

Morphosemantics, constructions, algorithmic typology and parallel texts

Bernhard Wälchli
bernhard@ling.su.se
Stockholm University

Bern (CH), September 20, 2012
Berner Zirkel für Sprachwissenschaft
Universität Bern, Institut für Sprachwissenschaft

Abstract

Unlike morphology (the internal formal structure of words) and semantics (the study of the meaning of words and sentences), **morphosemantics** is concerned with the link between marker and meaning. Traditional approaches to morphosemantics such as semiotics and construction grammar argue that the relationship between *image acoustique* and *concept* is symbolic. This works well if the links are known (in the “proficiency mode”). In this talk I argue that there is a statistical alternative which is particularly useful if the links are not known (in the “**discovery mode**”). Meanings and markers form **collocations** in texts which can be measured by means of collocation measures. However, there is a considerable **non-isomorphism** between marker and meaning. As is well known a marker can have many different meanings (polysemy). Somewhat less well known is that a meaning is often expressed by many different markers, both paradigmatically and syntagmatically (**polymorphy**).

To make meanings and markers commensurable, they must be converted into units of the same kind. This same kind is the set of contexts in a text or corpus where a marker or meaning occurs. If the distribution of a meaning in a corpus is known, its corresponding **marker complex** can be determined which consists of a paradigmatically and syntagmatically ordered set of simple markers. The markers considered here are **surface markers** of two types: **word forms** and **morphs**

(continuous character strings within word forms). More abstract marker types such as lexemes, grammatical categories and word classes might often be better markers than surface markers, but they are not available in the discovery mode.

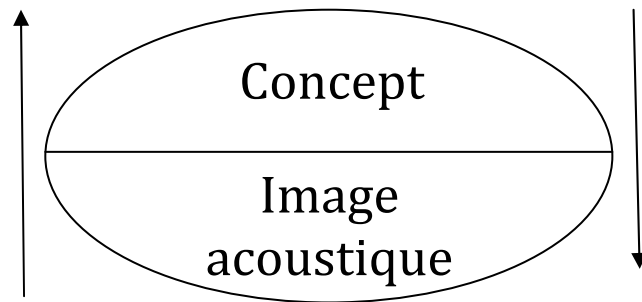
Marker complexes are a simple construction type. A procedural approach to construction grammar is adopted where marker complexes are viewed as an intermediate stage in a processing chain of increasingly more complex construction types from simple markers via marker complexes to syntactic constructions. Marker complexes have the advantage that they can be extracted automatically from massively **parallel texts**, i.e. translations of the same text into many languages, such as the New Testament used here. In parallel texts the same meanings (with certain restrictions) are expressed across different languages. This means that a **functional domain** can be defined as a set of contexts where a certain meaning occurs.

The same procedure is applied to cross-linguistically similar material and the procedure applied to cross-linguistic data is fully explicit and therefore replicable. It can be implemented in a computer program and run without the intervention of a typologist (**algorithmic typology**). The underlying idea is that the procedure of extraction is invariant (**procedural universal**) whereas the extracted structures can be highly variable depending on the texts and languages to which they are applied.

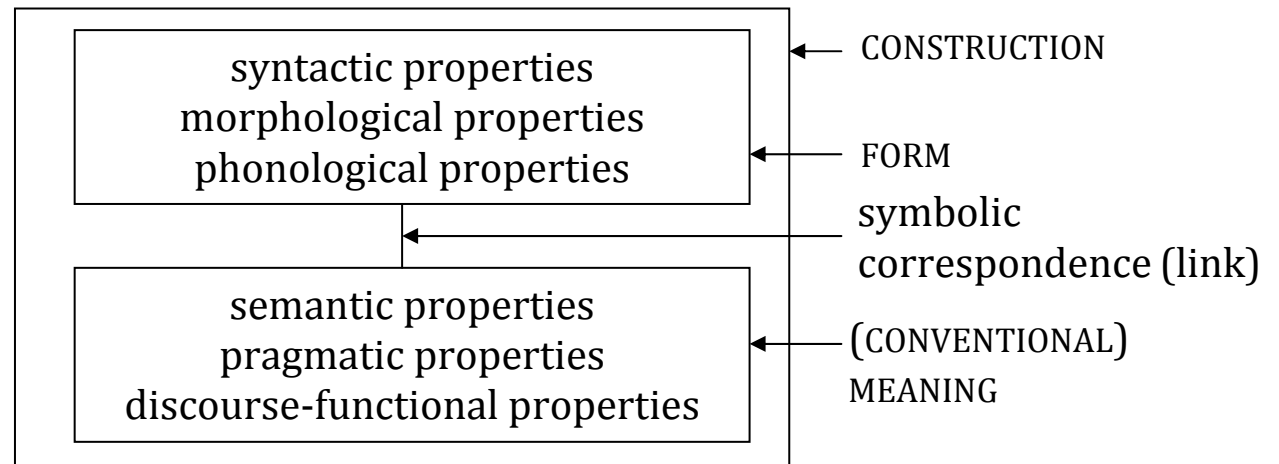
The talk considers to what extent surface markers are sufficient as input for the identification of constructions in a range of grammatical and lexical domains in a world-wide convenience sample of somewhat more than 50 languages. One of the domains considered in more detail is **comparison of inequality**. Comparison of inequality is expressed in most languages of the sample by an at least bipartite marker complex consisting of the parts standard marker ('than') and predicate intensifier ('more', '-er'). It will be argued here that both of them are intrinsic parts of the comparative construction. These findings are not fully in accordance with Leon Stassen's typology of comparison – a classical study in functional domain typology – which is based exclusively on the encoding of the standard NP. Other domains considered in the talk include negation, 'want', future, and predicative possession.

Traditional approaches to morphosemantics

De Saussure: Semiotics



Croft: Radical Construction Grammar



Morphology

the internal formal structure of words

Semantics

the study of the meaning of words and sentences

Morphosemantics

the link between marker and meaning

An alternative to symbolic links: collocations

“[a] collocation is an expression consisting of two or more words that correspond to some conventional way of saying things” (Manning & Schütze 1999: 151)

strong tea

powerful drug

Meaning and markers are different kinds of things. However, in texts they both manifest themselves as distributions. Distribution is the medium in which meaning can be turned into marker and vice versa.

Meanings and markers collocate

Collocation measures, e.g., t-score

$$T = \frac{\text{prob}(A, B) - \text{prob}(A) \times \text{prob}(B)}{\sqrt{\frac{1}{n} \times \text{prob}(A, B)}} = \frac{\frac{a}{n} - \frac{x}{n} \times \frac{y}{n}}{\sqrt{\frac{1}{n} \times \frac{a}{n}}}$$

Natural and non-natural meaning (Grice 1957, written 1948)

Natural meaning	Non-natural meaning
“Those spots mean measles”	“The lighted sign on the roof means that the cab is free”
Can be restated as “The fact that he has those spots means that he has measles”	Is compatible with “But it isn’t in fact free – the driver has made a mistake”.
<i>x means that p entails p</i>	<i>x does not mean that p entails p</i>

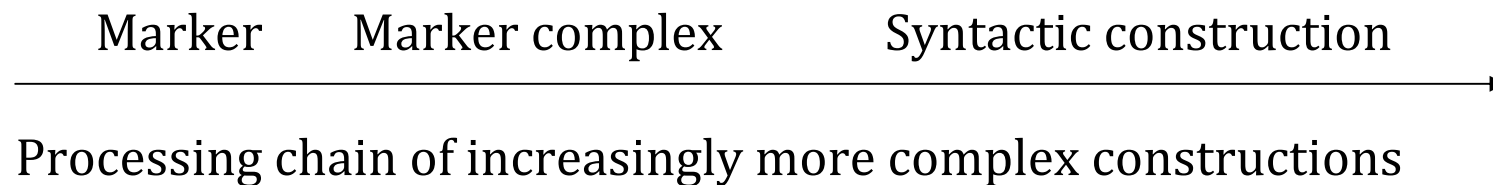
Collocations are natural meaning

Constructions in Construction Grammar

Goldberg (2006: 5) “[A]LL LEVELS OF GRAMMATICAL ANALYSIS INVOLVE CONSTRUCTIONS: LEARNED PAIRINGS OF FORM WITH SEMANTIC OR DISCOURSE FUNCTION, including morphemes or words, idioms, partly lexically filled and fully general phrasal patterns” (emphasis removed, BW)

morpheme:	e.g. <i>pre-, ing</i>
word	e.g. <i>avocado, anaconda, and</i>
idiom (partly filled)	e.g. <i>jog <someone's> memory, send <someone> to the cleaners</i>
Ditransitive	Subj V Obj ₁ Obj ₂ (e.g. <i>he gave her a fish taco, he baked her a muffin</i>)
Passive	Subj aux VPpp (PP _{by}) (e.g. <i>the armadillo was hit by a car</i>)

A dynamic approach to constructions



An onomasiological approach

	Meaning	Form
Onomasiological	given	wanted
Semasiological	wanted	given

In typology: functional domain

According to Miestamo (2005: 293) a functional domain is “any domain of related (semantic or pragmatic) functions that (one or more) language(s) encode with the formal means they possess”

Non-isomorphism of markers and meanings:

Polysemy and Polymorphy

a meaning is often expressed by many different markers

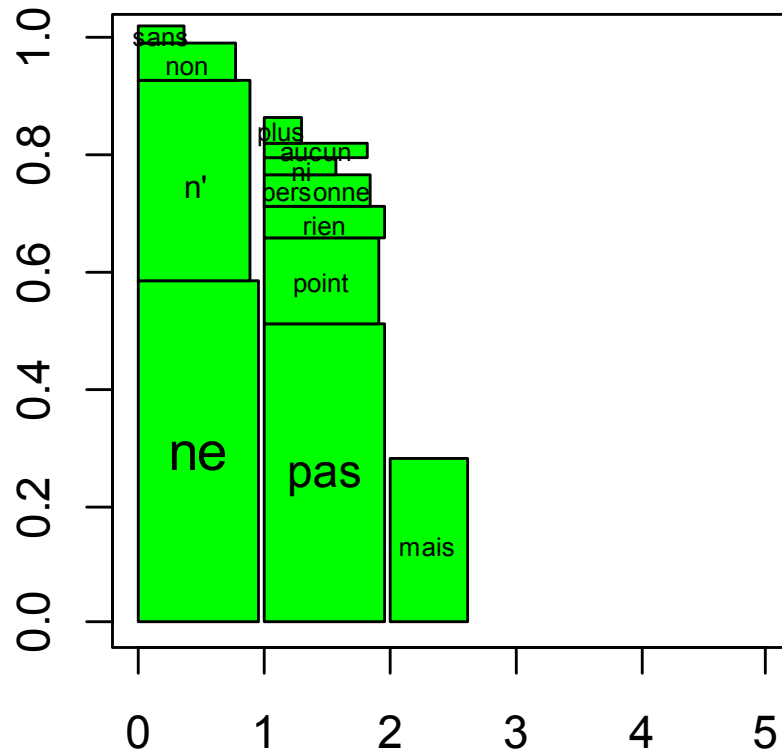
Spanish *querer* has more than one meaning ‘want’, ‘love/desire’

Negation (in formal semantics simply \neg) in French has more than one marker: *ne, pas, point, non, rien, sans* etc.

Syntagmatic and paradigmatic polymorphy

Marker complex: paradigmatically and syntagmatically ordered set of simple markers

Negation - French



Basic conventions

read bottom-up left-to-right

Slots: columns, left to right

Amplitude: vertical extension

Dedication: horizontal extension

Word form: green

Morph: yellow,

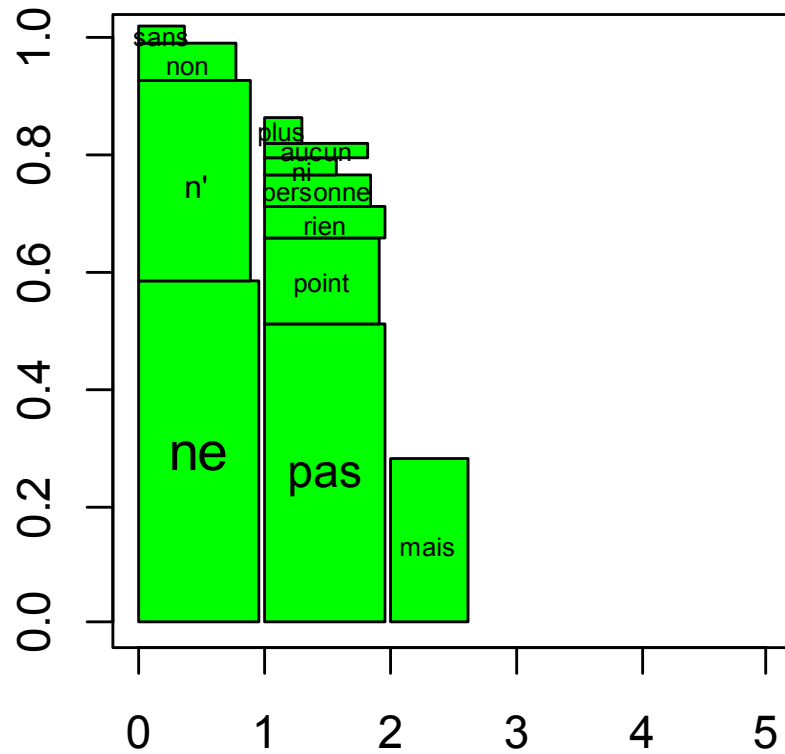
is word boundary

Word order: does not figure

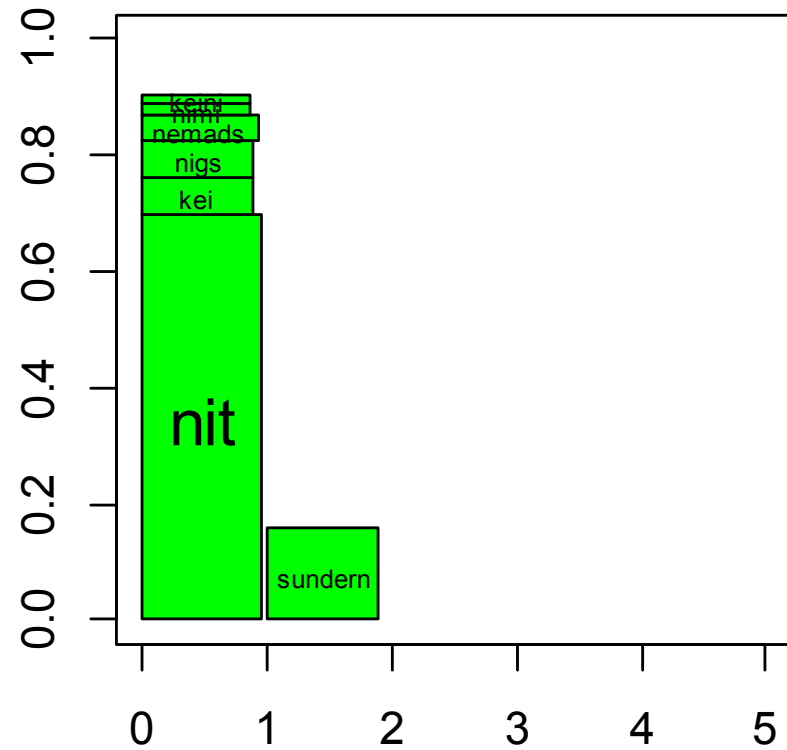
Syntax: does not figure

Lexemes, grams: do not figure

Negation - French



Negation - Alemannic

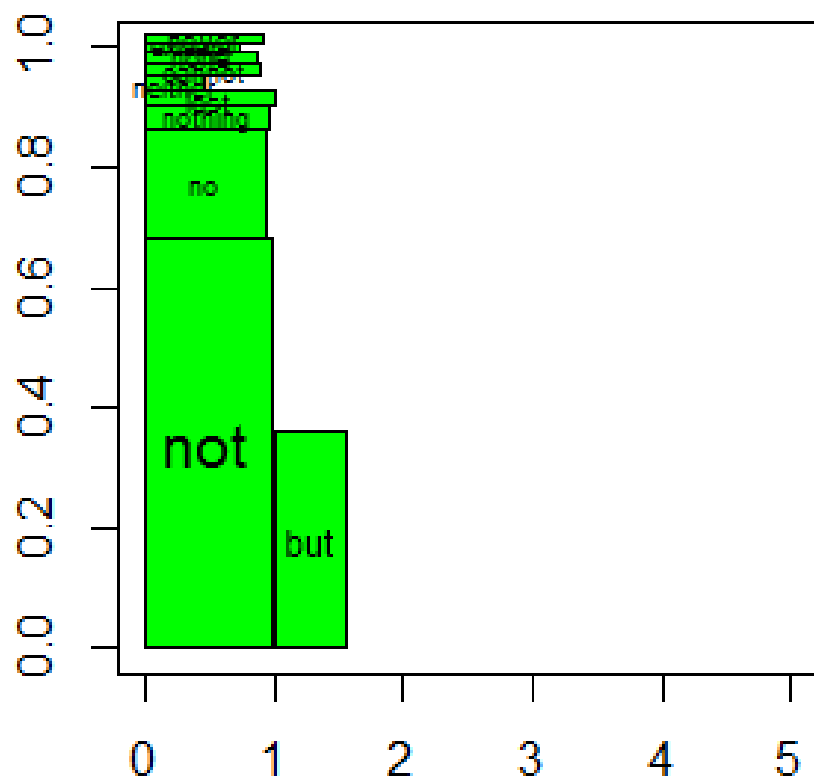


French: [ne|n'|non|sans]1 [pas|point|rien|personne|ni|aucun|plus]2 [mais]3

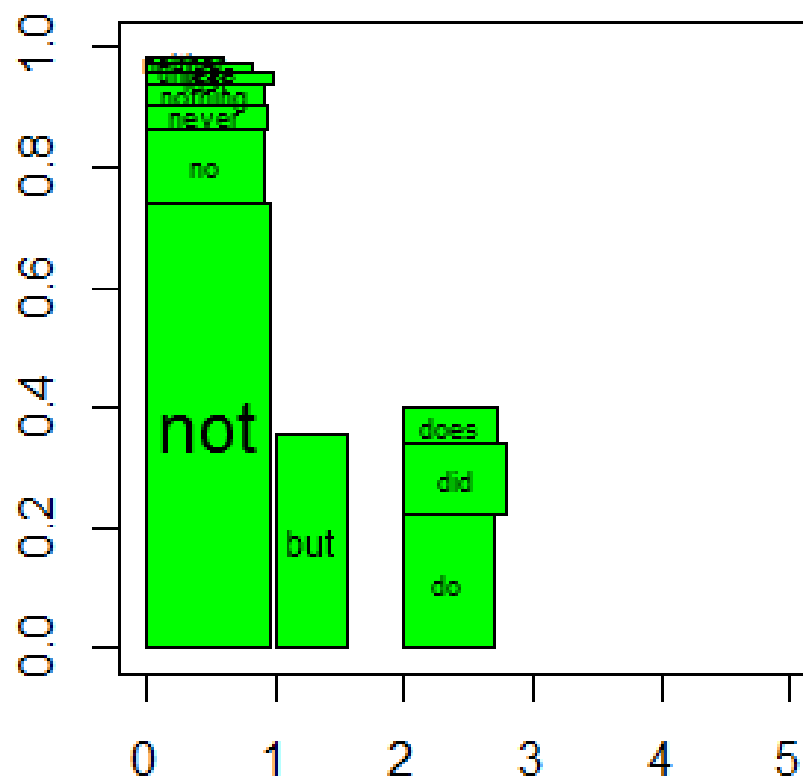
Alemannic: [nit|kei|nigs|nemads|nemi|keini]1 [sundern]2

40005017 [ne]1 croyez [pas]2 que je sois venu pour abolir la loi ou les prophètes je suis venu [non]1 pour abolir [mais]3 pour accomplir

Negation - English



Negation - English2



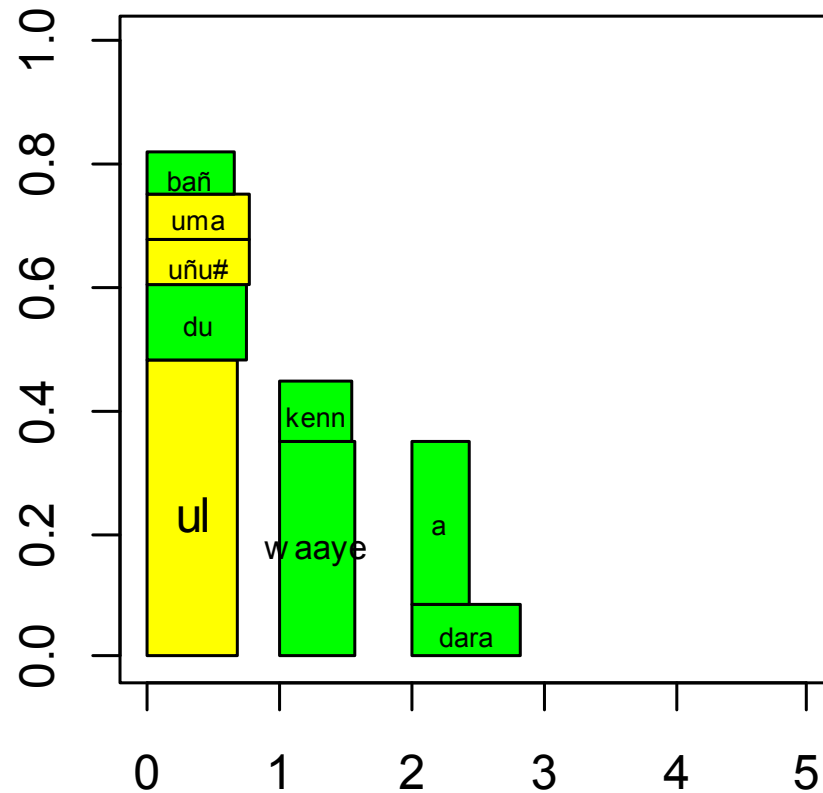
English: [not|no|nothing|lest|neither|cannot|none|except|never]1 [but]2

