

Caritive markers in the parallel Bible corpus

Kirill Kozhanov, Natalia Perkova

“Caritive constructions in the languages of the world”

November 30, 2020

St Petersburg

Introduction

The massively parallel Bible corpus

- Parallel texts are (ideally) most suitable for language comparison, as they per definition embed identical semantic components (including grammatical semantics).
- Linguistic typology and linguistic diversity require massively parallel texts (Cysouw & Wälchli 2007). The best existing candidate for a (relatively) represented massively parallel text is the Bible corpus (Cysouw & Mayer 2014).

The Bible corpus in typological studies

(Wälchli 2010), (Wälchli & Cysouw 2012): the domain of motion (local phrase markers and motion predicates)

- relatively small samples: 153 and 100 languages (101 doculect); Gospel of Mark — 190 and 360 motion situations
- manual annotation: what marker/verb is used
- probabilistic semantic maps (MDS) after distance matrices between contexts
- Interpretation of clusters in the semantic space

The Bible corpus in typological studies

(Wälchli & Cysouw 2012: 686)

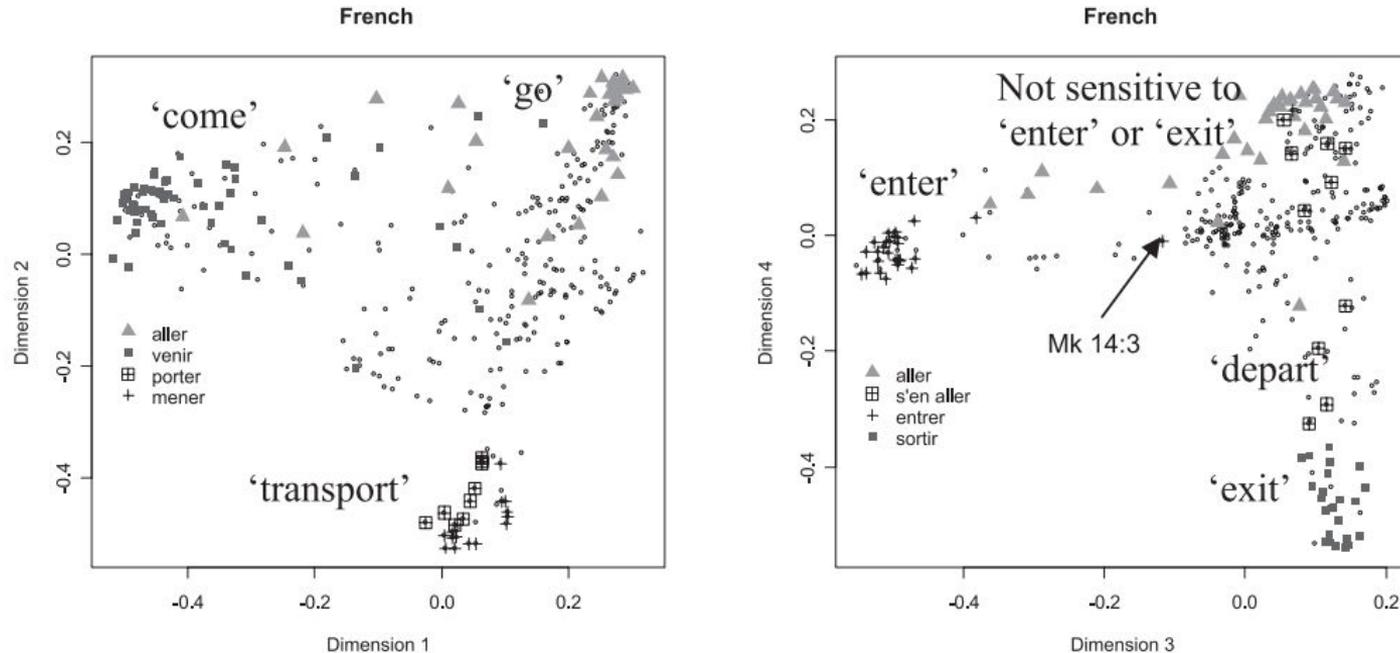


Figure 1. *The first four dimensions illustrated with French categories*

The Bible corpus in typological studies

(Dahl & Wälchli 2016), (Vafaeian 2018), (Dahl 2020): aspectual categories — perfects and iamitives, progressives

