

Uppsala University, 1 March 2022

Colexification, syllexification, and related things: Comparative concepts and possible explanations

MARTIN HASPELMATH

Max Planck Institute for Evolutionary Anthropology

I. Overview

Two kinds of deviations from one-to-one correspondence
between meanings and shapes:

(1) *colexification*

| | | |
|-------------|---------------|---|
| e.g. German | <i>Tasche</i> | 'bag; pocket' |
| English | <i>go</i> | 'gehen (go by foot); <i>fahren</i> (go by vehicle)' |

(2) *cogrammification*

| | | |
|-------------|----------------------|---|
| e.g. German | <i>ich singe</i> | 'I am singing; I sing' (PROGRESSIVE; HABITUAL) |
| English | <i>to Washington</i> | 'nach Washington (ALLATIVE); dem Washington (DATIVE)' |

(3) *syllexification*

| | | | |
|-------------|---------------|--------------------|------------------------------------|
| e.g. German | <i>Onkel</i> | 'mother's brother' | (cf. Swedish <i>mor-bror</i>) |
| English | <i>kitten</i> | 'young cat' | (cf. German <i>Katzen-junges</i>) |

(4) *syngammification*

| | | | |
|------------|----------------------|--|---------------------------|
| e.g. Latin | <i>libr-orum</i> | 'of book-s' | ("cumulative inflection") |
| French | <i>décriv-ai-ent</i> | 'they were describing' (-ai = imperfective + past') | |
| | | cf. Russian <i>opis-yva-li</i> [describe-IMPF-PST-3PL] | |

Three goals of this talk:

- present a systematic overview of the phenomena and the earlier terminology
- suggest some novel terminology that is reasonably transparent
- ask how limits on colexification and syllexification might be explained

2. Coexpression

2.1. Key terminology

Definitions of *colexification* and *cogrammification*:

- (5) **colexification** (of two meanings A and B):
= expression of either A or B by a root (= a minimal lexical form; Haspelmath 2020)
- (6) **cogrammification** (of two meanings A and B):
= expression of either A or B by a grammatical marker
- (7) **coexpression** (of two meanings A and B):
= expression of either A or B by a form or construction

| | |
|---------------------------|-------------------------------------|
| <i>colexification</i> : | coined by François (2008) |
| <i>coexpression</i> : | coined by Hartmann et al. (2014) |
| <i>cogrammification</i> : | coined on 1st March 2022 in Uppsala |

- (8) **coexpression diagram**
= a graphic representation of possible coexpression types

Coexpression diagrams are widely known as “semantic maps”
(e.g. Georgakopoulos & Polis 2018).

2.2. What is coexpressed: COMPARISON MEANINGS

It is often said that the semantic map model is neutral between polysemy and indeterminacy (monosemy), and also between polysemy and homonymy.

This neutrality is possible because the meanings which are coexpressed are not language-particular meanings – they are **comparison meanings**, i.e. a kind of comparative concept.

When we say that German *Tasche* colexifies ‘bag’ and ‘pocket’, we do not imply that it has two different meanings. We simply say that it can be used for ‘bag’ and for ‘pocket’, without implying anything about polysemy or vagueness.

Semantic maps are often said to capture “polysemy patterns”, but this is not accurate on the usual understanding of the term *polysemy* (= the presence of several related senses)

The literature also talks about **multifunctionality**, where the vague term *function* is perhaps taken as standing for a comparison meaning. (Some authors have also used the term *use* as a noncommittal term, e.g. Bybee et al. 1994: 44.)

An alternative to *semantic map* is **conceptual space** (Croft 2001), which implies that the meanings that are lexified differently in different languages are really universal **concepts**. This makes good sense because **concepts** are usually thought of as language-independent, clearly contrasting with language-particular **signata**.