English2: [not|no|never|nothing|lest|unless|neither]1 [but]2 [do|did|does]3

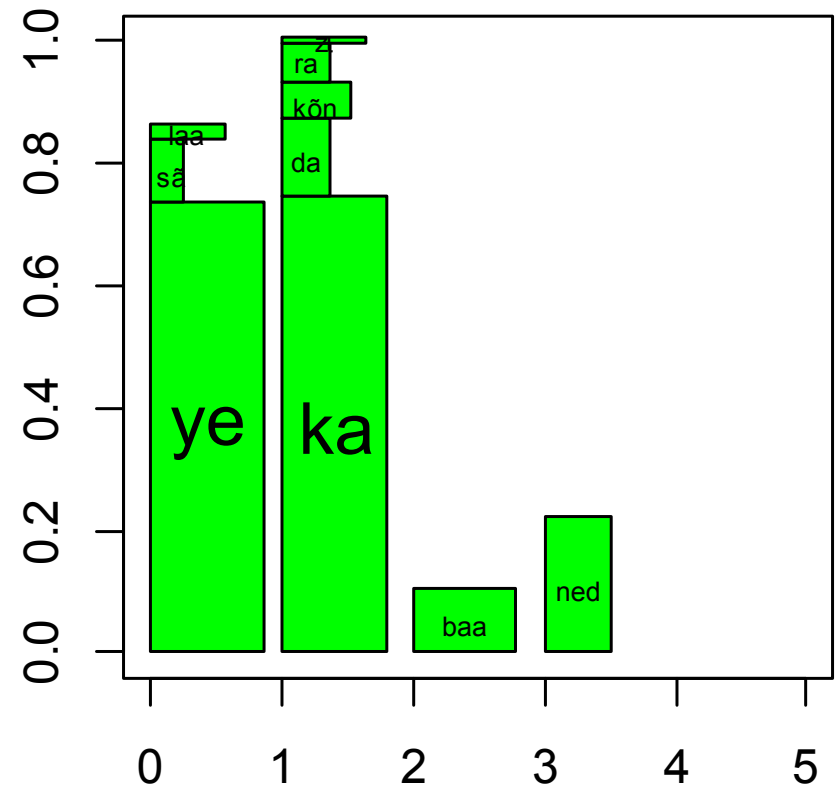
40005017 think [not]1 that i came ... i came [not]1 to destroy [but]2 to fulfil

40005017 [do]3 [not]1 think that i have... i have [not]1 come to destroy them [but]2 to fulfill them

Negation - Wolof



Negation - Moore



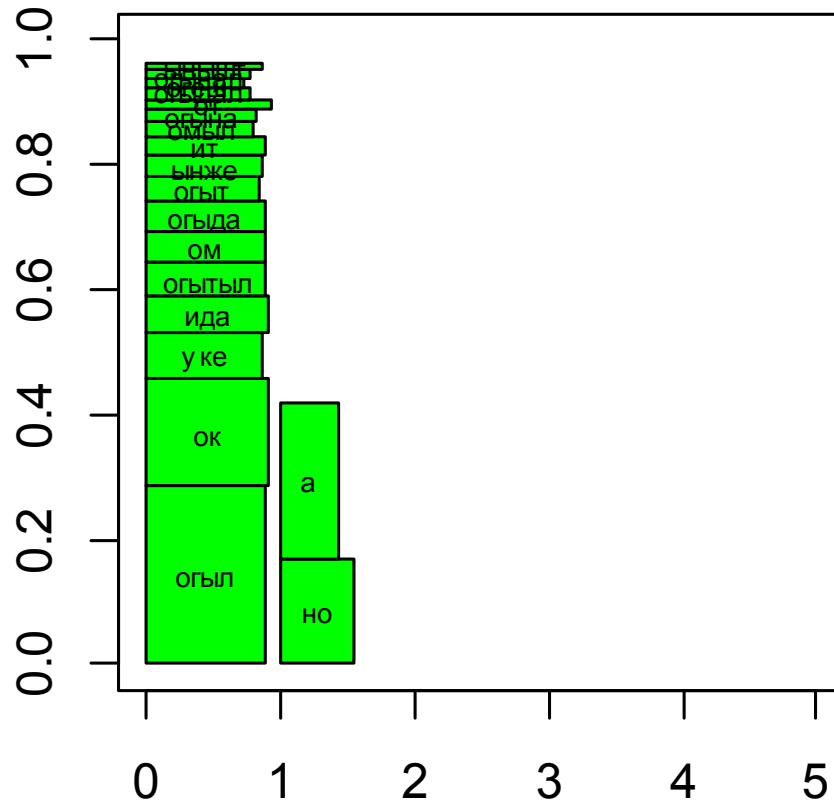
Wolof: [>ul<|du>uñu#<|>uma<|bañ]1 [waaye|kenn]2 [dara|a]3

Moore: [ye|sã|laa]1 [ka|da|kõn|ra|zi]2 [baa]3 [ned]4

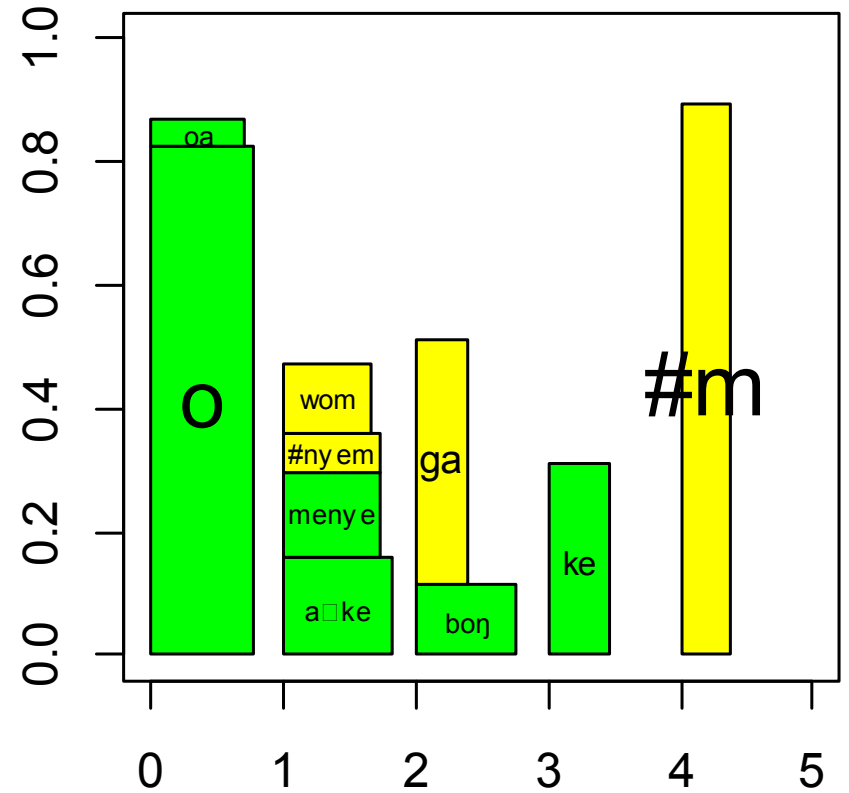
40005017 b[ul]1een defe ne ñëw ... ñëw[uma]1 ngir dindi leen [waaye]2 ngir ñu am ci man

40005017 [da]2 tags-y ... noy [ye]1 mam [ka]2 wa n na n sãam-b [ye]1 la yaa ti pids-ba

Negation - Mari



Negation - Ewe



Mari: [огыл|ок|уке|ида|огытыл|ом|огыда|огыт|ынже|ит|омыл|огына|от|огыдал|огеш|огынал|ынышт]1 [но|а]2

Ewe: [o|oa]1 [a ke|menye|>#nyem<|>wom<]2 [boŋ|>ga<]3 [ke]4 [>#m<]5

Algorithmic typology and procedural universals

The same procedure is applied to cross-linguistically similar material and the procedure applied to cross-linguistic data is fully explicit and therefore replicable. It can be implemented in a computer program and run without the intervention of a typologist (**algorithmic typology**).

The underlying idea is that the procedure of extraction is invariant (**procedural universal**) whereas the extracted structures can be highly variable depending on the texts and languages to which they are applied.

Comparison of Inequality

Stassen (1985) Comparison and Universal Grammar
Typology based on Standard of Comparison

Locative	‘Elephant big at/on horse’
Separative	‘ From horse elephant big’
Allative	‘Big elephant to horse’
Particle	‘Elephant big than horse’
Exceed	‘Elephant big exceeds horse/ exceeds horse in size’
Conjoined A	‘Elephant big, horse small’
B	‘Elephant big, horse not big’

Functional domain. Stassen, defined intensionally:

A construction having the semantic function of assigning a graded position on a predicative scale to two objects, standard and comparee are NPs

Here, defined extensionally:

For convenience, any verse containing English *than*

What about the predicate intensifier “more”, “-er”?

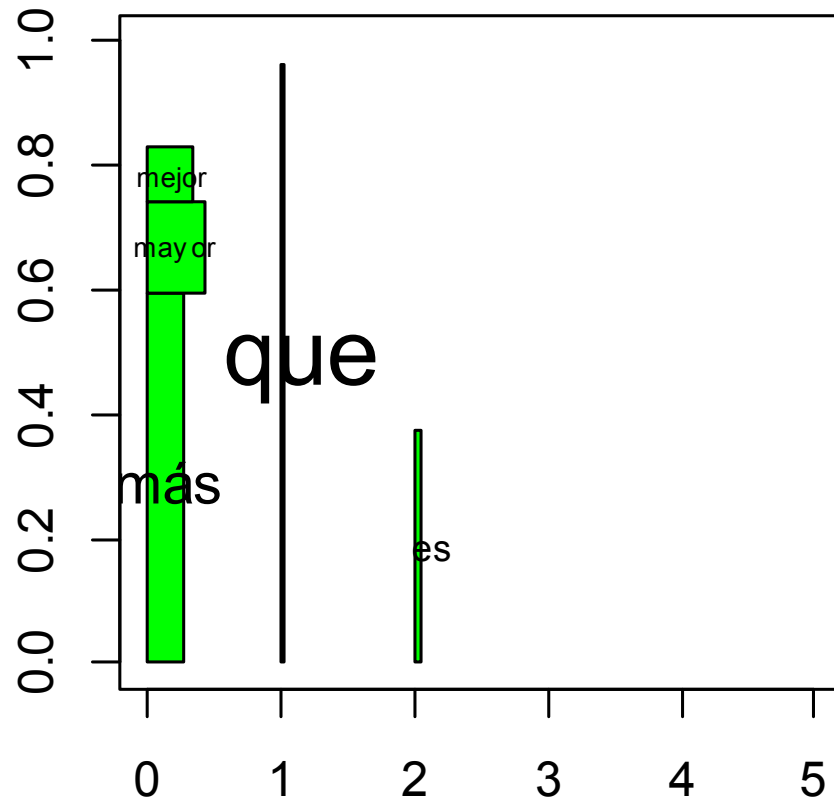
“a majority of the languages do not use such an overt marking” (Stassen 1985: 27)

“...I have not succeeded in finding an explanatory principle on the basis of which the presence or absence of this marking can be predicted. Hence, I will assume that the phenomenon of comparative-marking is irrelevant to our typology of comparative constructions, and that it must be explained in terms of (as yet unknown) regularities which are independent of those that determine the choice of a particular type of comparative construction. Therefore, I will not indicate systematically whether or not a given language requires morphological marking of the comparative predicate” (Stassen 1985: 28).

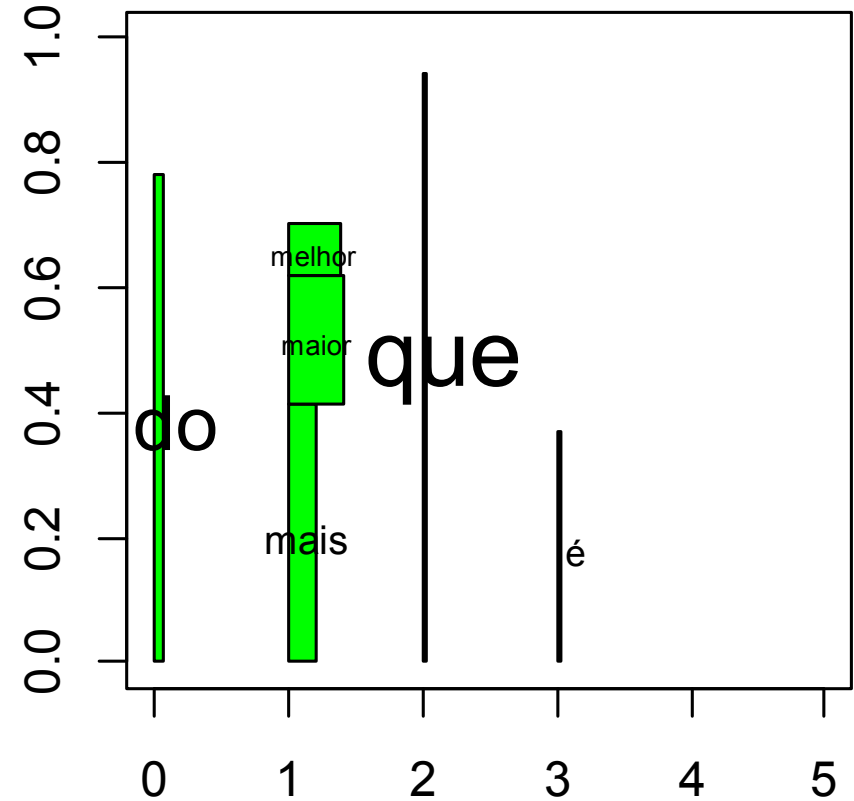
-> The data collection in Stassen’s typology is explanation-driven.

In the present approach, data collection is independent

Comparison - Spanish



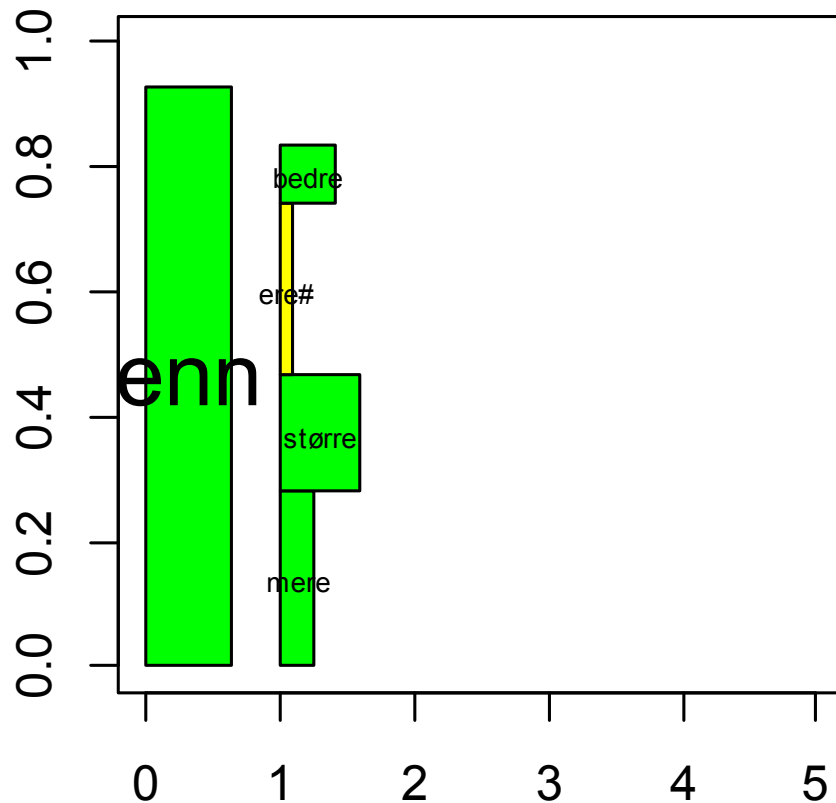
Comparison - Portuguese



Portuguese: **[do]1 [mais|maior|melhor]2 [que]3 [é]4**
 Spanish: **[más|mayor|mejor]1 [que]2 [es]3**

41001007 vem aquele [que]3 [é]4 [mais]2 poderoso [do]1 [que]3 eu de quem não sou digno
 41001007 viene tras mí el [que]2 [es]3 [más]1 poderoso [que]2 yo al cual no soy digno de

Comparison - Norwegian



Norwegian: **[enn]1 [mere|større|>ere#<|bedre]2**

Lithuanian: **[negu|už]1 [>esn<|daugiau|geriau|labiau|lengviau]2**

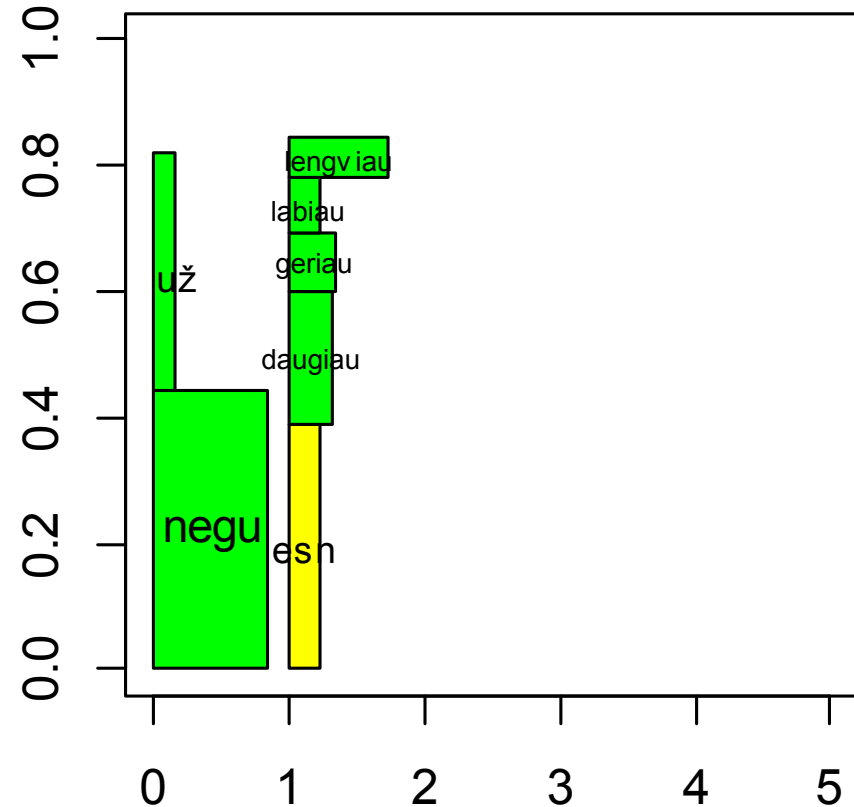
41001007

er mig kommer den som er **sterk[ere]2 [enn]1** jeg han hvis skorem jeg ikk

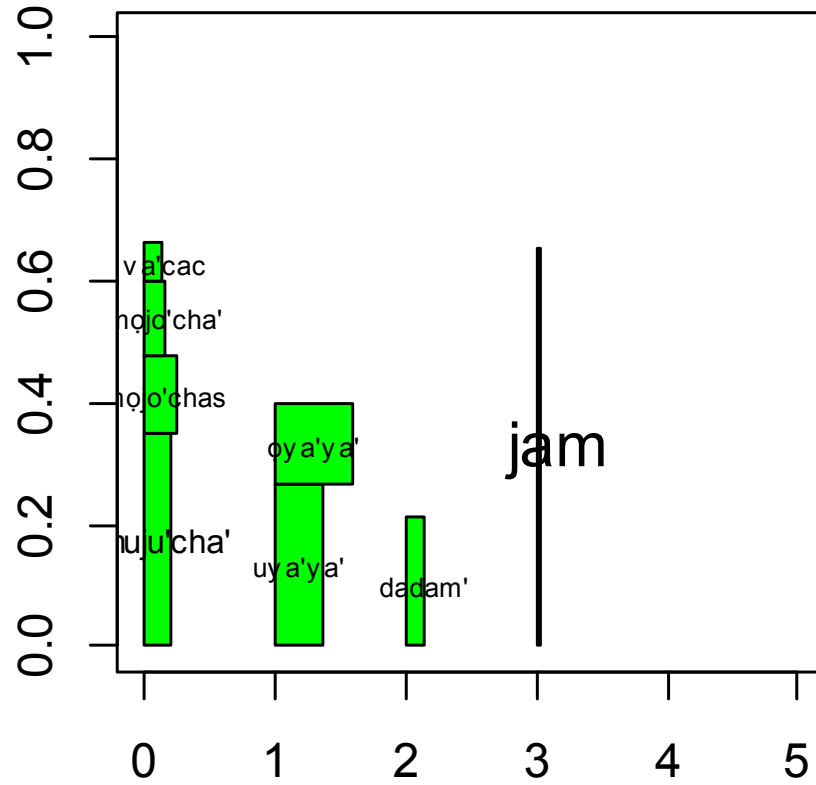
41001007

skelbė po manęs ateina **galing[esn]2is [už]1 mane** aš nevertas nusilenkęs

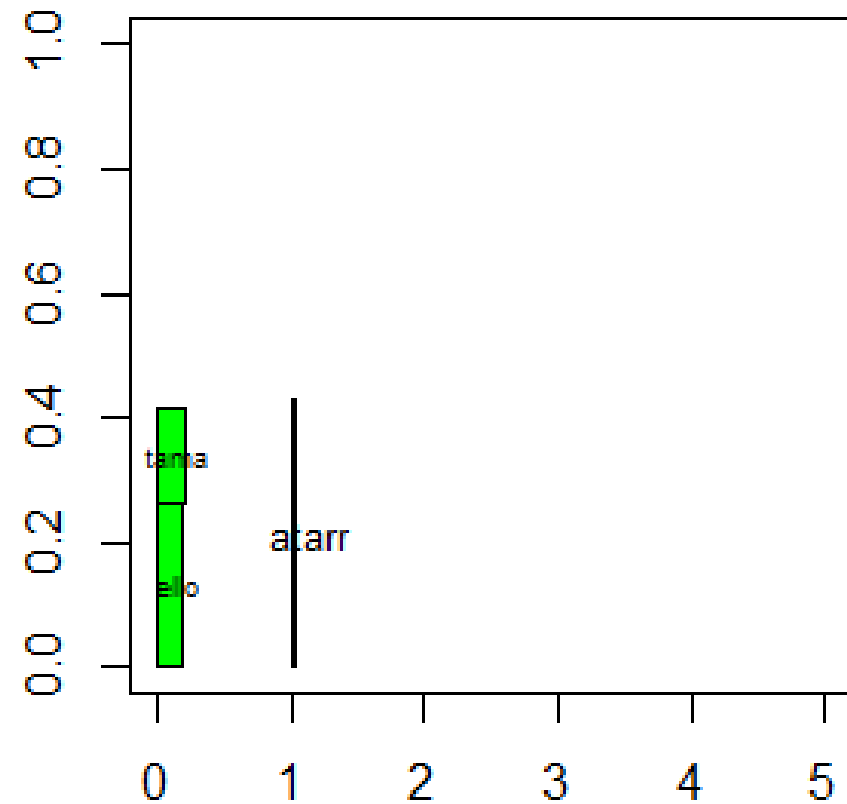
Comparison - Lithuanian



Comparison - Tsimane



Comparison - Yanesha



Tsimane: [muju'cha'|mojo'chas|mojo'cha'|va'cac]1 [uya'ya'|oya'ya']2 [dadam']3 [jam]4

Yanesha: [ello|tama]1 [atarr]2

No intensifier in the conjoined type?