- using an automatically word-aligned version of the corpus (NB: an English text as a starting point)
- E.g., (Dahl & Wälchli 2016): 1267 doculects, 1107 languages, 370 grams, union of contexts
- seed grams — preexisting knowledge of good gram candidates
- similarity wrt the generalized distribution of such grams (over the threshold)
- choosing best candidate segments with some filtering rules
- MDS for the union of contexts

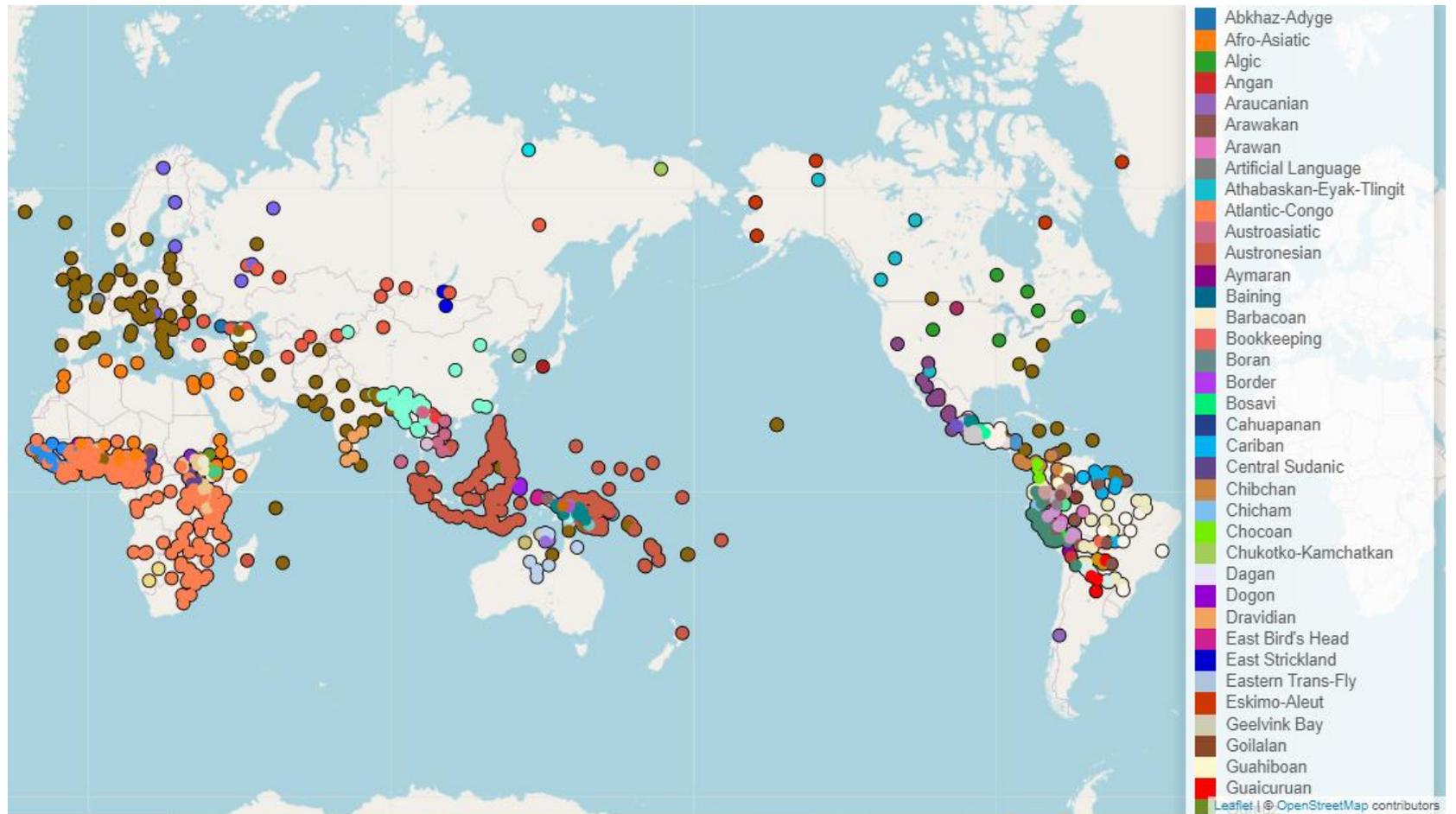
Methodology and data

Data for the current study

1273 languages, 1699 doculects (= translations)