But semantic maps can be constructed on the basis of specific sentences, e.g. in a questionnaire-based study (such as Dahl 1985) or in a parallel-text-based study (such as Dahl & Wälchli 2016). Concepts are general meanings independent of a particular context, but coexpression patterns can be studied in a context-dependent way, too – so the vague term **comparison meaning** seems more appropriate than “comparison concept”.

2.3. Some earlier terminology

The term **polysemy** goes back to Bréal (1897), though according to Nerlich (1990: 88), Bréal may have been influenced by Max Müller (1864):

“In the mythic period, thought and language were based on two tendencies: *polyonymy* and *homonymy*. **Polyonymy** consists in the fact that one and the same object which is perceived in various ways receives many names. **Homonymy** arises when objects perceived as different by the human mind nevertheless receive the same name” (F.M. Müller, *The Science of Language*, vol. II, Longmans, Green, and Co., London, p. 453-454).

Another common term is *categorization* – linguists say that different languages **categorize** particular domains (e.g. the human body, kinship relations, perception) differently by their lexical items (e.g. Koptjevskaja-Tamm et al. 2015: 434).

And linguists have often said that meanings are **lexicalized** in different ways in different languages.

For grammatical markers, the term **syncretism** has been used for quite some time in a synchronic sense (e.g. Plank 1991).

The term **colexification** spread quickly after it was proposed by François (2008), perhaps also because it was adopted by CLICS in 2014 (clics.cld.org; List et al. 2018; Rzymiski et al. 2020).

2.4. Some further terminology

- Lexical form F **colexifies** meanings A and B
- Form F **coexpresses** meanings A and B
- Language L **dislexifies** meanings A and B (François 2022)
= there are two different lexical forms (G and H) for A and B
- **partial colexification**, e.g.

| | | |
|--------|---------------------|----------|
| German | <i>Hand-tuch</i> | ‘towel’ |
| | <i>Leichen-tuch</i> | ‘shroud’ |
| | <i>Bett-tuch</i> | ‘sheet’ |

3. Synexpression

3.1. Key terminology

Definitions of syllexification and syngrammification:

- (8) **syngrammification** (of two meanings A and B):
= expression of both A and B in a grammatical marker (“synthetically”)

e.g. Latin *libr-orum*
book-PL.GEN
‘of book-s (plural + genitive)’

- (9) **sylllexification** (of two meanings A and B):
= expression of both A and B in a root (= a minimal lexical form)

e.g. English *stallion*
‘male + horse’

e.g. English *worse*
‘more + bad’

- (10) **synexpression** (of two meanings A and B):
= expression of both A and B in a form or construction

(***synexpression diagram**)

3.2. What is synexpressed: Comparison meanings

As in coexpression, when we say that a form **synexpresses** several meanings, we do not claim that these meanings should exist in the language in question – again, we are making statements with respect to comparison meanings.

In lexical semantics, there are probably many cases of synexpression of meanings that are not easy to render in a non-synexpressed way, e.g.

float = move on the surface of a liquid
(the language does not need to have a general word ‘liquid’)

Syngrammification is generally taken to imply that the meanings exist independently, e.g.

case + number Latin *libr-orum* ‘of books’

tense + aspect French *décriv-ai-ent*
(though imperfective aspect is never expressed independently of tense in French)

tense + person Latin *vid-i/vid-isti/vid-istis/vid-erunt*
‘I saw/you.SG saw/you.PL saw/they saw’

But in the case of person and number, it is actually not clear which meanings are syngammified – there are a variety of proposals, e.g.

| | | |
|-----|----------------|----------------|
| I | [+spkr, -addr] | [+auth, +part] |
| you | [-spkr, +addr] | [-auth, +part] |
| she | [-spkr, -addr] | [-auth, -part] |

In grammatical systems, we hope to decompose the forms into **elementary** meaning distinctions; this is less often done in lexical semantics, but when it is proposed, there are big questions around which are the elementary meanings, e.g.

kill = 'cause to die' (??)
'cause directly to die' (??)

