Toward a standard list of grammatical comparative concepts:
The Grammaticon

Martin Haspelmath
(MPI-SHH Jena & Leipzig University)

1. Comparisons of typological databases

Even though it has been noted that comparative concepts for typology are merely instruments for research and may therefore differ across researchers (Haspelmath 2010), we want different databases to be comparable.

For example, we would like to compare data from *WALS* (Haspelmath et al. 2005/2013) to be comparable with data from *APiCS* (Michaelis et al. 2013), *SAILS* Online (Muysken et al. 2016), and *SSWL* (Owusu et al. 2013).

2. Comparisons of lexical databases: the Concepticon

The task is similar to the task of lexical comparison across languages by means of a set of comparison meanings.

For lexical databases, a standard ontology now exists: The **Concepticon** (List et al. 2016, **concepticon.clld.org**), which has over 2500 comparison meanings that bring together lexical concepts from diverse lexical data collections. This allows quick and automatic comparison of lexical forms from diverse databases.

The concepticon includes lexical concepts from 161 concept lists: Swadesh list, IDS list, SIL-Africa list, Sutton & Walsh Australian list, and so on.

Altogether, there are 30,716 lexical concepts, grouped together into 2,632 concept sets or metaconcepts.

Blust-2008-210-166	correct, true [english]
Borin-2015-1532-120	right(2) [english]
Bowern-2007-207-26	correct [english]
Bowern-2011-204-35	correct/true [english]
Bowern-2012-189-34	*correct/true [english]
Bowern-2012-205-35	*correct/true [english]
Buck-1949-1108-894	right (adj., vs. wrong) [english]
Bulakh-2013-870-184	correct (adj.) [english]
Chen-1996-100b-55	对 [chinese]

3. A counterpart of of Concepticon: the Grammaticon

We propose to set up a counterpart of the Concepticon for grammatical patterns, called Grammaticon, which will facilitate the comparison of different grammatical datasets.

Analogous to the lexical comparison meanings in the Concepticon

(the **concept sets**), the Grammaticon contains **metafeatures** which capture what is common in highly similar features of different databases.

e.g.
WALS: Order of Subject, Object and Verb: SVO
APiCS: Order of subject, object and verb: Subject-verb-object (SVO)
SAILS: The dominant constituent order in a transitive clause is: AVP
DiACL: What is the canonical (neutral) word order in a main clause? SVO
SSWL: Property 05_SVO

metafeature:
               dominant order in transitive clauses is A-V-P

e.g.
WALS: Inclusive/exclusive distinction in independent pronouns:
               No inclusive/exclusive
APiCS: Inclusive/exclusive distinction in independent personal
               pronouns: No inclusive/exclusive distinction
SAILS: Is there an inclusive/exclusive distinction in personal pronouns?
               – no
SAILS: Is there a distinction between inclusive and exclusive for
               personal pronouns? – no

metafeature:
               no clusivity distinction in independent personal pronouns

e.g.
Grambank: Can the recipient in a ditransitive construction be marked
               like the monotransitive patient? NO (0)
WALS: Ditransitive Constructions: The Verb 'Give':
               Indirect-object construction

metafeature:
               **the R-argument is not even partially aligned in coding
               with the P-argument**

4. Comparing Concepticon, Grammaticon, and Phoible

	Concepticon	Grammaticon	Phoible
language-specific entity	word	construction	phoneme
project-specific comparative concept	project-specific lexical meaning	project-specific grammatical feature	segment type of contributors (UPSID, SPA, etc.)
comparative meta-concept	(lexical) concept set	(grammatical) metafeature	Phoible segment type

5. Ontologies

The goal of allowing comparability, also for machines (i.e. interoperability),
is similar to that of the GOLD ontology (General Ontology for Linguistic Description,
Farrar & Langendoen 2003).

There was also a related attempt to come up with standard category labels, called
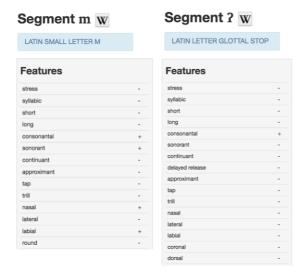ISOCat, but this seems to have been abandoned.

The GOLD ontology seems designed to be a general set of concepts for description,
but this is impossible, because each language has its own system.

If this is the goal of GOLD, it resembles the idea of Wierzbicka's and Goddard's
Natural Semantic Metalanguage (NSM), which can be used to describe any language,
but which has not proved very practical.

The Grammaticon's goal is more modest, in that it works entirely at the level of
comparative concepts. No claim is made that these concepts should be useful for the
description of particular languages.

6. Definitions of grammatical meatafeatures in terms of basic concepts

In Phoible, the segment types are defined by means of phonetic component features,
e.g.