(many thanks to Robert Östling and Murathan Kurfalı for the access to the data)

<https://rpubs.com/ipnos/697350> — the map with the languages represented in the corpus



Methodology

- automatic identification of elements with similar distributions (calculated after the presence/absence of particular strings in particular verses)

Scripts by Robert Östling (<https://github.com/robertostling/parallel-tools>)

```
python find_instances.py -e '^Без | без ' rus | python find_instances.py -a -e '^Sin | sin ' spa | python  
find_instances.py -a -e '^Without | without ' eng | python find_instances.py -a -e ' बिना | bina ' hin | python  
find_instances.py -a -e '^Be | be ' lit | python find_instances.py -a -e '^Bi | bi ' rmy | python  
find_equivalents.py --features=words,bigrams,subsequences --n-best=10 lav
```

— to find equivalents in Latvian for the average distribution (combined contexts) for all 6 languages

— looking for words, bigrams and subsequences, top 10 candidates

Methodology

originally 6 seed grams — only Indo-European

- English *without*, Spanish *sin*, Russian *без*, Hindi *बिना (binā)*, Lithuanian *be*, Romany *bi*
- also checked: only by English *without*, only by Spanish *sin*
- default settings, 10 best contexts

NB: manual checking of the results shows that *sin* as a starting seed marker gives better results (higher coefficients, better interpretability)

Possible explanations: a wider range of “semantic functions” of *sin* in the caritive semantic space; also — lower variability (?)

Prototypical caritive (absence of a companion)

46004008

spa-x-bible-reinavaleracontemporanea.txt

*Ya son ricos , y aun **sin nosotros** reinan .*

eng-x-bible-standard.txt46004008

*Already you have become rich ! **Without us** you have become kings !*

rus-x-bible-synodal.txt

*...вы уже обогатились , вы стали царствовать **без нас***

Circumstance caritive and attributive caritive (absence of a property)

65001012

*...pues **sin vergüenza alguna** comen pensando sólo en sí mismos . Son nubes **sin agua***

*...пиршествуя с вами , **без страха** утучняют себя . Это **безводные** облака , носимые ветром*

*...as they feast with you **without fear** , shepherds feeding themselves ; **waterless** clouds...*

Circumstance caritive (absence of a concomitant situation)

57001014

*pero no he querido hacer nada **sin tu consentimiento***

*но **без твоего согласия** ничего не хотел сделать*

*but I preferred to do nothing **without your consent***

Results

Lists of best caritive contexts

an extension for the scripts, returns tables with the following columns for the languages of the subsample:

- verse id
- number of languages where this verse has a caritive marker at least in 1 doculect
- number of all doculects that have a particular caritive marker in this verse
- the overall number of doculects that have this verse
- the coefficient of caritivity (ratio of 2 previous columns)

E.g., 49002012 6 52 55 0.9454545454545454

— we can ignore high coefficients if the number of languages is low

Lists of best caritive contexts

two top-lists compiled:

- the top 26 contexts with a caritive marker in at least 1 doculect in all 6 langs, the threshold is 0.5
- plus 37 contexts with a caritive marker in at least 1 doculect in at least 5 langs, the same threshold

Lists of best caritive contexts

- a few doculects to choose from
- a wider contextual space (more books than just New Testament)
- difference in the markers checked (here, rather separate words than morphemes like *-less* > the more limited space of contexts)

NB: we can always redefine the procedure by adding more seed markers and update our lists

This is still work in progress, as the further use of top lists requires better understanding of what markers to focus on and check for their presence/absence in the contexts.