3.3. Some earlier terminology

- fused expression
- portmanteau morph (Hockett 1947)
- cumulative expression / cumulative exponence (Matthews 1972)
(vs. *separative / separatist*; Plank 1999; Bickel & Nichols 2007)
- unmotivated vs. motivated expression (Ullmann 1957)
- polyexponence / coexponence (Bickel & Nichols 2005)
- conflation (Talmy 1985; 2000)
- presyntactic feature bundling (Distributed Morphology, e.g. Matushansky 2006)
- “fused suppletion” (e.g. *bad/worse*; contrasting with *good/bett-er*)
- synthetic vs. analytic lexicalization (Ježek 2016: 7-8)
- descriptive vs. labeling techniques for naming objects and events
(Seiler 1975; e.g. German *Lehr-er* ‘teacher’ vs. *Arzt* ‘doctor’)

3.4. Some further terminology

- Lexical form F **sylllexifies** meanings A and B
(e.g. English *mare* sylllexifies ‘horse’ and ‘female’)
- Form F **synexpresses** meanings A and B
(e.g. French *mieux* synexpresses ‘well’ + ‘comparative’)
- Language L **perilexifies** meanings A and B
= there are two cooccurring lexical forms (F and G) corresponding A and B

| | | |
|------------------------|----------|--------------------------------------|
| English <i>run</i> : | combines | motion + manner |
| German <i>reiten</i> : | | motion + instrument (,ride a horse’) |
| English <i>limp</i> : | | motion + manner + instrument |

Ježek (2016: 7-8)

| | | |
|---------------------------------------|-----|---------------------------------------|
| “synthetic” Italian (syllexifying) | vs. | “analytic English” (perilexifying) |
| <i>cenare</i> | | <i>have dinner</i> |
| <i>tardare</i> | | <i>be late</i> |
| <i>addormentarsi</i> | | <i>fall asleep</i> |

4. What explains the limits on coexpression and synexpression?

4.1. Coexpression: Colexification and cogrammification

A. Conceptual closeness (similarity) may explain coexpression

Haiman’s Isomorphism Hypothesis:

“Different forms will always entail a difference in communicative function. Conversely, recurrent identity of form between different grammatical categories will always reflect some perceived similarity in communicative function” (Haiman 1985: 19).

Croft (2001; 2003): conceptual spaces give us access to...

“the geography of the human mind, which can be read in the facts of the world’s languages in a way that the most advanced brain scanning techniques cannot even offer us” (Croft 2001: 364)

B. Likelihood of semantic extension explains coexpression

Cristofaro (2010):

coexpression is explained by tendencies of language change
(i.e. this is a mutational explanation; Haspelmath 2019)

C. Coexpression is constrained by clarity

if the range of meanings becomes too large, coexpression becomes rare (?)

e.g. König & Siemund (1999), on coexpression of

self-intensifier – reflexive – anticausative

Claim: only **self-intensifier & reflexive**, OR **reflexive & causative**,
but not all three!

4.2. Synexpression: Syllexification and syngammification

A. Periexpression is characteristic of certain language types

- e.g. Ullmann (1953; 1957) French seems to have more synexpression than German
 Seiler (1975): Cahuilla (Uto-Aztecán) has more periexpression than English

French tends to have “unmotivated” words,
 while German has more motivated words:

| | | |
|---------|--------------|-----------|
| French | German | |
| dé | Fingerhut | ‘thimble’ |
| gant | Handschuh | ‘glove’ |
| patin | Schlittschuh | ‘skate’ |
| entrer | hineingehen | ‘enter’ |
| divorce | Scheidung | ‘divorce’ |

