	
to/from	-in-to/from				etc

4.2.15.5 cf. Hungarian: in something, on top of something, on top of something else*

4.2.15.6 Refs: Abondolo, Daniel. 1998. The Uralic Languages. Routledge. Alice Harris can provide refs for Caucasian lgs.

Grünthal, Riho. Finnic adpositions and cases in Change. Mem Soc. Finno-Ou. (Refs about case in balto-finnish in general / Veps)

4.2.15.7 If you have a hierarchy as below, also: Core:Oblique opposition should be supported higher up (or as a feature of the case system). The Core:Oblique opposition includes a different notion of "Core" than the Core:Loc/Mo:Other.

4.2.16 [6.2.1] CoreCase: {?Grammatical cases}

4.2.16.1 Category of case that includes Nominative, dative, instrumental, ergative absolutive.

4.2.17 [6.2.2] LocationalCase/ MotionCase: {We suggest combining these two and defining them in feature based terms when a language needs to distinguish between motions vs. locations}

4.2.18 [6.2.3] ? (SemanticCase): Instrumental, Genitive, Possessive, Comitative, Partitive*, (in the semantic sense. This is a core case in Finnic (marks the object) - which shows why Case should be described in terms of features. One way to deal with the Finnic case is to say that Finnic "partitive" is not a prototypical partitive, but rather combines (GOLD) features of partitive and Accusative.) Barry Blake, Case.

4.2.19 [Figure 1] Entailment relations among *CoreCase* feature values: (where?? Affective case (Georgian) - corresponds to Ru, Ger. locative (*mir gefällt das*) - "experiencer datives", stimulus

4.2.20 [6.2.27] EssiveCase: {Finnic essive notion is different.}

4.2.21 [6.2.42-59] {The following is also an area which needs to be x-ref'd to historical info.}

4.2.22 [Figure 2] Entailment relations among *Evidentiality* feature values (seems based on Willet, Thomas(?) 1988 article in *Studies in Language*)

4.2.22.1 cf. Aikhenvald 2003, also grammars of Dixon ...

4.2.22.2 currently based completely on Palmer (2001), yet challenged by several researchers, whether or not it is totally linked to modality

4.2.22.3 instead, it also relates to the source of information, and only secondarily modality

4.2.22.4 Orthogonal reference: Form + etymon is frequently related to the perfect.

4.2.22.5 (relegate to a COPE?)

4.2.23 [6.6] GenderFeature: we recommend indeed putting this under a higher class.

4.2.23.1 Human vs. Non-human systems,

4.2.23.2 systematic/semantic vs. arbitrary systems

4.2.23.3 (Corbett, Greville: "controller gender vs. target gender") Steve Wexler and Laura Szlotich, Aronoff M. 1994 *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press, Chapter ?2)

4.2.23.4 distinguishes declension clearly from gender

4.2.24 [6.6.2] FeminineGender: A grammatical class of nouns whose {members tend to be perceived of as female} includes

4.2.25 [6.6.4] MasculineGender: (Cross-ref: historical etyma {Masc. gender in IE is different from a "std" semantically-based masc. gender, see Corbett}) A grammatical class of most nouns {whose members tend to be perceived of as}referring to but not necessarily limited to males. OR : ...a gender to which males are typically assigned.

4.2.26 [6.6.5] Other: Any number of other genders (semantically or formally defined), including neuter, vegetable, women, fire, and dangerous things, etc.

4.2.26.1 {Need to separate out the relevant notions so that linguists can combine the ones that they need}

4.2.26.2 (Refs: Corbett, Greville. 1991. *Gender* Foley 1991. *Yimas grammar*. Stanford)

4.2.26.3 Dobrin, Lise. 1998, The morphosyntactic reality of phonological form. *Yearbook of Morphology 1997*, ed. by Geert Booij and Jaap van Marle, 59-81. Dordrecht, Foris.

2003. Kibrik, Alexander. *Nominal Inflection Galore* EUROTYP ed. by Franz Plank.

4.2.27 [6.9] NumberFeature:

4.2.27.1 (Add Corbett, Greville G. *Number*.)