No! Rather no standard marker and more than one intensifier instead.

Yanesha' (Matthew 3:11)

Ña-pa' ñeñ [atarr]² ahuamencaĩ-asha' na-ña-pa' ama [tama]¹ ahuamencaĩ-eyay-no.
he-TOP REL **much** strong-PROP I-SEQ-TOP, **not** **that.much** mighty-NEG.SUFF-MIDD
'that {cometh after me} is mightier than I'

atarr P1; *ama tama* P2 'much P1, not that.much P2'

atarr P1; *ello metan(err)an* P2 'much P1, more/again/separated surpass(again) P2'

Tsimane (Matthew 3:11)

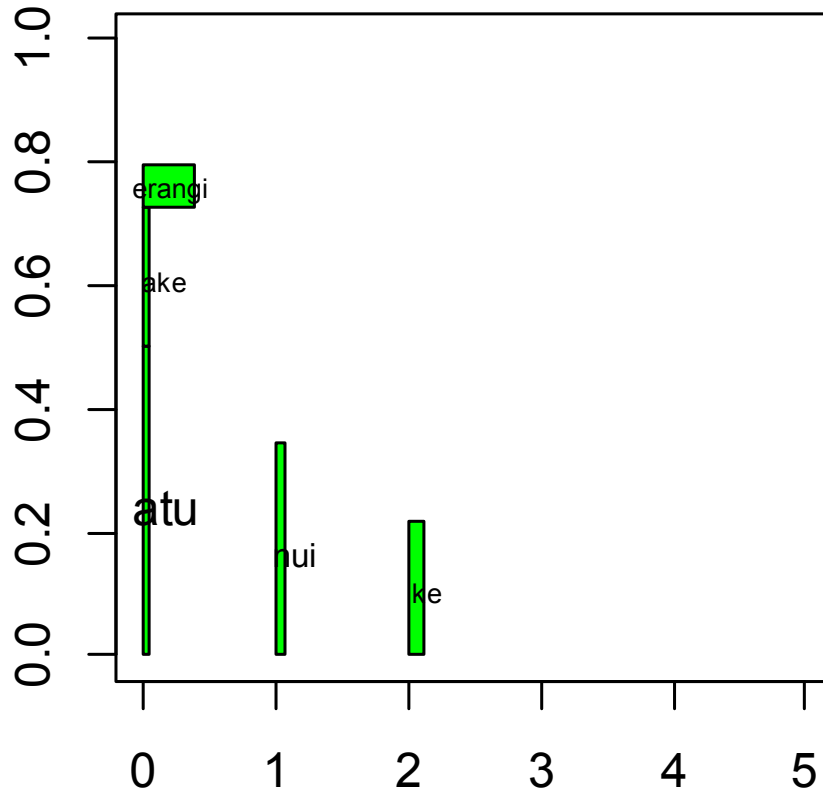
Mu' qui ra' atsi [muju'cha']¹ fer bu'yi-ty,
he/that so.that FUT come **more** strong be.in .a.position-MASC
[jam]⁴ jeñej yu, [uya'ya']² yu...
not like I **less** I

Tsimane has three intensifier slots

P1 [dadam']³ [muju'cha']¹ ... [oya'ya']² P2 'P1 better more ... less P2'

42012007 [dadam']³ mu' [muju'cha']¹ äräjjinac mi'in [jam]⁴ jeñej jaijtyi' in [oya'ya']² ma'jotacsi

Comparison - Maori



Maori: [atu|ake|erangi]1 [nui]2 [ke]3

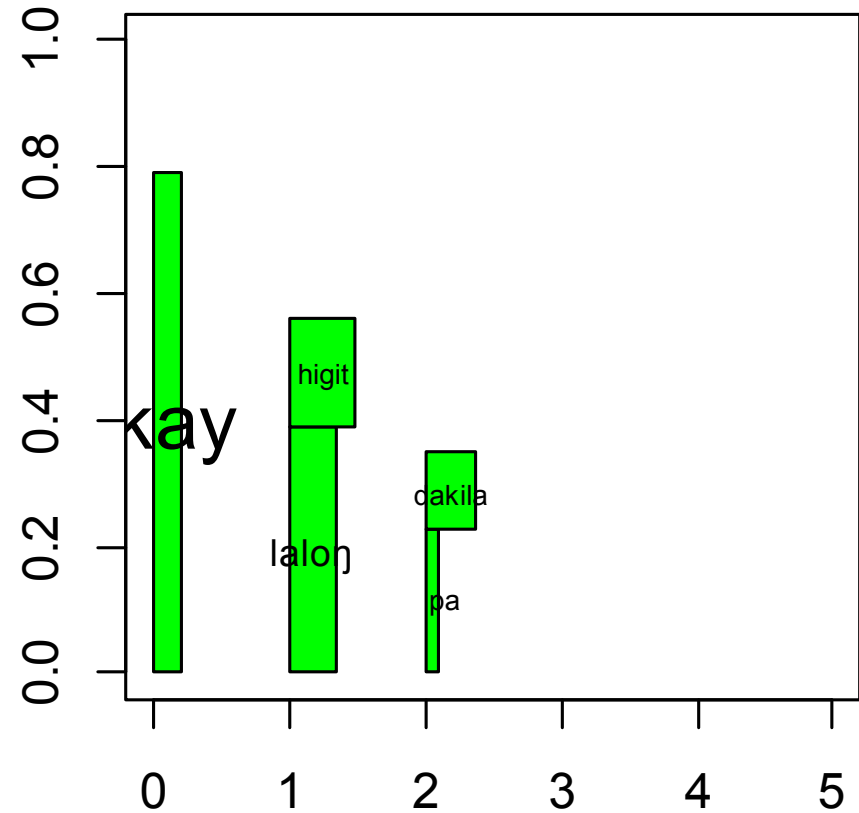
Tagalog: [kay]1 [laloŋ|higit]2 [pa|dakila]3

41001007 ...i ahau tetahi he kaha rawa **[ake]1 i** ahau e kore ahau e tau ki

41001007 ...sumusunod sa hulihan ko aŋ **[laloŋ]2** makapanjarihan **[kay]1 sa** akin hindi ako

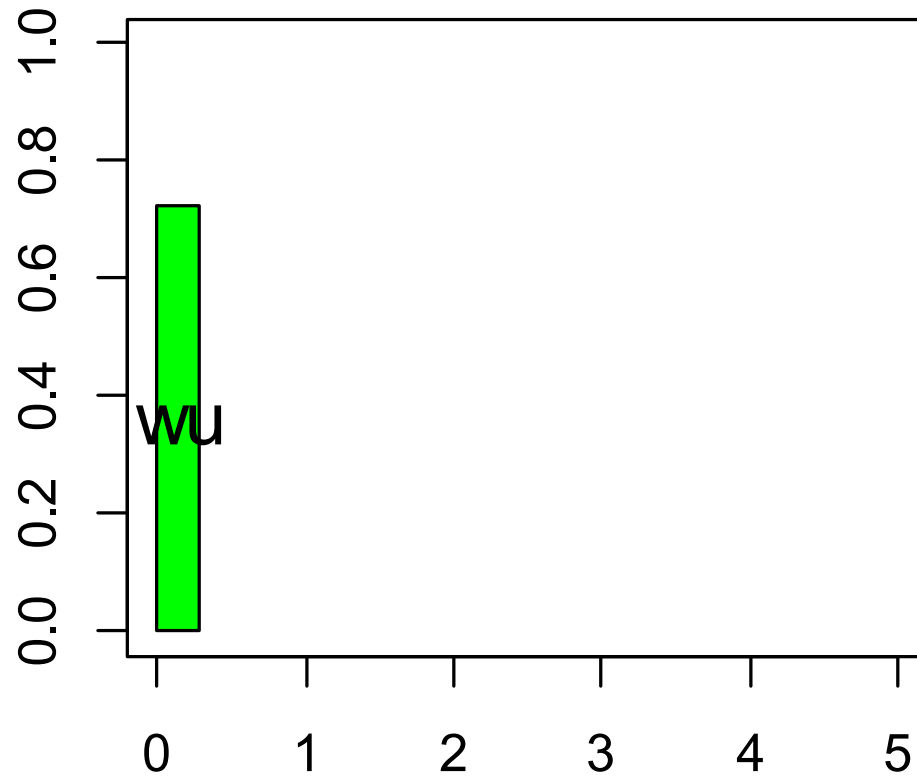
41012033 ...sa kaniyaŋ sarili ay **[higit]2 [pa]3 [kay]1 sa** lahat naŋ maŋa handog...

Comparison - Tagalog

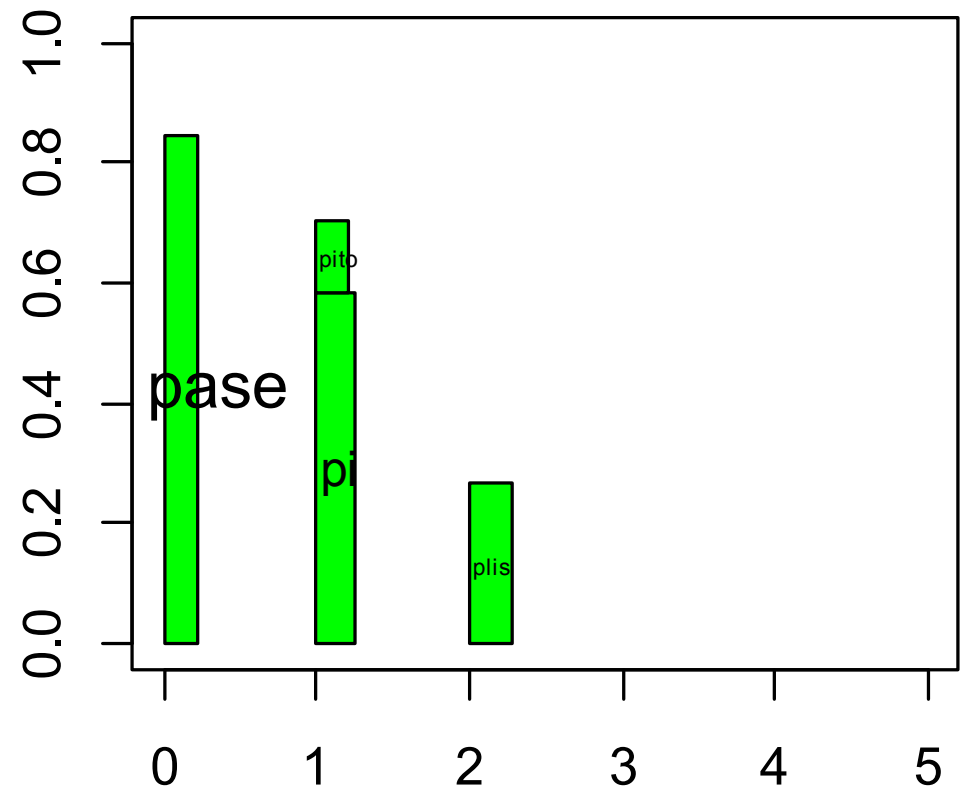


atu 'away', *ake* 'up', *nui* 'big', /kee/ 'different'

Comparison - Ewe

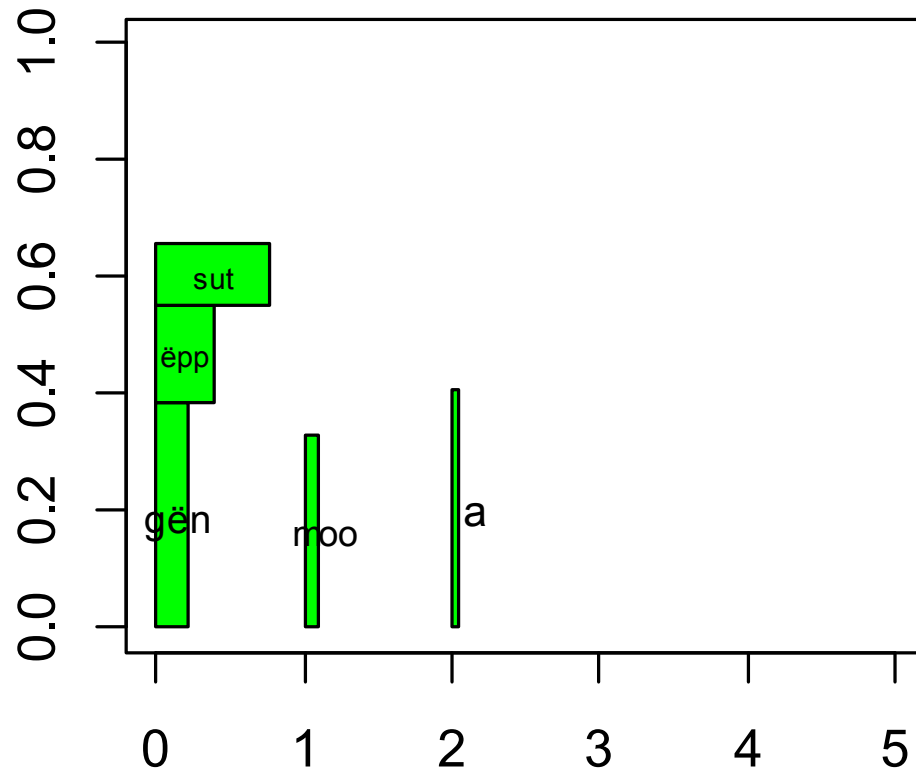


Comparison - Haitian

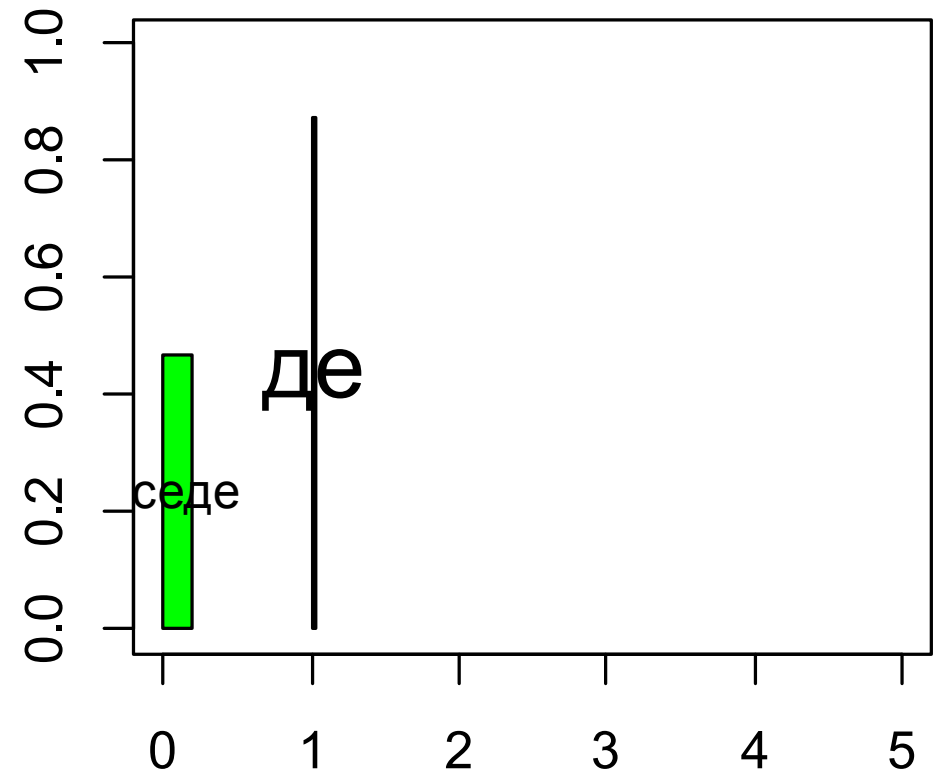


41001007 moun k'ap vin apre m' lan gen [plis]3 pouvwa anpil [pase]1 m' mwen pa bon
40010031 nou pa bezwen pè menm nou vo [pi]2 [plis]3 [pase]1 anpil ti zwazo
40011009 t wi mwen menm mwen di nou li [pi]2 [plis]3 [pase]1 yon pwofèt
41009043 l' jete [pito]2 ou antre nan lavi a ak yon sèl men [pase]1 pou ou rete ak tou de men

Comparison - Wolof



Comparison - Erzya



40003011 kiy ñëw sama gannaaw [moo]2 ma [ëpp]1 kàttan

41010025 giléem jaar ci bën-bënu pusa [moo]2 [gën]1 [a]3 yomb boroom alal dugg ci

40003011 мельган сыцясь мон[де]2нь [седе]1 пек виев

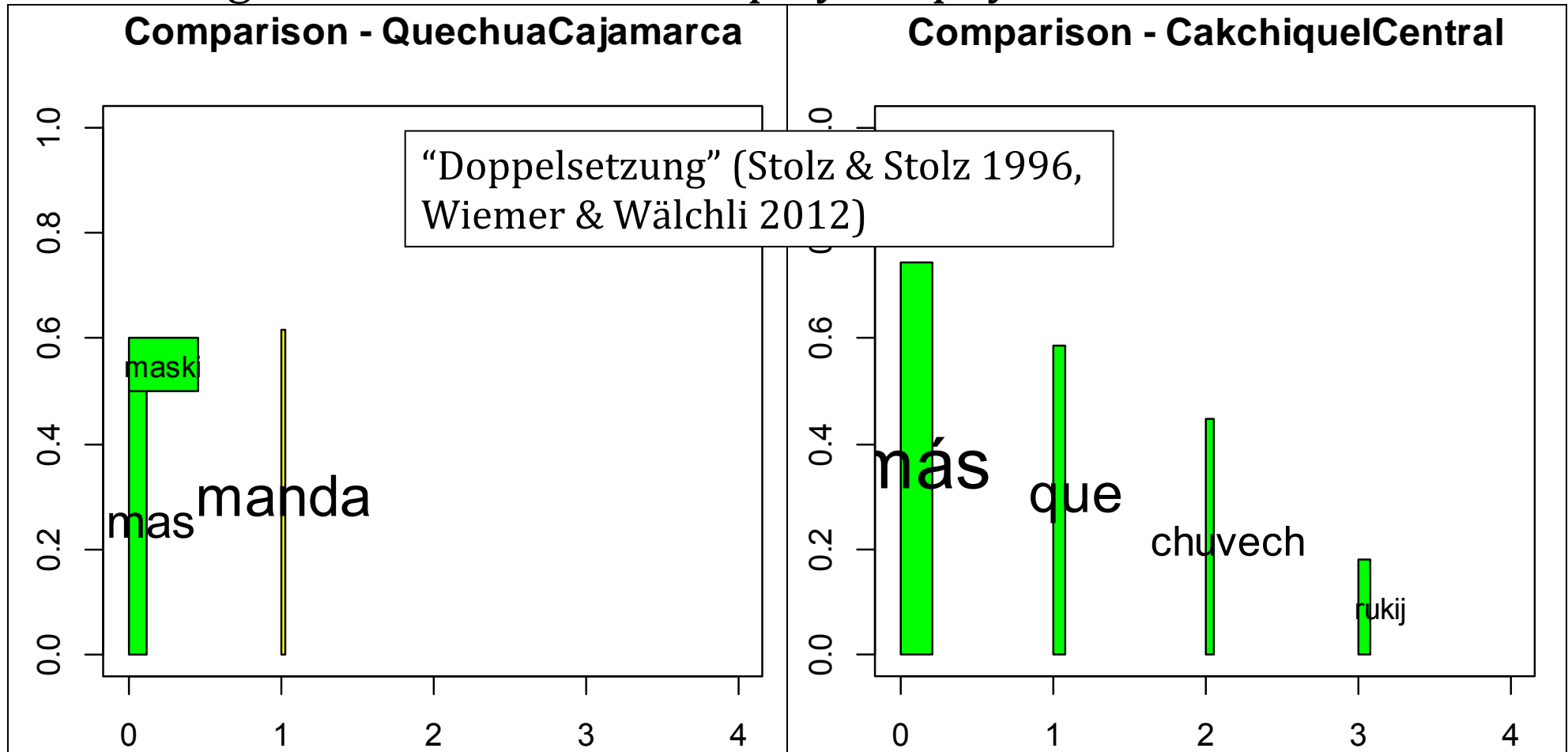
“Double marking” in comparison is described in Ansaldo (1999) as one type of comparison that needs to be accounted for by particular explanations. However, “double marking” is clearly the rule rather than the exception.