Segment m W

LATIN SMALL LETTER M

Features

stress	-
syllabic	-
short	-
long	-
consonantal	+
sonorant	+
continuant	-
approximant	-
tap	-
trill	-
nasal	+
lateral	-
labial	+
round	-

Segment ʔ W

LATIN LETTER GLOTTAL STOP

Features

stress	-
syllabic	-
short	-
long	-
consonantal	+
sonorant	-
continuant	-
delayed release	-
approximant	-
tap	-
trill	-
nasal	-
lateral	-
labial	-
coronal	-
dorsal	-

Concept sets

Showing 1 to 100 of 2,632 entries

← Previous | 1

ID	Name	Definition
Search	Search	Search
0	<NA>	Set of all concepts not yet mapped to a meaningful concept set
1	CONTEMPTIBLE	Deserving of contempt or scorn.
2	DUST	Any kind of solid material divided in particles of very small size.
3	BRAVE	Having or characterized by courage.
4	COURTYARD	An area wholly or partly surrounded by walls or buildings.
5	GAZELLE	An antelope of the genus Gazella mostly native to Africa and capable of running at high speeds for long periods.
6	EARTHQUAKE	The violent shaking of the ground produced by deep seismic waves, originating from the epicentre.
7	GATHER	To heap up; to collect or gather (e.g. work, magazines, etc.).
8	CURSE	An appeal or prayer for evil or misfortune to befall someone or something.

Likewise, many of the Grammaticon's metafeatures are not immediately clear, so they should be accompanied by definitions of the main terms, e.g.

Metafeature: the R-argument is not even partially aligned in coding with the P-argument

Related concepts
- P-argument
- R-argument
- argument coding

Features

Showing 1 to 2 of 2 entries

← Previous | 1 | Next →

Name	Featurelist
Search	Search
Can the recipient in a ditransitive construction be marked like the monotransitive patient? NO (2)	Grambank
Ditransitive Constructions: The Verb 'Give': Indirect-object construction	WALS

Metafeature: no clusivity distinction in independent personal pronouns

Related concepts
- clusivity
- independent personal pronoun

Features

Showing 1 to 4 of 4 entries

← Previous | 1 | Next →

Name	Featurelist
Search	Search
Inclusive/exclusive distinction in independent personal pronouns: No inclusive/exclusive distinction	APiCS
Inclusive/exclusive distinction in independent pronouns: No inclusive/exclusive	WALS
Is there a distinction between inclusive and exclusive for personal pronouns? no	SAILS
Is there an inclusive/exclusive distinction in personal pronouns? no	SAILS

Concept: P-argument

A P-argument is the patient argument of a physical-effect verb ('kill', 'break').

Related metafeatures:

- P-argument nominals are flagged only if they are definite and animate
- P-argument nominals are flagged only if they are animate
- P-argument nominals are flagged only if they are definite
- P-argument nominals are always flagged
- verb indexes exist neither for A nor for P
- verb indexes exist both for A and for P
- verb indexes exist only for P
- the R-argument is not even partially aligned in coding with the P-argument
- the R-argument is at least partially aligned in coding with the P-argument

Defining concepts:

- argument

Derived concepts:

- accusative flag
- applicative marker
- core argument
- object

These definitions often involve the task of **retro-defining** the concepts used by the database designers, because they are not always clearly defined.

Such retro-definitions are not unproblematic, because we do not always know well which decisions the original database designers would have chosen.

But the task of retro-defining is similar with lexical databases, so we are not faced with a new kind of problem.

These basic concepts could eventually develop into something like "an IPA for comparative grammar research".

Note that this is a very different philosophy from the widespread idea that terms like "head", "antipassive", "perfect", "polysynthesis" or "clitic" are inherently fuzzy. They are merely as fuzzy as we want them to be.

The reality of languages is certainly continuous, both on grammar and in phonetics. But the best-known comparative concepts of segmental phonetics (the IPA symbols) are not fuzzy, and likewise the best-known grammatical terms could be non-fuzzy.

7. Conclusion

I conclude that new tools for lexical comparison (i.e. Concepticon) and for phonetic-segment comparison (i.e. Phoible) have shown us a way how to lift grammatical comparison to a new level.

Comparability can be achieved through the "Grammaticon", a list of metafeatures that subsume features of existing and partly comparable grammatical databases.

This task is more difficult, and some features of existing databases have been recalcitrant. But the great majority of existing features can be incorporated in one way or another into Grambank.

We hope that we will have the resources to do this over the next few years.

References

Farrar, Scott & D. Terence Langendoen. 2003. A linguistic ontology for the Semantic Web. *GLOT International* 7(3). 97-100. (Available at http://www.linguistics-ontology.org/gold.html)

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Haspelmath, Martin, Matthew Dryer, David Gil & Bernard Comrie (eds.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press. (2013 online version available at http://wals.info)

Kortmann, Bernd & Lunkenheimer, Kerstin (eds.) 2013. *The Electronic World Atlas of Varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available at http://ewave-atlas.org)

List, Johann-Mattis & Cysouw, Michael & Forkel, Robert (eds.) 2016. *Concepticon*. Jena: Max Planck Institute for the Science of Human History.

Muysken, Pieter, Harald Hammarström, Olga Krasnoukhova, Neele Müller, Joshua Birchall, Simon van de Kerke, Loretta O'Connor, Swintha Danielsen, Rik van Gijn & George Saad. 2016. *South American Indigenous Language Structures (SAILS) Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available at http://sails.clld.org)