4.2.27.2 Need to separate number out from nouns

4.2.27.3 Need to include Pluractionals (similar to Frequentative, particularly in Africanist literature)

4.2.27.4 Prototypical Languages: Chechen, Ainu, Svan, Yurok, Bole

4.2.27.5 Cusic, David D. 1981 The verbal plurality of Aspect.

Shuh, Russell. ...Chadic. Wood, Esther. In progress diss , a typological survey of pluractionals. ??ejwood@berkeley.edu or ask Jeff Good.

4.2.28 [6.10] PersonFeature: (cross-ref. to honorifics)

4.2.28.1 Refs: Siewierska 2004, Person. CUP.

4.2.29 [6.10.6] ThirdPerson: (needs to include, somewhere, the category of Impersonal:

4.2.29.1 NPs To deal with impersonal 3p reference (Fr. *on*) - usage issue.

4.2.29.2 Vs: -n-ta constructions Kibort, Anna. In-process dissertation. *Passive and Passive-like constructions in English and Polish*. online)

4.2.29.3 Langs: Ancient Egyptian

4. Working Group 6

5.1 Introduction

5.1.1 Defining the highest levels of an ontology has consequences throughout the entire system, so it includes often the hardest decisions. Consequently, the discussion of the group revolved about difficult questions concerning definitions of elementary concepts such as sign, feature, form, function, structure. Another topic was the tradeoff between loosely defined categories everyone seems to be happy with at first glance, but nobody can rely on at second, and rigidly defined categories that set sharp cuts but next to everyone has some complaint about. But there was some agreement on areas where lack of terminological agreement is especially visible and where consequently ontological clarifications would be especially welcome and needed, e.g., to further discussion across the boundaries of different linguistic schools.

5.2 Discussion

5.2.1 There was agreement that linguistic signs are abstract entities with (necessarily so) property structure (different aspects) and (possibly) meronomic structure (different parts such as morphemes, words, phrases). The aspects of a given sign can be partitioned according to GOLD into

5.2.1.1 form (phonological or orthographic aspect),

5.2.1.2 content (semantic and pragmatic aspect), and

5.2.1.3 structure (morphological and syntactic aspect).

5.2.2 It has been pointed out that of these three terms only content is more or less unambiguous where both form and structure have various confusingly different

uses in linguistics. When linguists talk about the form-function distinction, e.g., what they mean is most of the time what is called the structure-content distinction in the terminology used above. Similarly, since there is both phonological structure and semanto-pragmatic structure it has to be emphasized that the notion of structure meant above includes only grammatical structure.

5.2.3 Linguistic signs as abstract entities can be instantiated (or realized) by sign tokens in which their form has a concrete realization as a phonetic event (in the case of a sign of oral language; for other modalities correspondingly). The question of the realizations of content and structure in sign tokens has not been addressed.

5.2.4 A considerable part of the discussion was devoted to spelling out the distinction between content and structure (alias form and function). Bill Croft warned about blurring the border. A pronoun with the value female of the structural feature gender has this feature independent of its use in an exophoric pronoun that refers to a person with female sex (here the morphological aspect codetermines the pragmatic – in the sense of referential – aspect of that sign) or as an anaphor referring back to a noun with female gender (here the morphological aspect of this sign is in line with the morphological aspect of another sign and is independent of the pragmatic aspect).

5.2.5 Structural features are in general labeled according to one of the semantic and pragmatic uses (ideally their default use) their possessors (signs with these features) have. The group agreed that it is hard to find structural features that are defined without recourse to content features (augments may be a case in point; clear instances are positional features such as: *sentence initial position*).

5.2.6 Among the questions that remained open was if the wording used above is correct (it is the signs that bear the features, e.g., the English sign *she* bears the feature 'female') or if it would be more precise to say that not the signs but their aspects bear features (the structural aspect of, e.g., the English sign *she* bears the feature 'female'). Both views become compatible if features are uniquely correlated with aspects (which seems plausible), because then one can define 'X has feature Z' just as shorthand for 'aspect Y of sign X has feature Z', where Y is the unique aspect Z belongs to.