In a vast majority of the languages surveyed there are both standard markers and intensifiers extracted in comparison

- unless there is a standard marker extended into the intensifying domain or
- unless there are two different types of intensifiers in the conjoined type

In a clear majority of languages, comparison is syntagmatically polymorphous

Borrowing of function words and polymorphy



40003011

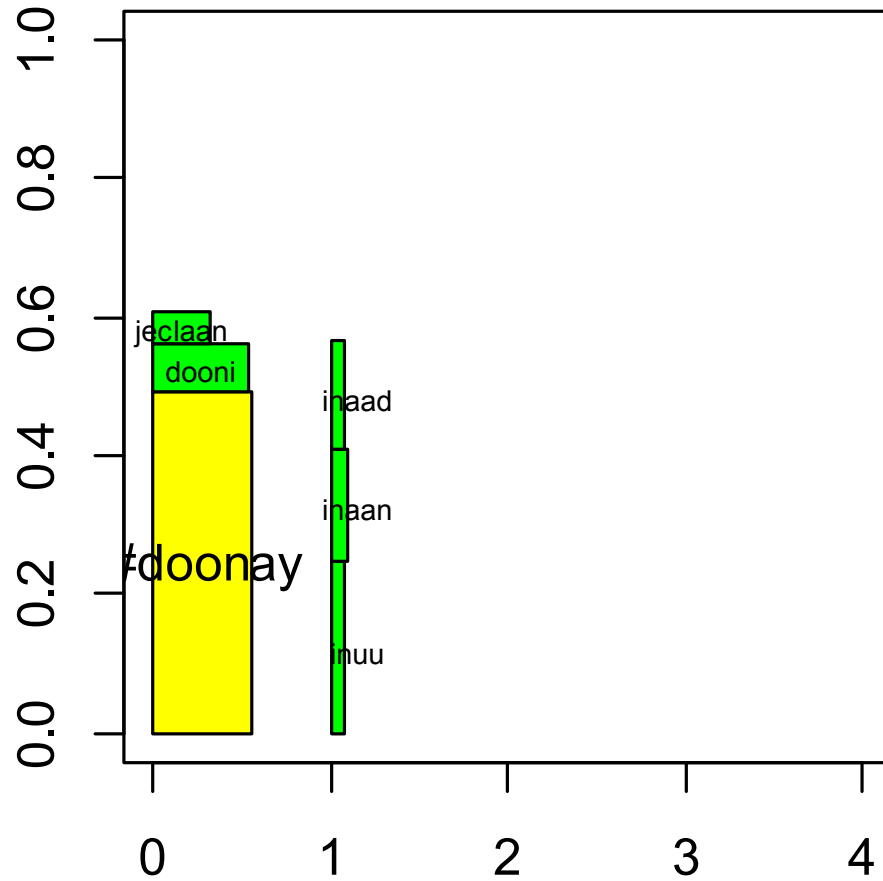
illapa piru rinnam shamuq noqa[manda]2 suq [mas]1 pudirniyuq

41010025

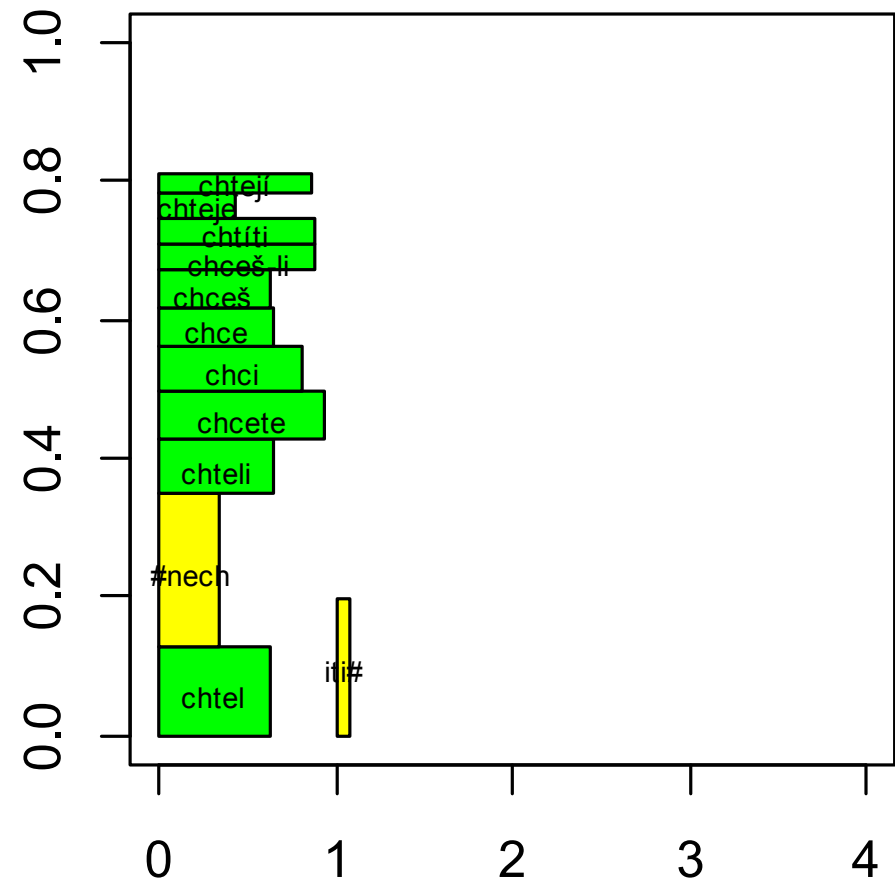
xa [más]1 laek ... camello ... jun bak [que]2 [chuvech]3 jun beyon noc

Next domain: WANT

Want - Somali



Want - Czech



Some marker complexes for ‘Want’:

Czech:	[chtě >#nech< chtěli chcete (5 more forms)]1 [>iti#<]2
Zulu:	[>thand<]1 [>#uku]2
Wolof:	[bëgg >#bëgg<]1 [a]2
Greek (Modern):	[vα θέλω]1 [θέλει θέλεις θέλω θέλετε (6 more forms)]2
Somali:	[>#doonay< dooni jeclaan]1 [inuu inaan inaad]2
Saramaccan:	[kë]1
Greenlandic West:	[>uma<]1
Maltese:	[>rid< ried riedx]1

Haspelmath’s (2005) typology of ‘want’ (simplified) and marker complexes

	Complement subject implicit	Complement subject expressed overtly	Desiderative affix
Marker complex without subordinator	Saramaccan	Maltese	West Greenlandic
Marker complex with subordinator	Czech, Zulu, Wolof	Modern Greek, Somali	

Khanina (2008, 2010) vs. Goddard and Wierzbicka (2010).

- Khanina claims that ‘want’ is not universal in the sense of “being treated only as particular type of a more general situation” (2008: 845).
- In her variety sample, she finds that 95 of 136 desideratives (her cover term for all ‘want’ expressions) “are regularly used to express other situations than pure ‘want’” (2008: 847).
- According to her this is a challenge for Natural Semantic Metalanguage where ‘want’ is considered to be a semantic prime, “i.e. an indivisible unit of meaning with a lexical exponent in all languages” (Goddard and Wierzbicka 2010: 108).
- For Natural Semantic Metalanguage it is important to distinguish between polysemy and semantic generality. For instance, Spanish, *querer* has two (or more) meanings ‘want’ and ‘love/like’ rather than one general meaning ‘want/love/like’. Khanina, however, argues that multiple meanings of ‘want’ expressions are best analyzed as macrofunction by default.
- Khanina questions the universality of ‘want’. According to Goddard and Wierzbicka (2010) this is due only to an underestimation of polysemy.

Parallels to the present approach

- Goddard and Wierzbicka (2010: 114): “a semantically primitive meaning will always be expounded by means of specifically lexical material, by a ‘segmental sign’, and not (for example) by reduplication, or ablaut, or solely through a grammatical construction.”
- Khanina’s approach is similar to the present one in that she explicitly chooses desideratives in European languages (the practical meta-languages of most descriptive grammars) as her point of departure. She also speaks of translational equivalents of Standard Average European. If I here choose Classical Greek *ethelo* ‘want’ to define the domain, the basic rationale is very similar.
- Natural Semantic Metalanguage proceeds to a large extent onomasiologically as far as semantic primes are concerned. For all concepts which are claimed not to be semantic primes, however, Natural Semantic Metalanguage rather takes a semasiological stance, but this does not need to concern us here since ‘want’ is claimed to be a semantic prime.

Discovery mode vs. proficiency mode

In the approach taken here, we operate in the discovery mode. This means that we cannot make any distinction between **polysemy** and macrofunction since **there is no established marker-meaning relationship**. If we want to find out how a meaning is expressed cross-linguistically, what is given is one meaning and all potential marker candidates. The marker-meaning relationship cannot be given, otherwise we would not **find out**. If the marker-meaning relationship is given, we already **know** what the meaning of a form is.

However, I do not assume that there should be an isomorphism between marker and meaning, as long as there is a collocation of meaning and marker, a relationship can be established irrespective of whether there is polysemy in a narrow sense or macrofunction. In the material considered here there is no problem to establish a meaning-marker relationship in the 'want'-domain in virtually all languages considered.

WANT=SAY, but in very different ways

Kobon (Mark 10:51)

“Yip nihön g-aŋ, a gi-mön, au-ab-ön?” ö g-a...
1SG.OBJ what do-IMP3SG, QUOT do-SS2SG come-PRS-2SG QUOT do-RMPST3SG

“Amgö u kauyaŋ niŋ-nam, a g-em, au-ab-in,” a g-a.
eye that again see-PRESCR1SG QUOT do-SS1SG come-PRS1SG QUOT do-RMPST3SG

[Jesus answered him,] "What do you want me to do for you?" [The blind man said to him, "Rhabboni,] that I may see again." Literally: He like: "You come like 'Do me what' ...

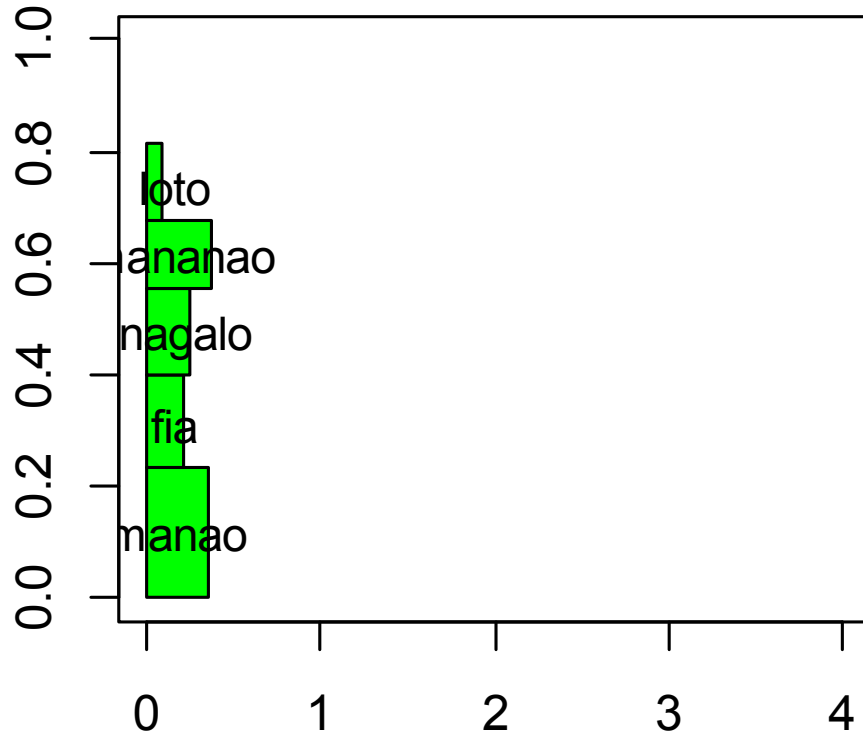
Bukiyip (Mark 10:51)

“Nyak ny-a-kli i-nek-um-enyu moneken?” ...
2SG 2SG-REA-say 1SG:IRR-do-BENEF-2SG.OBJ what

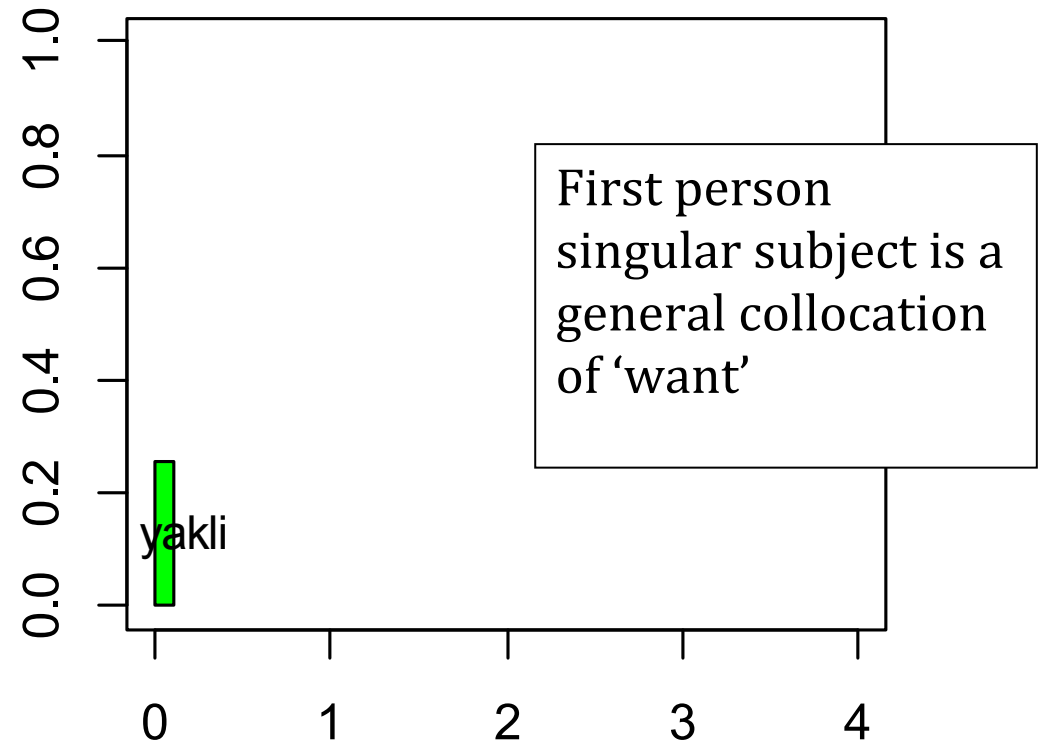
“...yek y-a-kli i-na-tulugun.”
1SG 1SG-IRR-say 1SG:IRR-REFL-look/see

[Jesus answered him,] “What do you want me to do for you?” [The blind man said to him, “Rhabboni,] that I may see again.”

Want - Samoan



Want - Bukiyip



Considerable paradigmatic polymorphy.

Samoan <manao> /mana'o/ (verb) 'want, desire', *fia* (particle preposed to verb) expresses a wish, *finagalo* (noun honorific) 'wish', <mānana'o> /mana'o/ (verb) plural of *mana'o* 'want, desire', *loto* (noun) 'heart, will'.

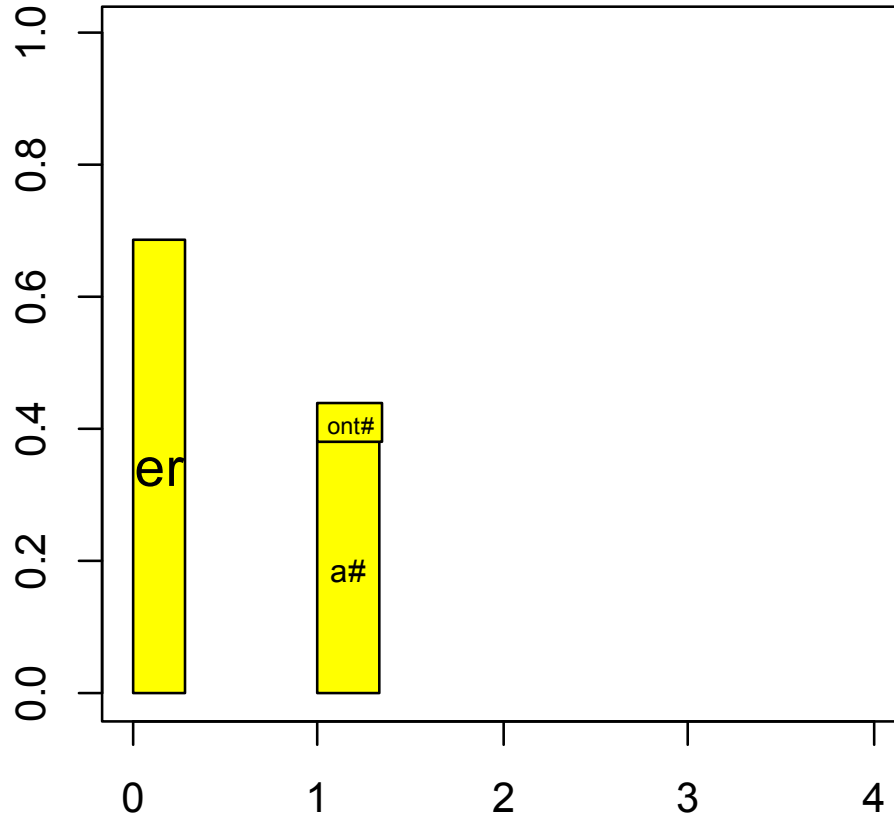
A local approach to semantic decomposition

- Every utterance (or part of utterance) has an **exemplary contextual meaning** of its own that is unique (see, e.g., Wälchli and Cysouw 2012 for discussion).
- This exemplary meaning is usually highly similar to that of many other utterances, which is why it is possible to identify clusters of utterances with highly similar meaning: these share **ranges of meaning**.
- For every range of meaning there is a **local decomposition** of exemplary contextual meaning into two components: **the general meaning of the range versus everything else**.
- If the meaning range is **lexical**, the exemplars will most obviously differ in their **grammatical meanings**. If the meaning range is grammatical, the variable elements will most obviously be lexical. This yields an appearance of a global division of meaning into lexical and grammatical meaning. However, this division is not in any way rigid.
- Grammatical meanings and lexical meanings are treated alike. No difference between lexical and grammatical typology.