B. High absolute frequency explains synexpression, low frequency explains periexpression

- | | | |
|--------------------------|---------------------------|--|
| e.g. kinship terms | <i>padre vs. madre</i> | <i>herman-o vs. herman-a</i> (Spanish) |
| e.g. male-female animals | <i>dog vs. bitch</i> | <i>lion vs. lion-ess</i> |
| e.g. comparatives | <i>bad vs. worse</i> | <i>expensive vs. more expensive</i> |
| e.g. quality nouns | <i>big vs. size</i> | <i>narrow vs. narrow-ness</i> |
| e.g. ordinal numerals | <i>one vs. first</i> | <i>seven vs. seven-th</i> |
| e.g. person & number | Russian <i>ty vs. vas</i> | <i>on vs. oni</i> (you.SG/you.PL, ‘he/they’) |
| e.g. number & case | Latin <i>ego vs. me</i> | <i>nos vs. nos</i> (‘I/me’, ‘we/us’) |

Moreover, it seems clear that different cultural preferences lead to different syllexification tendencies in particular domains, e.g. rich kinship terms in languages which use kinship terms frequently:

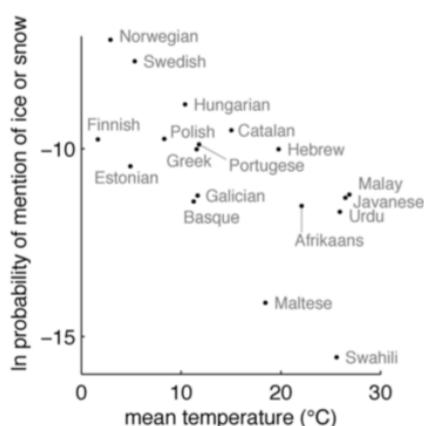
Evans (2011):

| | ♂ referent | | ♀ referent | |
|-----------|------------|-----------|------------|-----------|
| | ♂ speaker | ♀ speaker | ♂ speaker | ♀ speaker |
| (elder) | 1 | 3 | 5 | 7 |
| (younger) | 2 | 4 | 6 | 8 |
| (elder) | brother | | sister | |
| (younger) | | | | |
| (elder) | kagak | | | |
| (younger) | adik | | | |
| (elder) | ani | | otōto | |
| (younger) | ane | | imōto | |
| (elder) | thabuju | kularrind | kularrind | yakukathu |
| (younger) | duujind | kularrind | kularrind | duujind |

Or rich terms for frozen water in languages spoken at different latitudes, in areas with different average temperatures (Regier et al. 2016):



Speakers of languages like Swedish actually do speak more about snow or ice than speakers of languages like Maltese, at least on Twitter:



A final observation

If a language makes **more fine-grained distinctions** with its morphs than another one (e.g. English *snow* vs. *ice*, French *entrer* 'go in' vs. *sortir* 'go out', Japanese *otōto* 'elder sister' vs. *imōto* 'younger sister'),

then often this also means that there is **less colexification** (and more dislexification) – and if syllexification (= making fine-grained distinctions) is due (in part) to frequency of use, then **frequency is relevant also to colexification**.

Thus, frequency of use seems to be important not only for asymmetric coding ("markedness"; see Haspelmath 2021), but also for understanding coexpression and synexpression patterns.