Lists of best caritive contexts

eng-x-bible-kingjames.txt

- 49002012** *That at that time ye were **without Christ***
- 50002014** *Do all things **without murmurings and disputings** :*
- 42006049** *...like a man that **without a foundation** built an house upon the earth*
- 58007020** *And inasmuch as not **without an oath** he was made priest*
- 42022035** *And he said unto them , When I sent you **without purse , and scrip , and shoes** , lacked ye any thing ?*

Lists of best caritive contexts

Some relevant Old Testament contexts

18006006

*Can that which is unsavoury be eaten **without salt** ?*

*Едят ли **безвкусное без соли** , и есть ли вкус в яичном белке ?*

18024007

*They cause the naked to lodge **without clothing** , that they have no covering in the cold .*

*нагие ночуют **без покрова и без одеяния на стуже** ;*

Top candidates for caritive markers

A list of counterparts to the seed sample

- Using the initial seed marker set, we searched for the counterparts to the caritive markers that have most similar distributions to the generalized distribution over our seed markers.
- We crawled lists of 10 most likely counterpart candidates for each translation with scores representing the similarity between the markers.
- The united distribution over 6 markers is skewed towards very specific collocations, e.g., ‘without + fault’, ‘without + sheep’.
- Therefore we also crawled lists for English *without* and Spanish *sin* as seed markers; the latter seems to give more adequate results.

One of the Portuguese translations (by 6 langs)