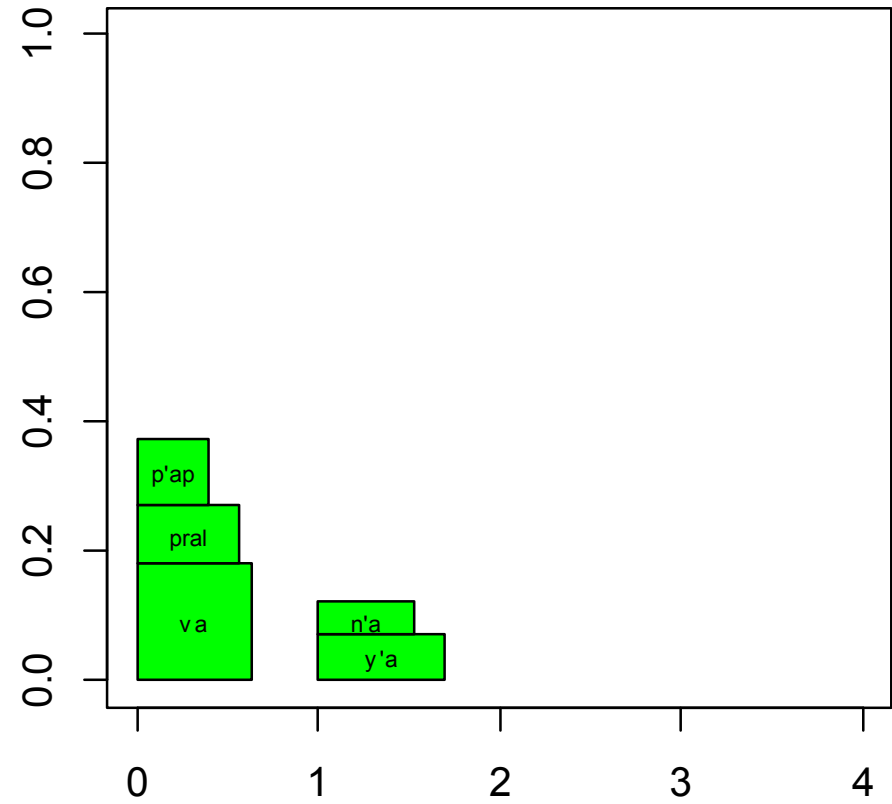
Future tense

(Indonesian *akan*)

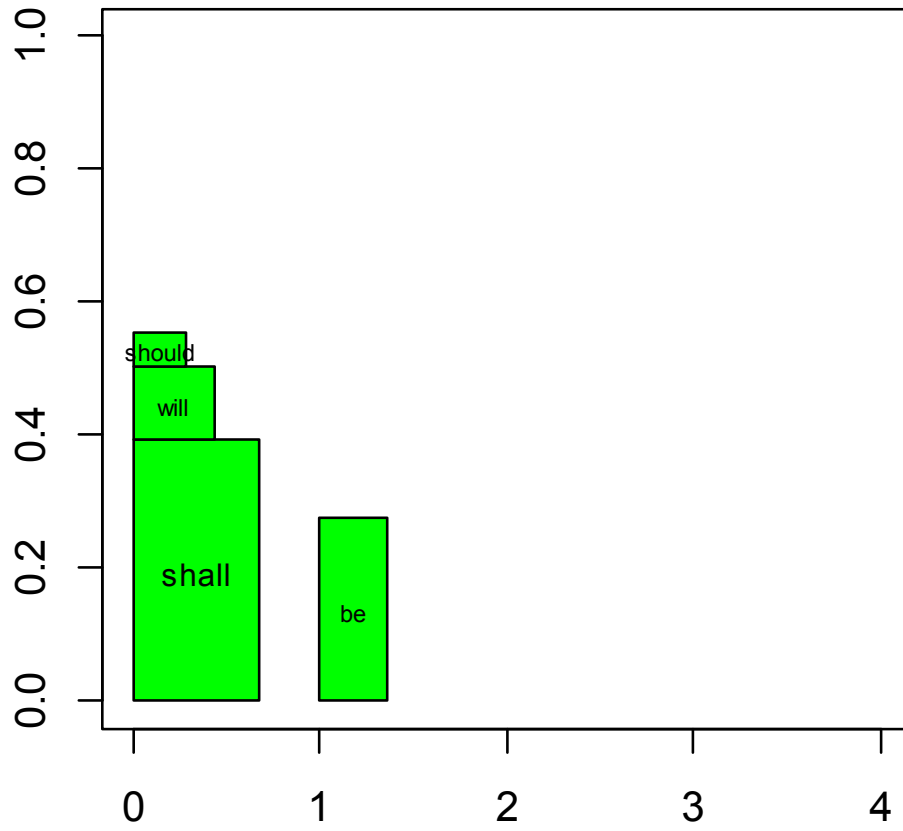
Future - French



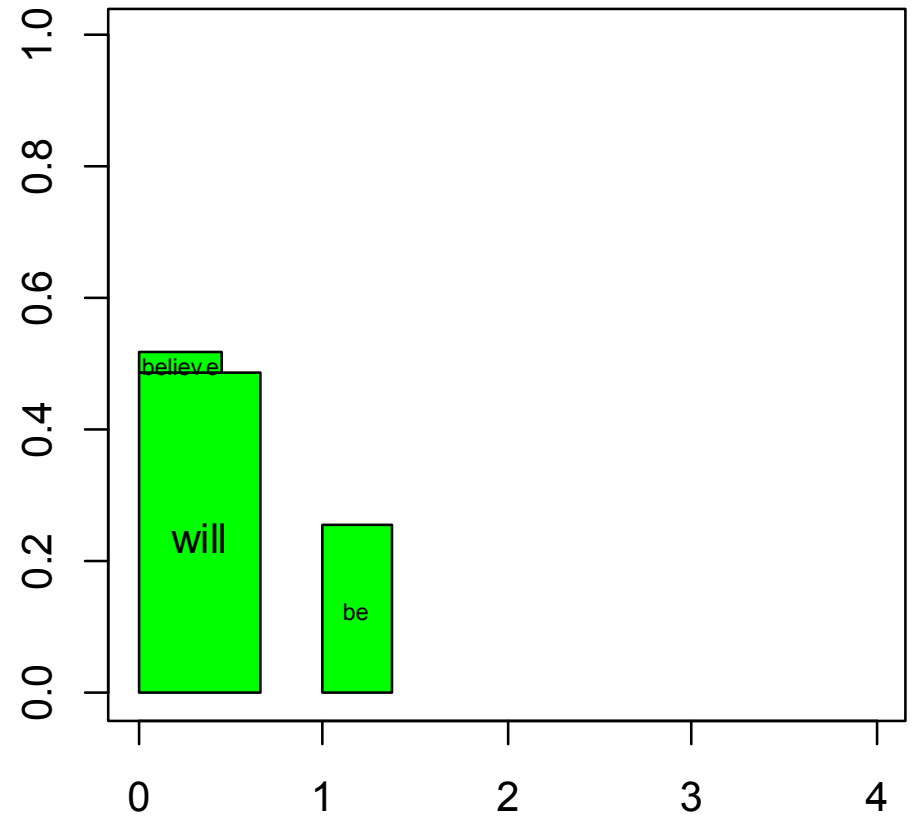
Future - Haitian



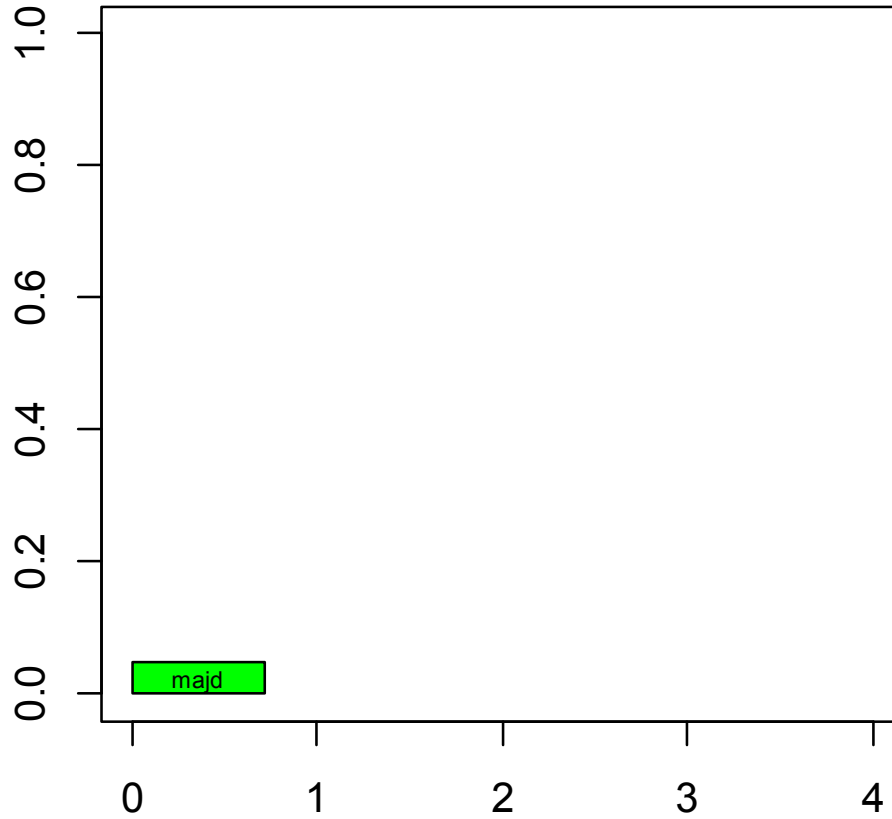
Future - English



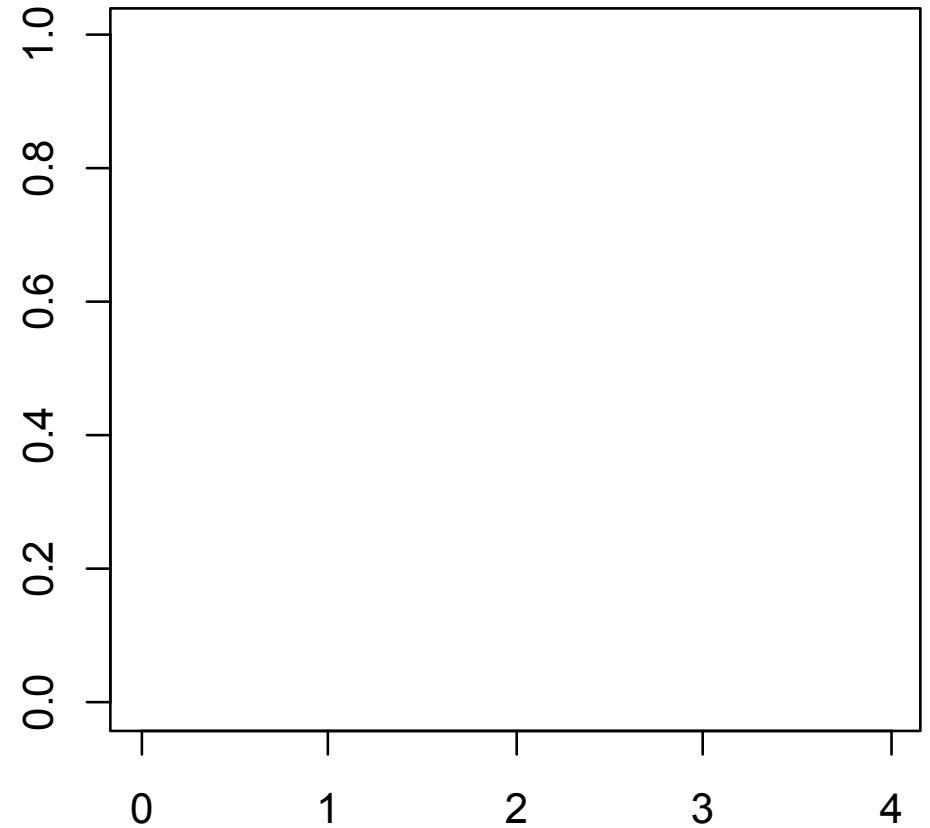
Future - English2



Future - Hungarian

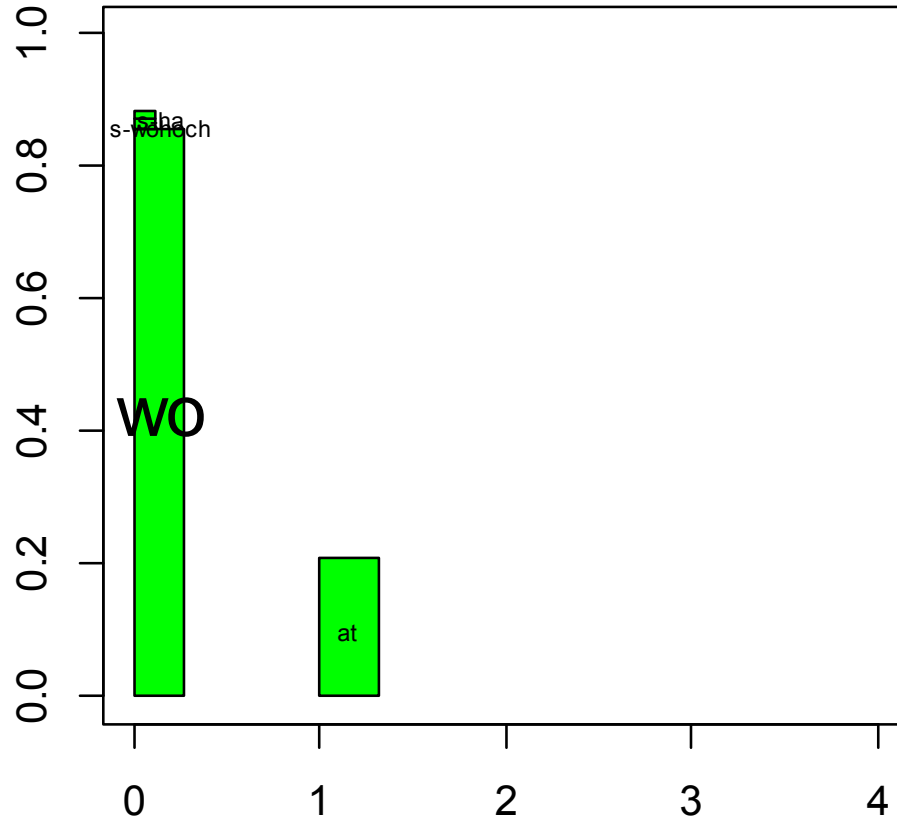


Future - Finnish

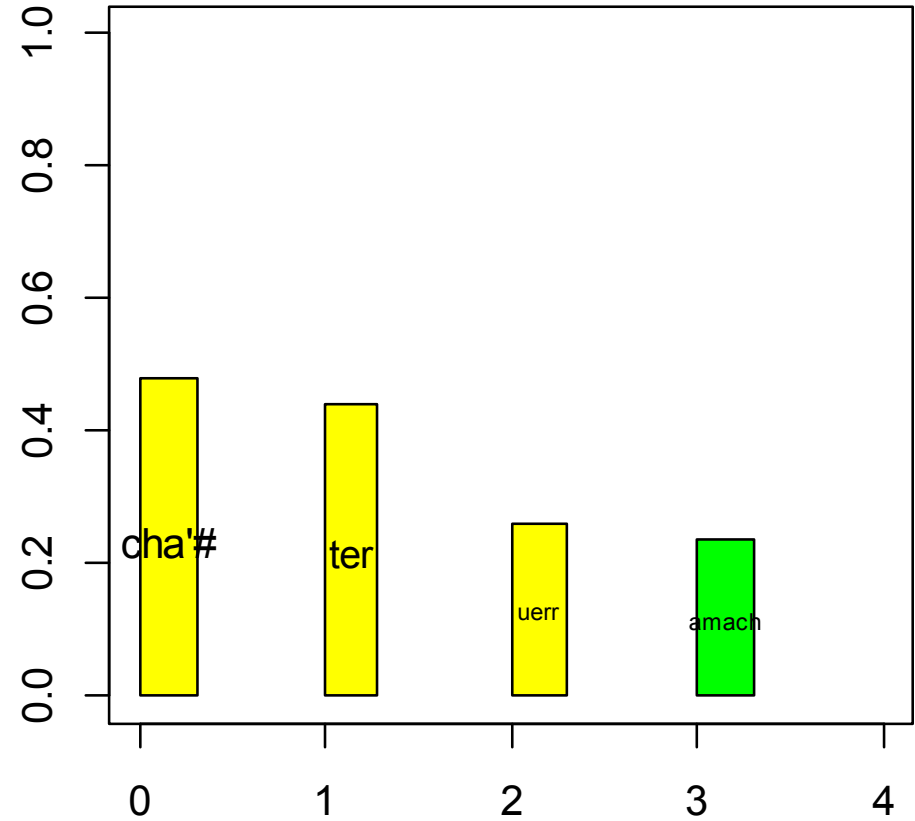


According to Dahl (1985: 105) the number of languages without future tense category is fairly small.

Future - Papago



Future - Yanesha



In Papago the potential modal marker *wo* is extracted along with *at* consisting of *a*-non-imperative mood and *-t* contemporary tense (used in past, present and future contexts opposed to zero pre-experiential and *d* remote past).

Predicative Possession

Stassen (2009): four basic types of predicative possession:

Type		Defining marker
• Locational	' At/to possessor [there] is/exists a possessee'	Possessor
• With	'Possessor is/exists with a possessee'	Possessee
• Topic	'[As for] Possessor, possessee is/exists'	None
• Have	'Possessor has a possessee'	Predicate

Predicate

Negation

Possessor

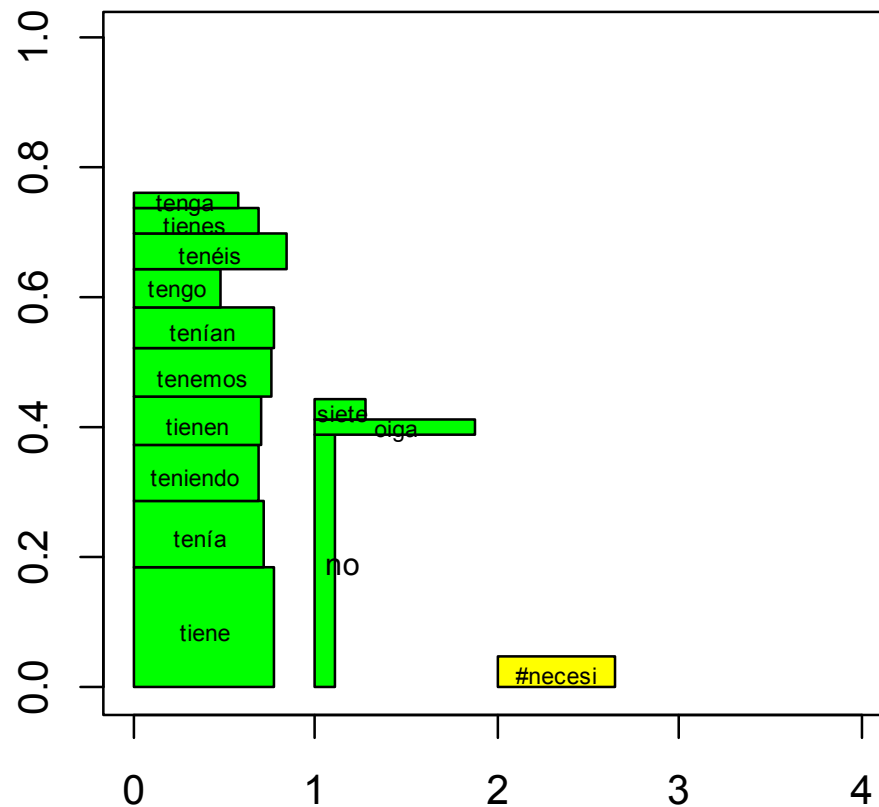
Indefinite

Possessee

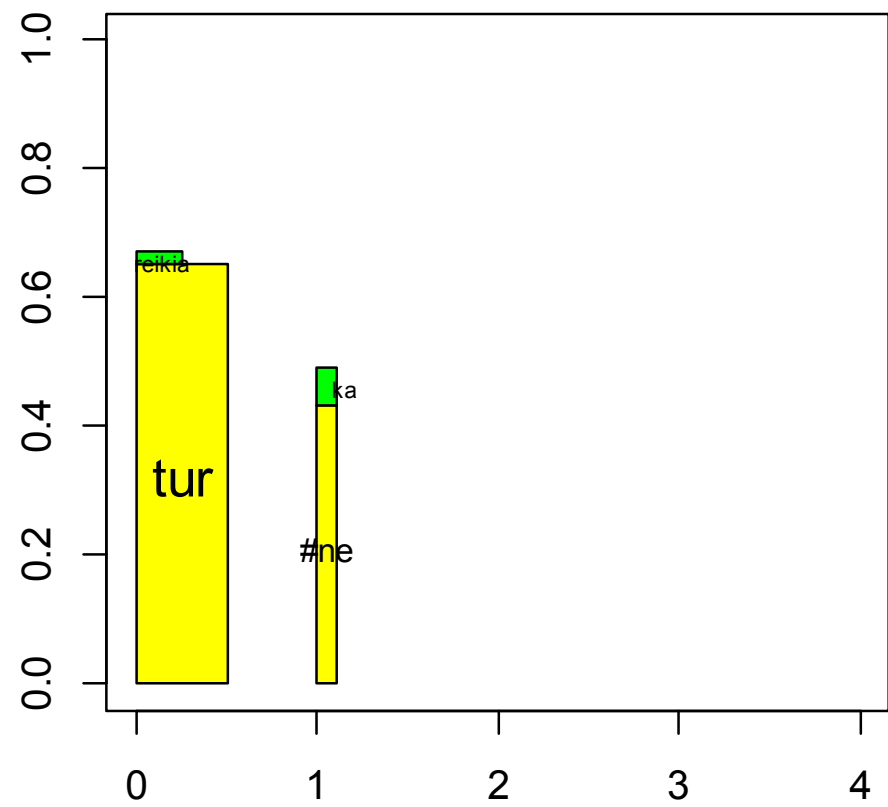
Negation and indefinite (indefinite article of possessum) are general collocations of predicative possession