References

- Bickel, Balthasar & Nichols, Johanna. 2005. Inflectional synthesis of the verb. In Martin Haspelmath, David Gil & Bernard Comrie, Matthew S. Dryer (ed.), *The world atlas of language structures*. Oxford: Oxford University Press. (<https://wals.info/chapter/22>)
- Bickel, Balthasar & Nichols, Johanna. 2007. Inflectional morphology. In Shopen, Timothy (ed.), *Language typology and syntactic description*, vol. III: Grammatical categories and the lexicon, 169–240. Cambridge: Cambridge University Press.
- Bybee, Joan L. & Perkins, Revere & Pagliuca, William. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: The University of Chicago Press.
- Cristofaro, Sonia. 2010. Semantic maps and mental representation. *Linguistic Discovery* 8(1). 35–52. (doi:10.1349/PSI.1537-0852.A.345)
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. 2nd edition. Cambridge: Cambridge University Press.
- Dahl, Östen & Wälchli, Bernhard. 2016. Perfects and iamitives: Two gram types in one grammatical space. *Letras de Hoje* 51(3). 325–348.
- Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.
- Evans, Nicholas. 2011. Semantic typology. In Song, Jae Jung (ed.), *The Oxford handbook of linguistic typology*, 504–533. Oxford: Oxford University Press.
- François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Vanhove, Martine (ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (Studies in Language Companion Series 106), 163–216. Amsterdam: Benjamins.
- François, Alexandre. 2022. Lexical tectonics: Mapping structural change in patterns of lexification. (*to appear*).
- Georgakopoulos, Thanasis & Polis, Stéphane. 2018. The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass* 12(2). e12270. (doi:10.1111/lnc3.12270)
- Haiman, John. 1985. *Natural syntax: Iconicity and erosion*. Cambridge: Cambridge University Press.
- Hartmann, Iren & Haspelmath, Martin & Cysouw, Michael. 2014. Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38(3). 463–484. (doi:<https://doi.org/10.1075/sl.38.3.02har>)
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and crosslinguistic comparison. In Tomasello, Michael (ed.), *The New Psychology of Language*, vol. 2, 211–243. New York: Lawrence Erlbaum. (10.5281/zenodo.831410)
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. (doi:10.1017/S0022226720000535)
- Hockett, Charles F. 1947. Problems of morphemic analysis. *Language* 23(4). 321–343. (doi:10.2307/410295)
- Ježek, Elisabetta. 2016. *The lexicon: An introduction*. Oxford: Oxford University Press.
- König, Ekkehard & Siemund, Peter. 1999. Intensifiers and reflexives: A typological perspective. In Frajzyngier, Zygmunt & Curl, Traci S. (eds.), *Reflexives: Forms and functions*, 41–74. Amsterdam: Benjamins.
- Koptjevskaja Tamm, Maria & Rakhilina, Ekaterina & Vanhove, Martine. 2015. The semantic of lexical typology. In Riemer, Nick (ed.), *The Routledge handbook of semantics*. London: Routledge.
- List, Johann-Mattis & Greenhill, Simon J. & Anderson, Cormac & Mayer, Thomas & Tresoldi, Tiago & Forkel, Robert. 2018. CLICS²: An improved database of cross-linguistic colexifications assembling lexical data with the help of cross-linguistic data formats. *Linguistic Typology* 22(2). 277–306. (doi:10.1515/lingty-2018-0010)

- Matthews, Peter H. 1972. *Inflectional morphology*. Cambridge: Cambridge University Press.
- Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37(1). 69–109. (doi:10.1162/002438906775321184)
- Nerlich, Brigitte. 2003. *Change in language: Whitney, Bréal, and Wegener*. London: Routledge.
- Plank, Frans. 1991. Of abundance and scantiness in inflection: A typological prelude. *Paradigms: The economy of inflection*, 1–40. Berlin: Mouton de Gruyter. (doi:10.1515/9783110889109.1)
- Regier, Terry & Carstensen, Alexandra & Kemp, Charles. 2016. Languages support efficient communication about the environment: Words for snow revisited. *PLOS ONE* 11(4). e0151138. (doi:10.1371/journal.pone.0151138)
- Rzymiski, Christoph & Tresoldi, Tiago & Greenhill, Simon J. & Wu, Mei-Shin & Schweikhard, Nathanael E. & Koptjevskaja-Tamm, Maria & Gast, Volker et al. 2020. The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data* 7(1). 13. (doi:10.1038/s41597-019-0341-x)
- Talmy, Leonard. 1985. Lexicalization patterns. In Shopen, Timothy (ed.), *Language typology and syntactic description (Volume III)*, 57–149. Cambridge: Cambridge University Press.
- Talmy, Leonard. 2000. *Toward a cognitive semantics*. Cambridge, Mass.: MIT Press.
- Ullmann, Stephen. 1953. Descriptive semantics and linguistic typology. *Word* 9(3). 225–240.
- Ullmann, Stephen. 1957. *The principles of semantics*. Glasgow: Jackson, Son & Co.