Have-possessives are easiest to extract if they are recent (high dedication) when they have not managed yet to grammaticalize into something else

Pred. Possession - Spanish



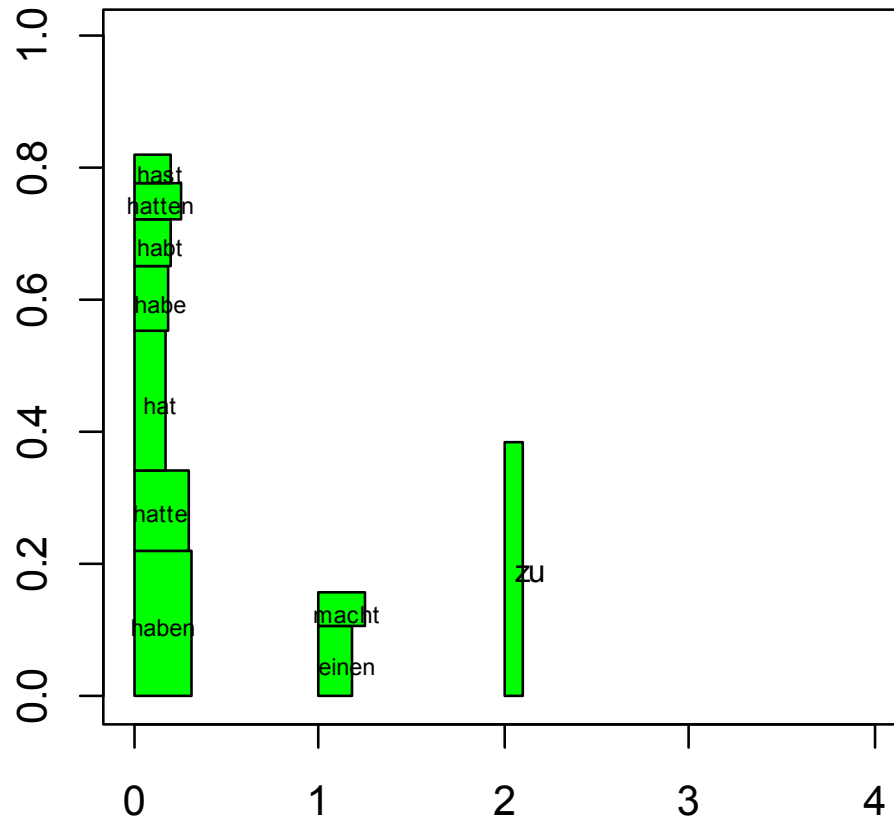
Pred. Possession - Lithuanian



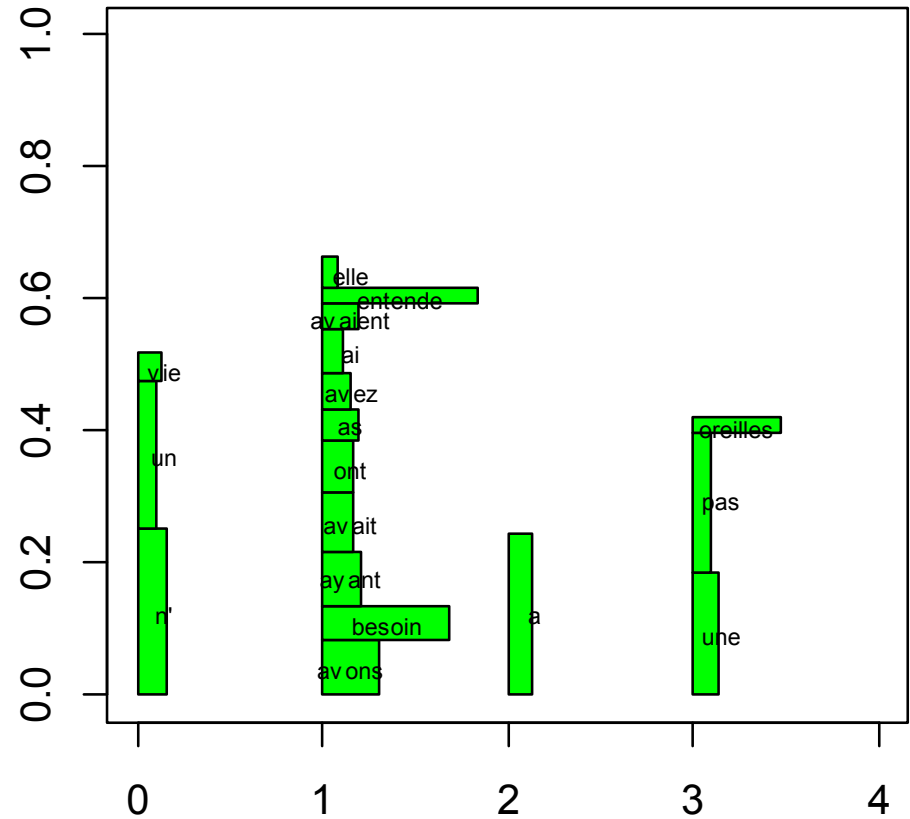
Spanish: [tiene|tenía|teniendo|tienen|tenemos|tenían|...]1 [no|oiga|siete]2 [>#necesi<]3

Lithuanian: [>tur<|reikia]1 [>#ne<|ka]2

Pred. Possession - German



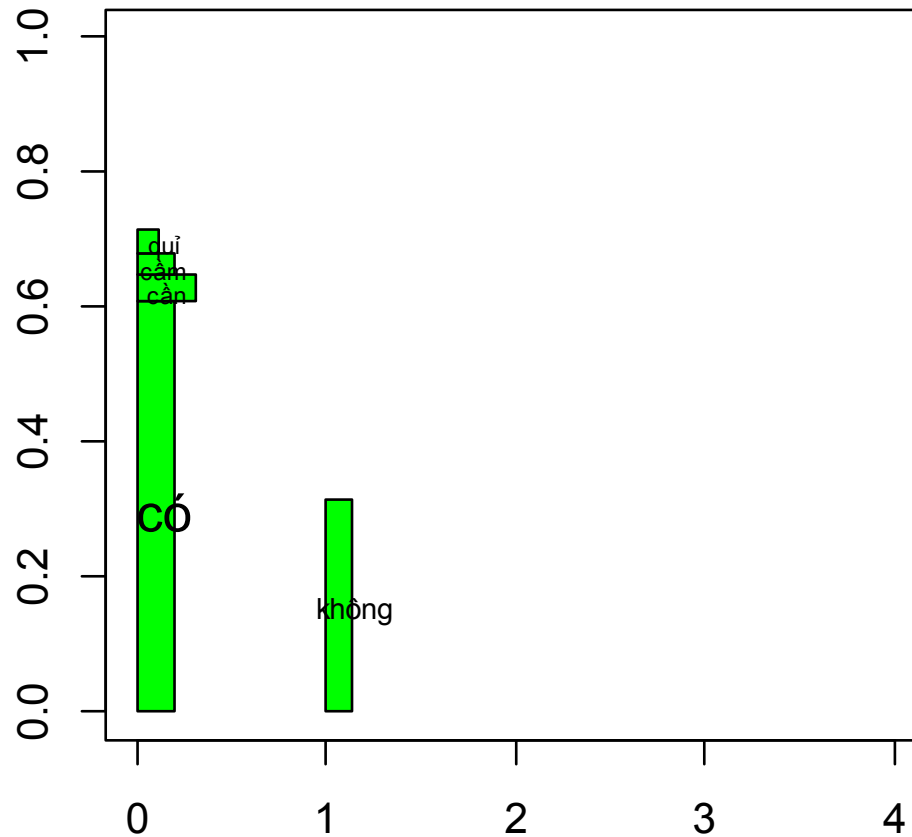
Pred. Possession - French



German: [haben|hatte|hat|habe|habt|hatten|hast]1 [einen|macht]2 [zu]3

French: [n'|un|vie]1 [avons|besoin|ayant|avait|ont|as|avez|ai...]2 [a]3 [une|pas|oreilles]4

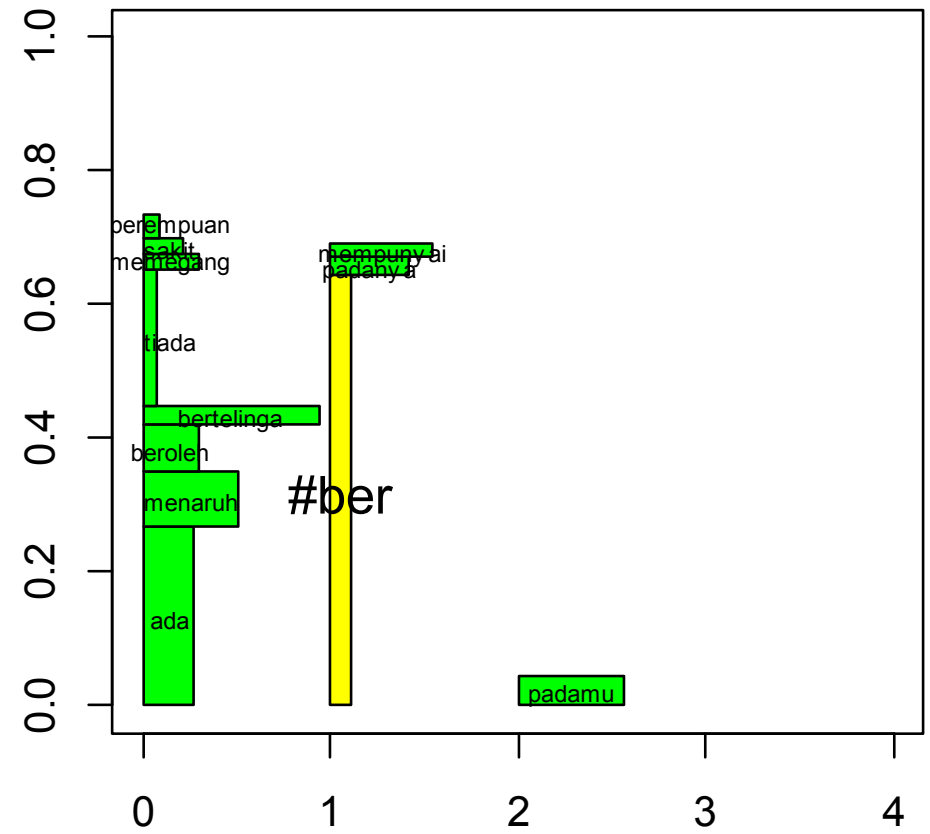
Pred. Possession - Vietnamese



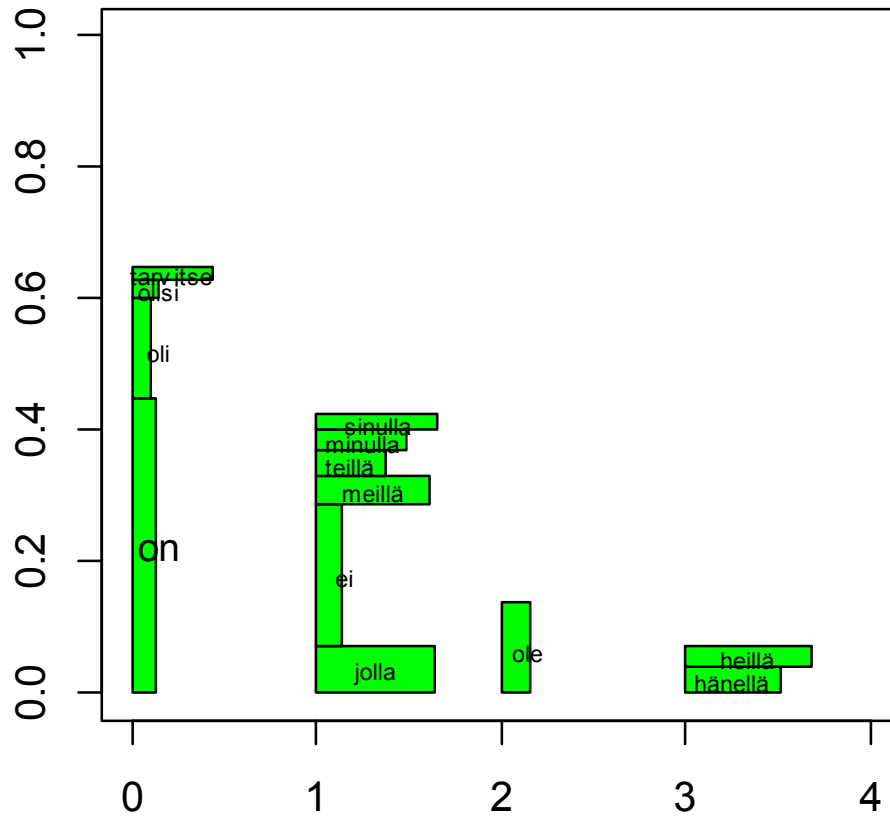
Vietnamese: [có|cần|cầm|quĩ]1 [không]2

Indonesian: [ada|menaruh|beroleh|bertelinga|tiada|...]1 [>#ber<|padanya...]2 [padamu]3

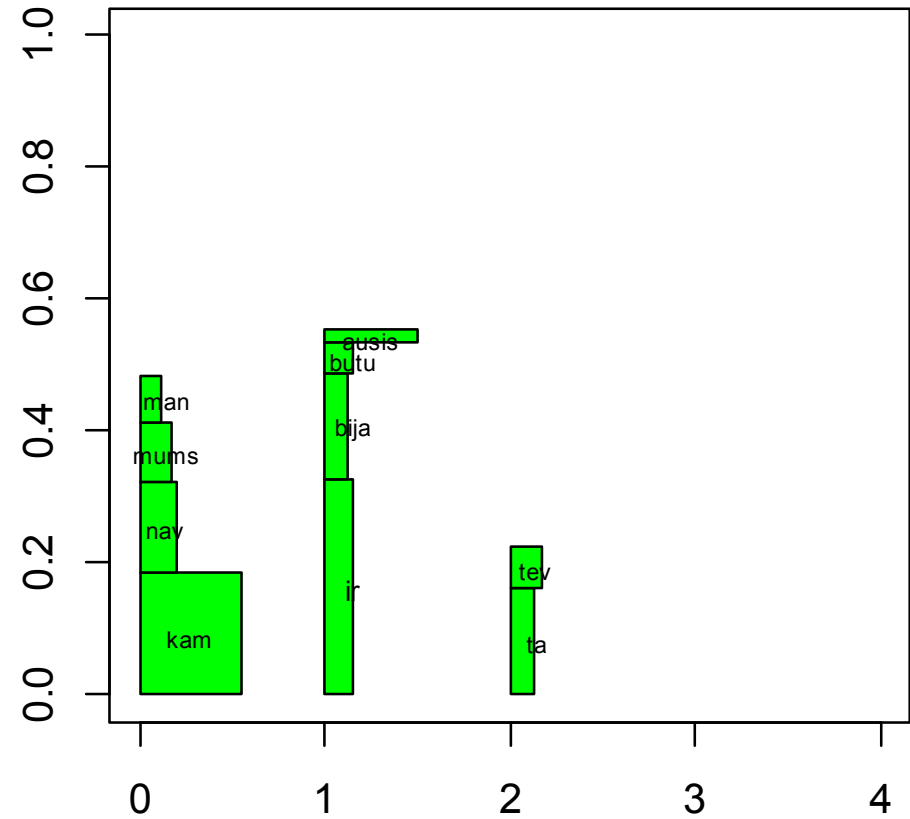
Pred. Possession - Indonesian



Pred. Possession - Finnish



Pred. Possession - Latvian

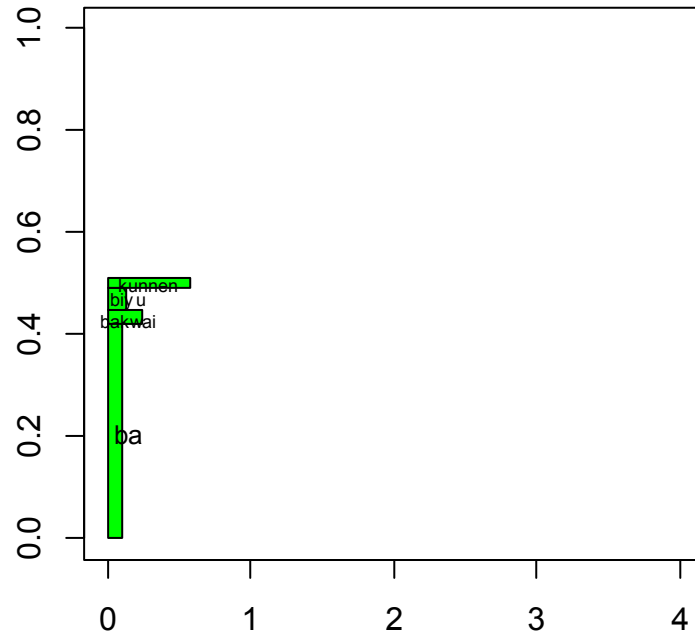


Finnish: [on|oli|olisi|tarvitse]1 [jolla|ei|meillä|teillä|minulla|sinulla]2 [ole]3 [hänellä|heillä]4

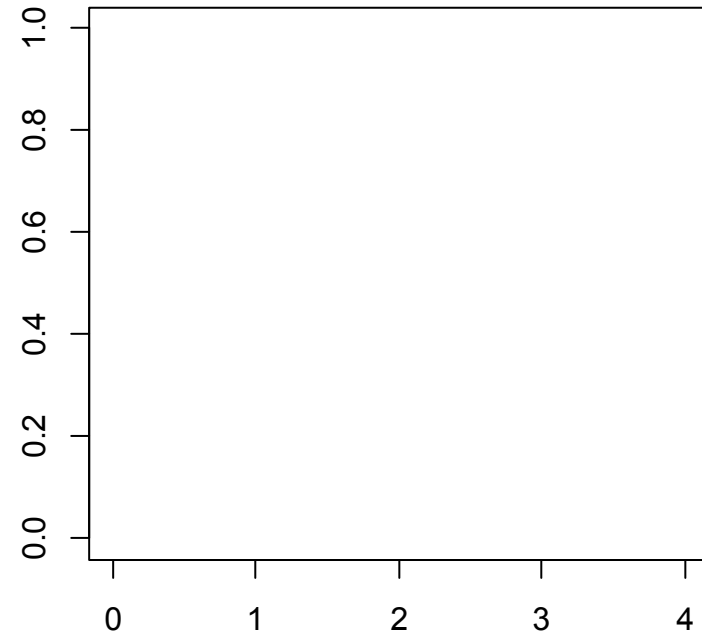
Latvian: [kam|nav|mums|man]1 [ir|bija|būtu|ausis]2 [tā|tev]3

With the 'with' possessive extraction often fails:

Pred. Possession - Hausa



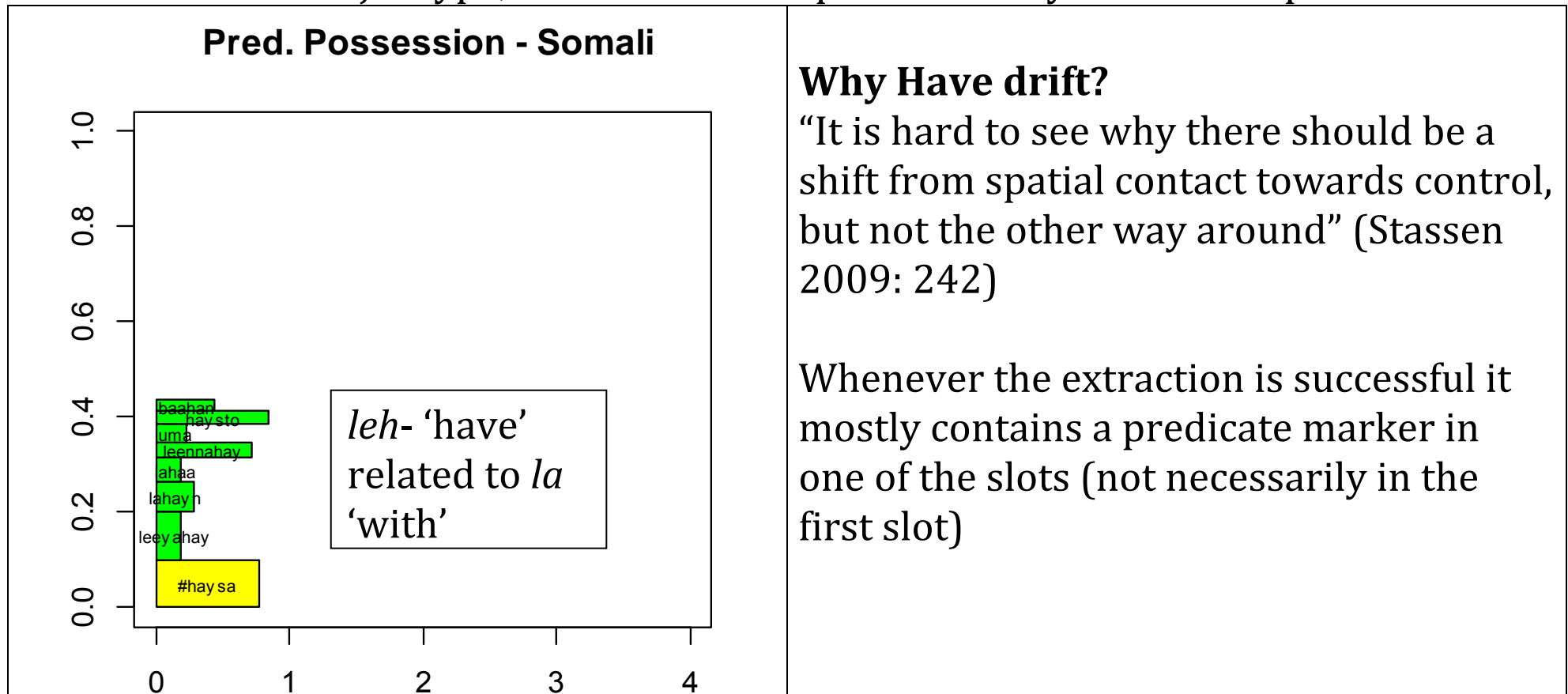
Pred. Possession - Hixkaryana



Hausa: [ba|bakwai|biyu|kunnan]1 'not/seven/two/ear'

Stassen (2009):

- Predicativization: reanalysis of the categorial and syntactic status of the phrase which contains the possessee
- Transitivity/'Have'-drift: if a language starts to reanalyse its possessive in the direction of a major type, the intended output will always be a Have-possessive



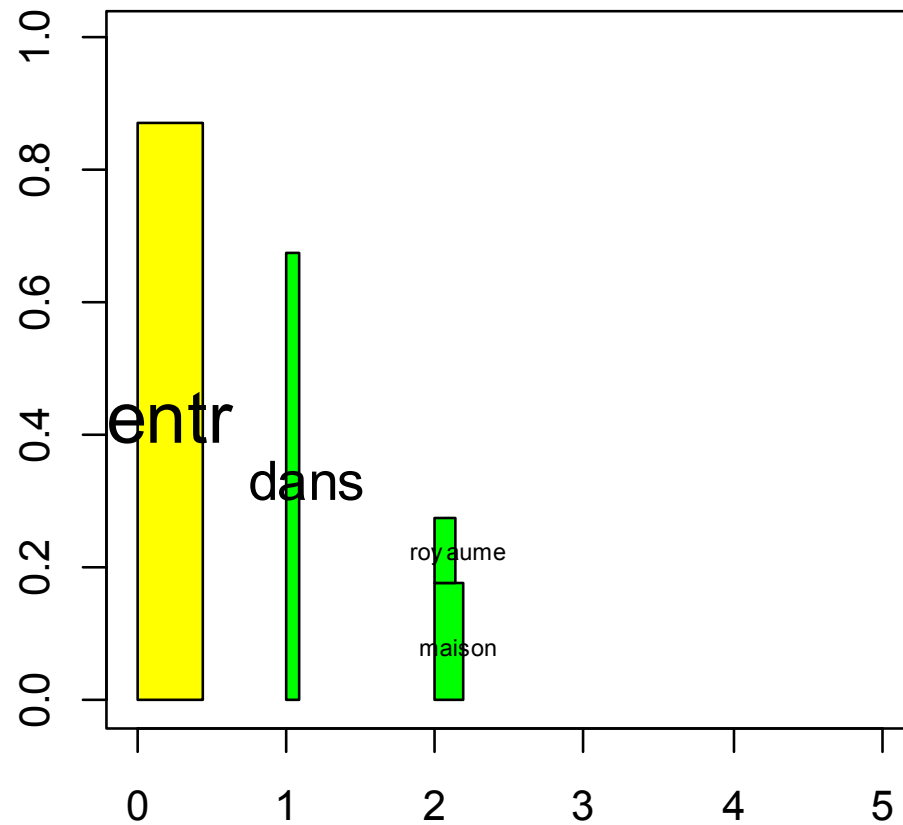
Why Have drift?

“It is hard to see why there should be a shift from spatial contact towards control, but not the other way around” (Stassen 2009: 242)

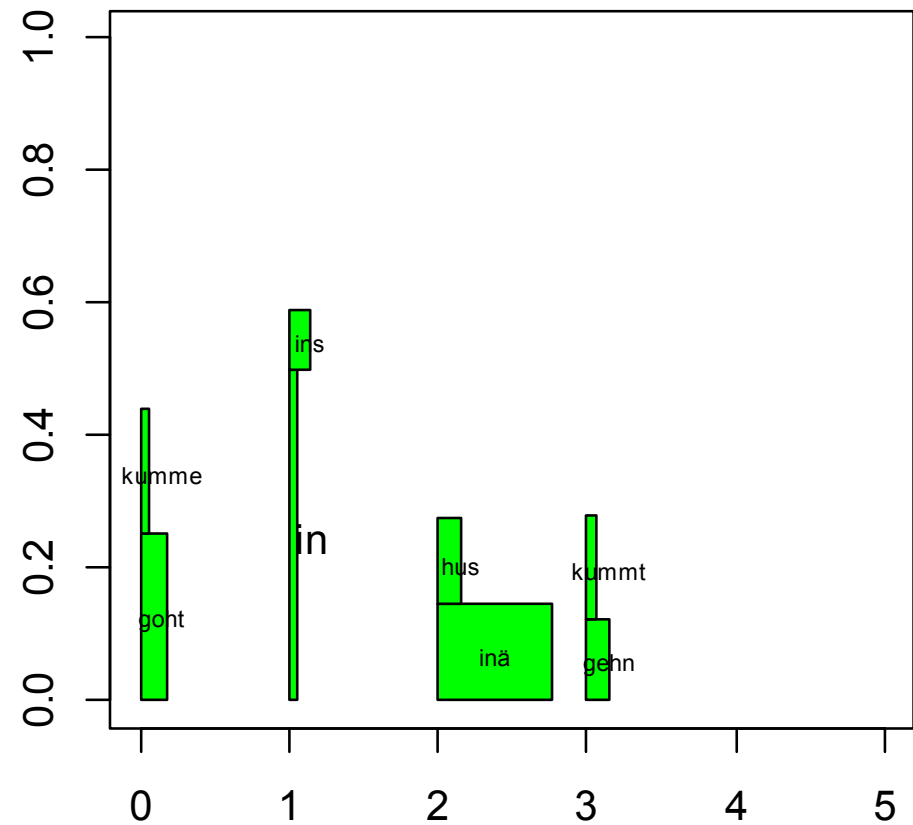
Whenever the extraction is successful it mostly contains a predicate marker in one of the slots (not necessarily in the first slot)

Domain defined by Classical Greek lemma *eiserchomai* 'enter'

Enter - French

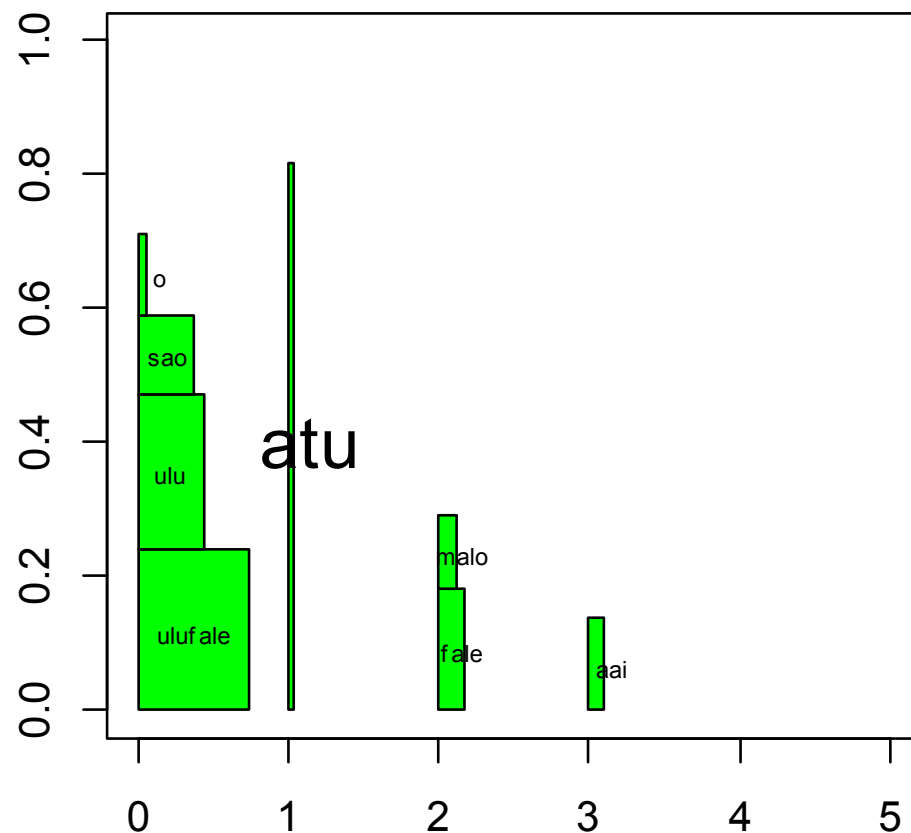


Enter - Alemannic

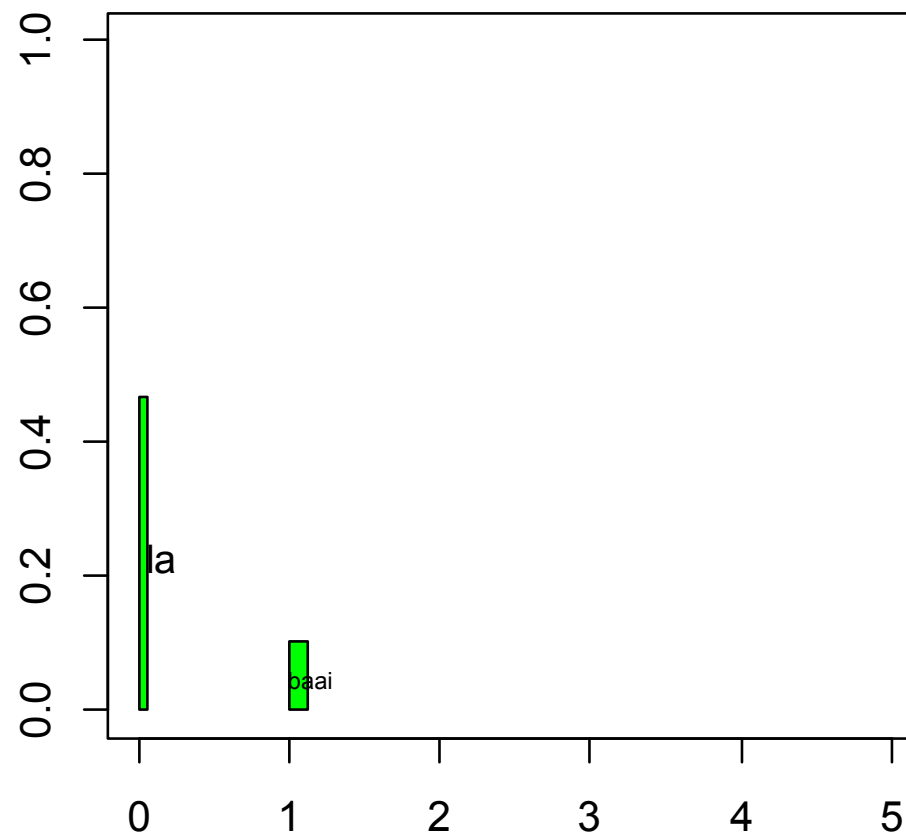


'House' as part of the meaning of 'Enter'

Enter - Samoan



Enter - Dinka



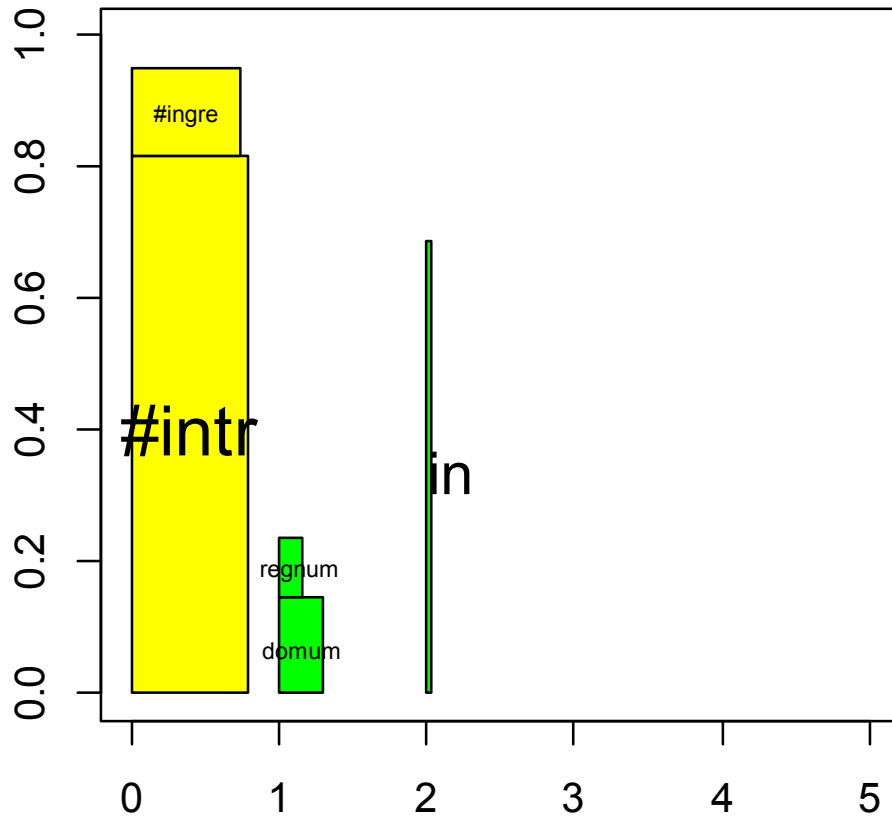
In a convenience sample of 51 languages there is an average of 2.4 slots per language. In 36 languages (42%) there is at least one noun extracted (mostly 'house')

In the vast majority of languages both verbal (V) and adnominal/adverbial (AN/AV) components contribute to the encoding of ‘enter’.

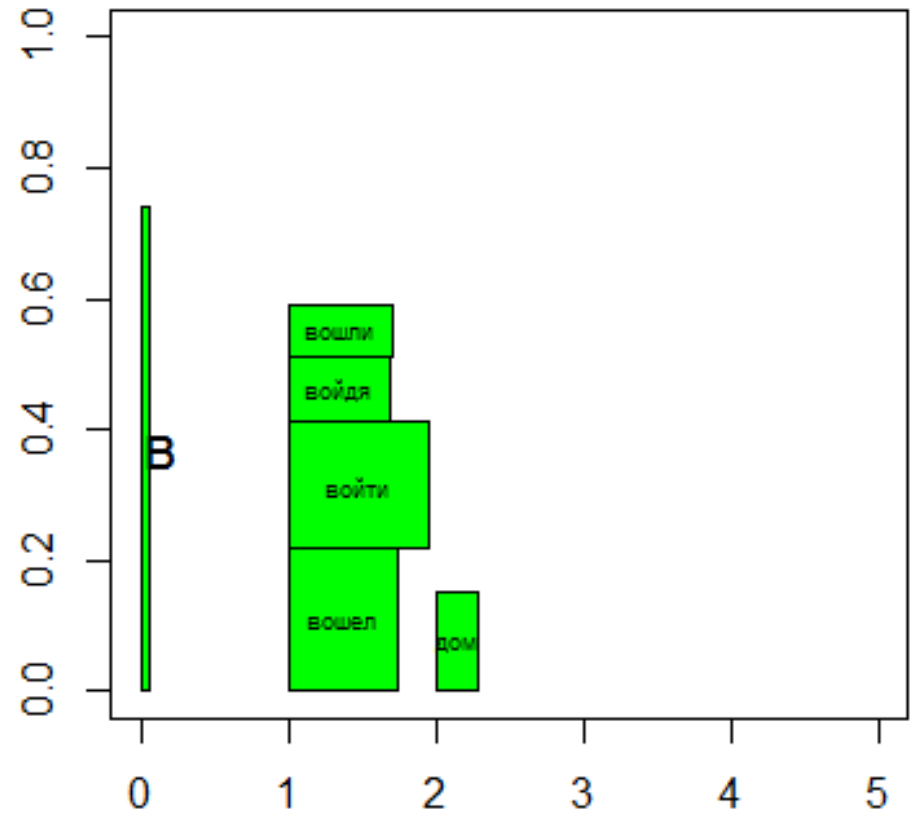
V	Somali, Maltese , Hausa, Ful Adamawa, Vietnamese, Tagalog, Mandarin, Burarra, Yine
V AN/AV	Basque, Kannada, Albanian, Alemannic , Greek (Modern), Hindi, French, Italian, Latin , Portuguese, Romanian, Spanish, Korean, Buriat, Kalmyk, Tatar, Turkish, Finnish , Komi, Mari, Mordvin (Erzya), Swahili, Zulu, Ewe, Wolof, Bambara, Moore, Yoruba, Dinka , Zarma, Haitian Creole, Saramaccan , Maori, Samoan, Indonesian, Malagasy, Lahu, Tobelo, Kuot, Wik Mungkan, Greenlandic (West), Mixe (Coatlán) Otomí (Mezquital), Trique, Paumarí, Quechua (Cajamarca), Aymara
AN/AV V	Avar, Welsh, Danish, German, English, Icelandic, Low Saxon, Norwegian, Swedish, Greek (Classical), Saami (Northern), Tok Pisin, Yanesha’
AN/AV	Latvian, Lithuanian, Croatian, Hungarian, Cakchiquel , Bribri
AN AV&V	Afrikaans, Dutch, Czech, Polish, Russian, Ossetic
V&AV	Mapudungun

boldface: only one slot, all other doculects have two or more slots

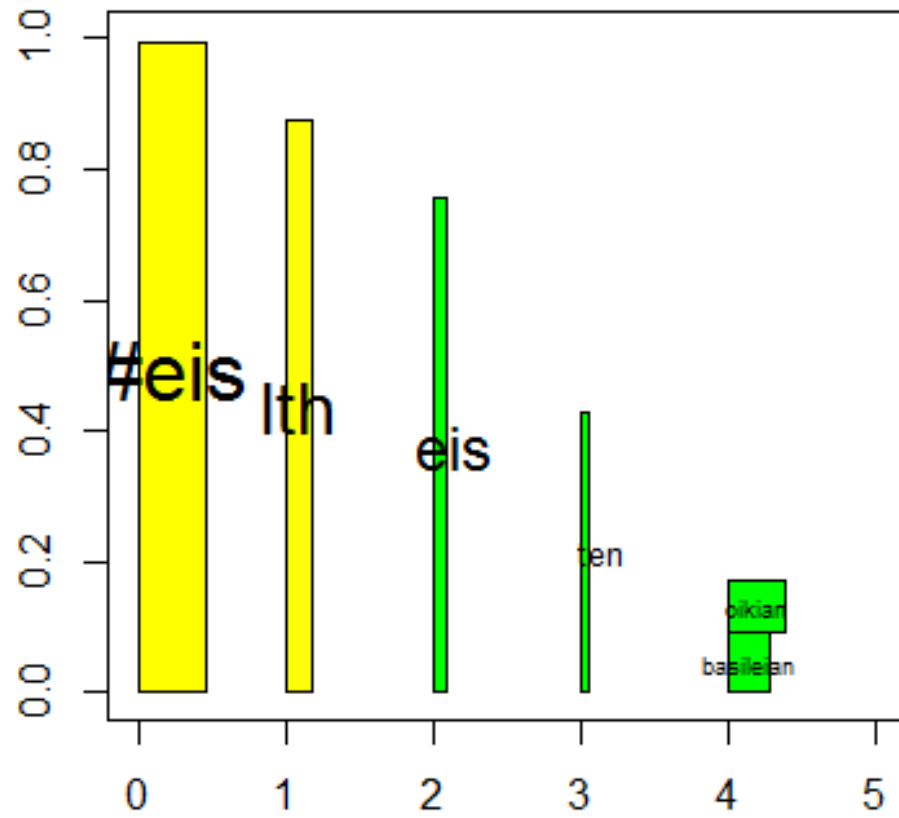
Enter - Latin



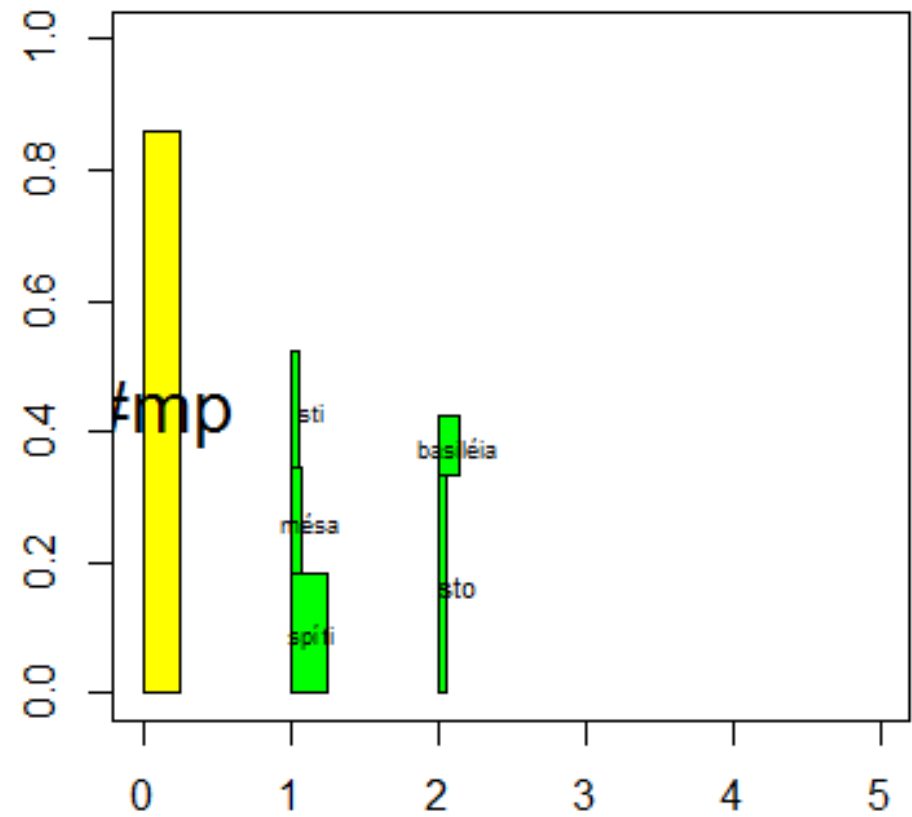
Enter - Russian



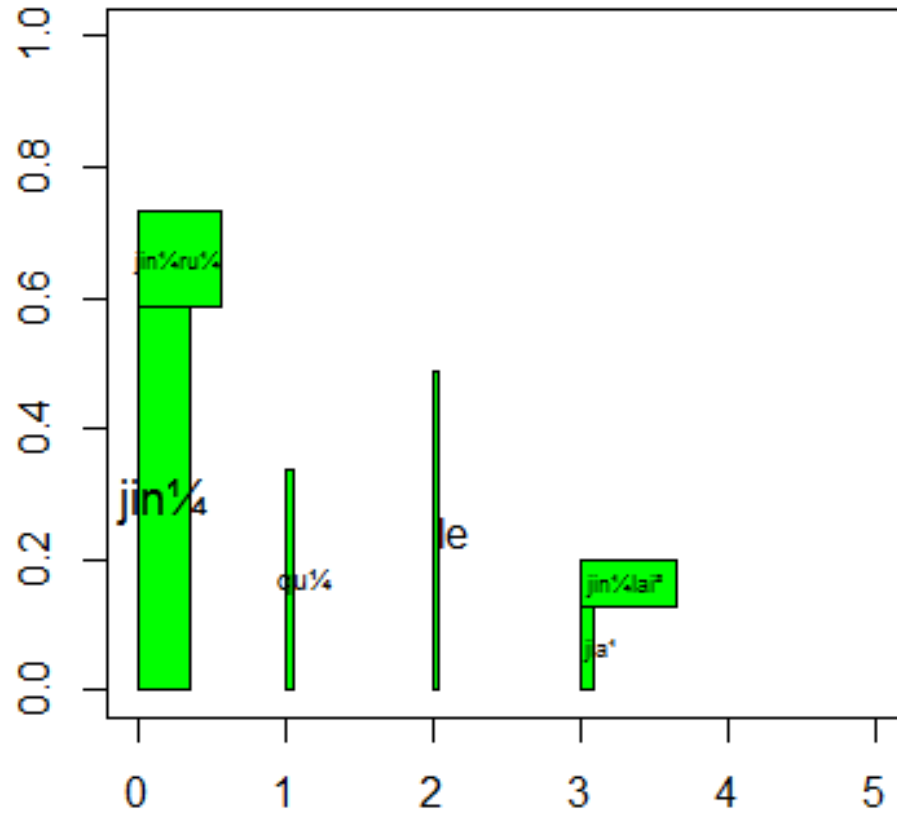
Enter - GreekClassical



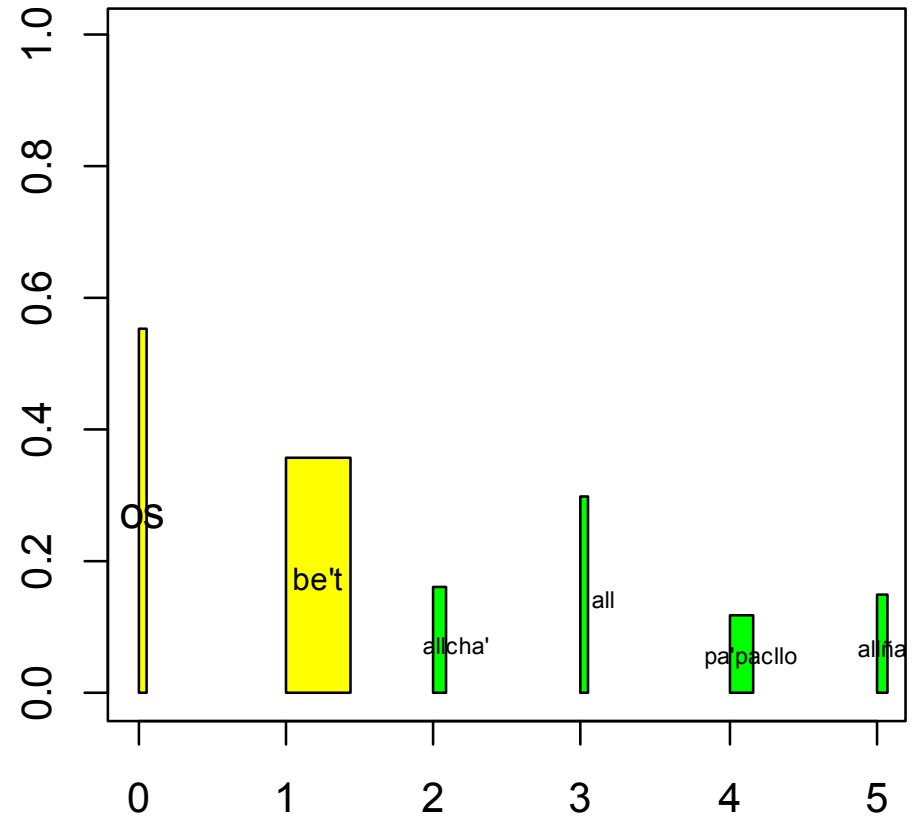
Enter - GreekModern



Enter - Mandarin



Enter - Yanesha



Some marker complexes for Enter: Verbal, Adverbial, Adnominal

Alemannic:	[goht kumme]1 [in ins]2 [inä hus]3 [gehn kummt]4
English:	[into in]1 [entered enter]2 [house kingdom came]3
GreekModern:	[> #μπ <]1 [σπίτι μέσα στη]2 [στο βασίλεια]3
GreekClassical:	[> #εισ <]1 [> λθ <]2 [εις]3 [την]4 [βασίλειαν οικίαν]5
Italian:	[> #entr <]1 [casa ne]2 [nella in]3
Latin:	[> #intr < > #ingre <]1 [domum regnum]2 [in]3
Spanish:	[> #entr <]1 [en]2 [casa reino]3
Russian:	[в]1 [вошел войти войдя вошли]2 [дом]3
Turkish:	[> #gir <]1 [evine içeri eve]2
Finnish:	[meni mennä]1 [sisään sisälle]2 [siellä valtakunta an]3 [tuli sinne]4
Hungarian:	[> #be <]1 [> ba# <]2
Maltese:	[daħal jidħol daħlu]1
Ewe:	[> ge <]1 [me]2 [ɖe]3
Haitian:	[antre]1 [kay wa]2 [kote lakay]3 [nan]4
TokPisin:	[insait]1 [go]2 [haus kingdom]3 [taun]4
Vietnamese:	[vào]1 [nhà]2
Tagalog:	[> asok# <]1 [bahay kaharian]2
Mapudungun:	[> #konp <]1 [ruka mülewe]2

Middle voice (“Reflexive”): trigger Croatian *se* > areal/genealogical effect

Croatian	se	1.0
Polish	się	0.5718
Czech	se	0.55836
Russian	>ся#<	0.52197
Russian	>сь#<	0.51321
Romanian	se	0.42018
Icelandic	>st#<	0.41464
Spanish	se	0.38316
Latvian	>ies#<	0.3582
Afrikaans	word	0.35431
Romanian	s'	0.35357
Italian	si	0.34838
Latin	>ur#<	0.34398
Portuguese	se	0.3165
Alemannic	sich	0.31625
Latvian	>ās#<	0.3152
LowSaxon	sich	0.31472
Danish	sig	0.30998
Swedish	sig	0.30734
German	sich	0.30669
French	se	0.30658

GreekClassical	>θη<	0.30607
Dutch	worden	0.29763
English3	be	0.29591
English	be	0.29451
English2	be	0.29373
Norwegian	sig	0.28876
Portuguese	>-se#<	0.273
SaamiNorthern	>uvv<	0.27032
Albanian	>ohē<	0.27002
Danish	>es#<	0.2656
GreekModern	>ηκ<	0.2586
Welsh	>ir#<	0.24938
Welsh	>#ym<	0.24607
Lithuanian	>si<	0.24275
Dutch	zich	0.23754
Norwegian	>es#<	0.23708
Yoruba	a	0.23662
Malayalam	>pped<	0.23651
Mari	>алт<	0.23095
Swedish	>as#<	0.22728
Hausa	yi	0.22458

Spanish	>se#<	0.22393
Romanian	vă	0.22252
Komi	>ӧдч<	0.21995
Albanian	u	0.21817
Dutch	wordt	0.21636
<i>Papago</i>	e	0.21098
LowSaxon	woare	0.21081
English2	>ed<	0.21043
GreekClassical	>αι#<	0.20949
Finnish	>ty<	0.20837
Lahu	la	0.20774
Hungarian	>ék#<	0.20648
SaamiNorthern	>oj<	0.20549
Alemannic	wird	0.20413
Zulu	>wa#<	0.20181
Tatar	>ыл<	0.20094
English3	was	0.19769
<i>QuechuaCajamarca</i>	>aka<	0.19682
<i>TokPisin</i>	kamap	0.19448
Kalmyk	>гд<	0.1938
English	were	0.19252
Icelandic	sig	0.19147
Norwegian	blev	0.19144

<i>Bribri</i>	e'	0.19057
Dutch	werd	0.19035
French	s'	0.18798
<i>Tagalog</i>	>anag<	0.18769
English3	>#re<	0.1859
<i>Yine</i>	>tka#<	0.18529
Alemannic	wäre	0.18476
Turkish	>nm<	0.18391
Norwegian	bli	0.18309
Swedish	bliva	0.18194
Czech	>no#<	0.18086
<i>Wolof</i>	>iku<	0.17981
Kannada	>iko<	0.17874
Romanian	te	0.17851
<i>Hixkaryana</i>	>os<	0.17817
<i>Kuot</i>	>#te<	0.17683
<i>Yoruba</i>	nigbati	0.17679
Dutch	>ver<	0.17521
English2	were	0.17355
German	werden	0.17214
<i>Somali</i>	la	0.17183
<i>GreenlandicWest</i>	>neqa<	0.17171
Romanian	de	0.17127

Middle voice (“Reflexive”)

Croatian	se	1.0
Polish	się	0.5718
Czech	se	0.55836
Russian	>ся#<	0.52197
Romanian	se	0.42018
Icelandic	>st#<	0.41464
Spanish	se	0.38316
Latvian	>ies#<	0.3582
Afrikaans	word	0.35431
Italian	si	0.34838
Latin	>ur#<	0.34398
Portuguese	se	0.3165
Alemannic	sich	0.31625
LowSaxon	sich	0.31472
Danish	sig	0.30998
Swedish	sig	0.30734
German	sich	0.30669
French	se	0.30658
GreekClassical	>θη<	0.30607
Dutch	worden	0.29763
English3	be	0.29591

English	be	0.29451
English2	be	0.29373
Norwegian	sig	0.28876
SaamiNorthern	>uvv<	0.27032
Albanian	>ohë<	0.27002
GreekModern	>ηκ<	0.2586
Welsh	>ir#<	0.24938
Lithuanian	>si<	0.24275
<i>Yoruba</i>	a	0.23662
Malayalam	>pped<	0.23651
Mari	>алт<	0.23095
<i>Hausa</i>	yi	0.22458
Komi	>өдч<	0.21995
<i>Papago</i>	e	0.21098
Finnish	>ty<	0.20837
Lahu	la	0.20774
Hungarian	>ék#<	0.20648
<i>Zulu</i>	>wa#<	0.20181
Tatar	>ыл<	0.20094
<i>QuechuaCajamarca</i>	>aka<	0.19682
<i>TokPisin</i>	kamap	0.19448

Can all lexical and grammatical meanings be addressed in this way?

No, probably not. Grams with extremely high text frequency (“inflectional categories”), such as plural, adnominal possession (‘genitive’), present, imperfective are difficult to address in most languages.

The present version of the approach is very crude, possible improvements:

- Lexemes and grams as marker candidates instead of word forms and morphs
- Cross-linguistic semantic prototypes as semantic triggers rather than word forms from particular languages instantiating a meaning (Dahl 1985)

There are many practical problems:

- Accidental collocations in a parallel text
- Lexical or grammatical meanings not attested in a parallel text

Is this modern typology?

“Modern typology is a discipline that develops variables for capturing similarities and differences of structures both within and across languages (qualitative typology), explores clusters and skewings in the distribution of these variables (quantitative typology), and proposes theories that explain the clusters and skewings (theoretical typology)” (Bickel 2007: 248)

qualitative > quantitative > theoretical

Here we rather use an inverse processing chain
theoretical > quantitative > qualitative

Theoretical consideration and quantitative analysis come first; the outcome is a descriptive measurement which must be evaluated qualitatively

Cross-linguistic description has been strongly neglected in typology. Description in typology should not be fully outsourced to field linguists, and it should be independent from explanation (data collection in typology should not be explanation-driven)

References

- Ansaldo, Umberto. 1999. Comparative constructions in Sinitic. Areal typology and patterns of grammaticalization. PhD Thesis. University of Stockholm.
- Berthele, Raphael. 2006. *Ort und Weg. Die sprachliche Raumreferenz in Varietäten des Deutschen, Rätoromanischen und Französischen*. Berlin: De Gruyter.
- Bickel, Balthasar. 2007 Typology in the 21st century: major current developments. *Linguistic Typology* 11 (1): 239–251.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
- Davies, John. 1981. *Kobon*. (Lingua Descriptive Studies, 3.) Amsterdam: North-Holland.
- Goddard, Cliff & Wierzbicka, Anna. 2010. ‘Want’ is a lexical and conceptual universal. *Studies in Language* 34(1): 108–123
- Goldberg, Adele E. 2006. *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Grice, Herbert Paul. 1957. Meaning. *Philosophical Review* 66(3): 377-388
- Haspelmath, Martin. 2005. ‘Want’ complement subjects. In Haspelmath, Martin & Dryer, Matthew & Gil, David & Comrie, Bernard (eds.) 2005. *The World Atlas of Language Structures*. (Book with interactive CD-ROM). Chapter 124. Oxford: Oxford University Press.
- Khanina, Olesya. 2008. How universal is ‘wanting’? *Studies in Language* 32(4): 818–865.
- Khanina, Olesya. 2010. Reply to Goddard and Wierzbicka. *Studies in Language* 34.1: 124-130
- Manning, Christopher D. & Schütze, Hinrich. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, Mass: MIT Press.
- Miestamo, Matti. 2005. *Standard Negation: The negation of declarative verbal main clauses in a typological perspective*. (Empirical Approaches to Language Typology 31.) Berlin/New York: Mouton de Gruyter.
- Saussure, Ferdinand de. 1968. *Cours de linguistique générale. Édition critique par Rudolf Engler*. Tome 1. Wiesbaden: Harrassowitz

- Sinha, Chris & Kuteva, Tanja. 1995. Distributed spatial semantics. *Nordic Journal of Linguistics* 18: 167-199.
- Sölling, Arnd. 2011. *Bewegungsverben in Nordamerika - Semantische Elemente in narrativen Texten*. Diss. phil. hist., Universität Bern,
- Stassen, Leon. 1985. *Comparison and Universal Grammar*. Oxford: Blackwell.
- Stassen, Leon. 2009. *Predicative Possession*. Oxford: Oxford University Press.
- Stolz, Christel, Stolz, Thomas. 1996. Funktionswortentlehnung in Mesoamerika. Spanisch-amerindischer Sprachkontakt (Hispanoindiana II). *Sprachtypologie und Universalienforschung (STUF)* 49: 86-123.
- Talmy, Leonard. 1972. *Semantic structures in English and Atsugewi*. Ph.D. dissertation, University of California at Berkeley.
- Talmy, Leonard. 1991. Path to realization: a typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society, February 15-18, 1991*, 480-519.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics. Vol. II: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.
- von Waldenfels, Ruprecht. Forthcoming. Explorations into variation across Slavic: taking a bottom-up approach. In Szmrecsanyi, Benedikt & Wälchli, Bernhard (eds.), *Linguistic variation in text and speech, within and across languages*. To be published in Walter de Gruyter's *Linguae et Litterae* series.
- Wälchli, Bernhard. Forthcoming. Algorithmic typology and going from known to similar unknown categories within and across languages. In Szmrecsanyi, Benedikt & Wälchli, Bernhard (eds.), *Linguistic variation in text and speech, within and across languages*. To be published in Walter de Gruyter's *Linguae et Litterae* series.
- Wälchli, Bernhard & Cysouw, Michael. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50.3: 671-710. (Theme issue edited by Koptjevskaja-Tamm, M. & Vanhove, M. (eds.), *New Directions in Lexical Typology*).
- Wiemer, Björn & Wälchli, Bernhard. 2012. Contact-induced grammatical change: Diverse phenomena, diverse perspectives. In Wiemer, B. & Wälchli, B. & Hansen, B. (eds.), *Grammatical Replication and Borrowability in Language Contact*, 3-64. Berlin: Mouton de Gruyter.

Appendix 1: Extract from the database (Trigger Vietnamese *đã*, 2545 tokens)

No	Domain	Doculect	Slot	Type	Marker	Amplitude	Dedication	Extraction
21	Perfect	German	1	W	hat	0.229971724788	0.648936170213	0.38526
22	Perfect	German	1	W	habe	0.0890669180019	0.549418604651	0.25564
23	Perfect	German	1	W	haben	0.0947219604147	0.463133640553	0.25541
24	Perfect	German	1	W	ist	0.213477851084	0.276556776557	0.23864
25	Perfect	German	1	W	hatte	0.0508953817154	0.421875	0.26517
26	Perfect	German	1	W	war	0.062205466541	0.236135957066	0.22472
27	Perfect	German	1	W	hast	0.0263901979265	0.427480916031	0.22856
28	Perfect	German	1	W	habt	0.0311027332705	0.308411214953	0.23182
29	Perfect	German	1	W	hatten	0.0254476908577	0.409090909091	0.24499
30	Perfect	German	1	W	sind	0.0433553251649	0.159169550173	0.21097
31	Perfect	German	2	M	ge	0.8821866164	0.320877613987	0.21669
32	Perfect	English	1	W	hath	0.123939679548	0.57423580786	0.25716
33	Perfect	English	1	W	had	0.116399622997	0.505112474438	0.24964
34	Perfect	English	1	W	have	0.144203581527	0.398956975228	0.25148
35	Perfect	English	2	W	which	0.16918001885	0.468057366362	0.24137
36	Perfect	English	2	M	ed#	0.538171536287	0.296931877275	0.22233
37	Perfect	English	2	W	made	0.0400565504241	0.291095890411	0.23456
38	Perfect	English	2	W	sent	0.0268614514609	0.322033898305	0.21955
39	Perfect	English	3	W	been	0.0725730442978	0.709677419355	0.22933
163	Perfect	Vietnamese	1	W	đã	1.0	1.0	1.0
164	Perfect	Vietnamese	2	W	cho	0.395852968897	0.356234096692	0.21582...

Appendix 2: R-code written by the Python program generating the visualization of marker complexes

```
plot(c(0,5),c(0,1),col="white",main="Perfect - German",xlab="",ylab="")
slot=0;par=0
ing=0.648936170213;ingg=0.648936170213;ed=0.229971724788;edd=0.3;str="hat"
rect(slot,par,slot+ing,par+ed,col="green")
text(slot+ingg/2,par+ed/2,str,cex=si*edd)
par=par+ed
ing=0.549418604651;ingg=0.549418604651;ed=0.0890669180019;edd=0.3;str="habe"
rect(slot,par,slot+ing,par+ed,col="green")
text(slot+ingg/2,par+ed/2,str,cex=si*edd)
par=par+ed
ing=0.463133640553;ingg=0.463133640553;ed=0.0947219604147;edd=0.3;str="haben"
rect(slot,par,slot+ing,par+ed,col="green")
text(slot+ingg/2,par+ed/2,str,cex=si*edd)
par=par+ed
ing=0.276556776557;ingg=0.3;ed=0.213477851084;edd=0.3;str="ist"
rect(slot,par,slot+ing,par+ed,col="green")
text(slot+ingg/2,par+ed/2,str,cex=si*edd)
...
```