por-x-bible-almeidaatualizada.txt

45.63 #cordeir#

604.85 sem

43.61 #cordeiro#

174.53 ,_sem

43.45 um_ano

102.31 #defeito#

35.92 #cordeiros#

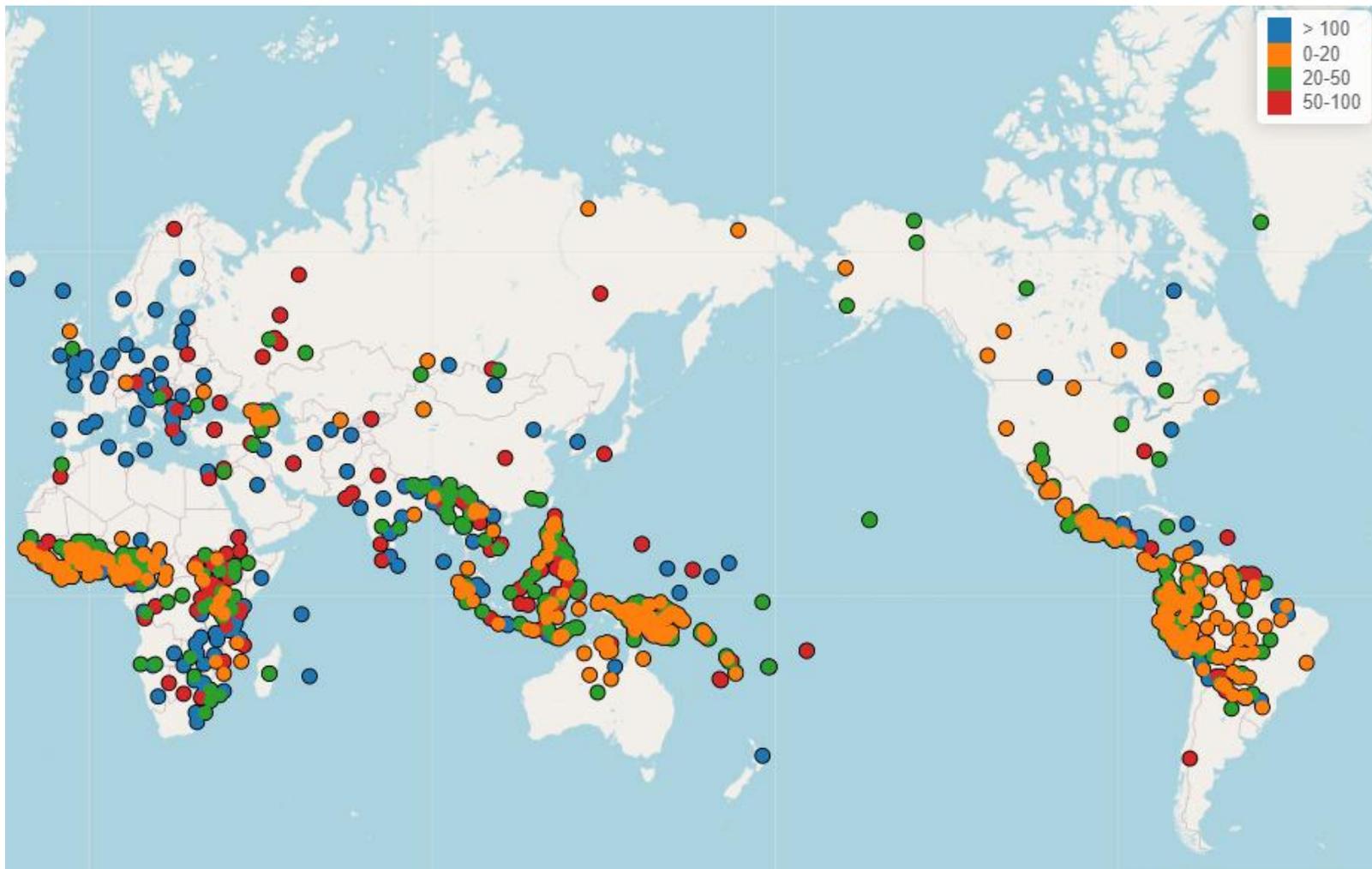
102.31 defeito

35.92 cordeiros

47.11 ano_,

Top candidates

- As a candidate threshold we chose to have the log transformed scores higher than 50, being an arbitrary choice after manual inspection of the data. It seems that similarity scores around 100 and above do help to identify valid candidates for caritive contexts.
- Map 2 illustrated the distribution of the doculects divided into groups by the log transformed score (find an interactive version here: <https://rpubs.com/ipnos/697442>)



Discussion

Caritive typology in the Bible texts

What we see in our results is first of all just indirect evidence on the similarity of caritive gram distributions.

An obvious question: what else can we get from such data?

- morpho(phono)logical properties (not much, really)
- semantic properties (only indirectly)
- syntactic properties (not really with unannotated data or without manual inspection)

Caritive typology in the Bible texts

Morphophonology

- we can extract strings, but they are not combined into variants as allomorphs
- also tricky for affixes and roots; NB: check the search for suffixes?
- Lamba (Bantu < Atlantic-Congo), the top-10 by Spanish *sin* , cf. Bemba, *-bul-* 'lack'

143.14 #akuwul#

91.17 #ulo'#

75.49 #kuwulo'#

67.33 #wakuwul#

65.18 #akuwulo'#

46.52 #apakuwul#

46.40 #apakuwul#

45.61 #ulo'k#

43.13 #kuwulo'k#

40.19 #wakuwulo'#

Caritive typology in the Bible texts

Semantics

- rather properties of absentees (animacy, referentiality, functions) and the contexts in general
- compatibility
- e.g., annotate most typical caritive contexts

still rather potential future work (hopefully in collaboration with the project team)

Caritive typology in the Bible texts

Compatibility

- with proper nouns (49002012 ‘without Christ’)
- with personal pronouns (46004008 ‘without us’)
- with other pronouns
- with kinship terms (58007003 ‘without father, without mother’)

NB: also contexts with standard negator / other negators / comitatives in adjacent clauses

Partly realized by the project team; we have a clear advantage: more languages, longer texts for some of them (not only Gospels, sometimes also Old Testament books)

Caritive typology in the Bible texts

Marker properties (~ grammatical semantics)

- specialized (dedicated)
- CAR = NEG.EX
- CAR = NEG.POSS
- CAR = NEG (standard)
- other variants, e.g., 'to lack' (a lexicalised verbal root; no formal negation)

Negative possessive

#lawore# in Vunjo (Bantu < Atlantic-Kongo)

lit-x-bible-ecumenical.txt

76.09 be

37.47 #netur#

22.24 #eturi#

NEG.EX = NEG.POSS

E.g., (Southern) Altai language (< Turkic): **joк**

Candidates: NEG copulas (Russian, Latvian, Sami), but both *néra* (NEG.be) and *netur#* (NEG.have) in Lithuanian; caritive markers; NB: no standard negators in the top list

Negation (vague)

teke in Bulu (Cameroon) < Atlantic-Congo

by SIN:

269.68 teke

43.74 ne_te

check against Russian, Lithuanian, Sami

= different types of negative polarity elements (NEG, 'nobody'), 'without', 'but'

Negation (vague)

Bulu (bum) < Bantu < Atlantic-Congo

rus-x-bible-synodal.txt

245.73 и_не

96.71 ни

92.97 #нет#

86.73 #кто#

83.94 нет

71.31 ,_не

51.41 #без#

39.84 #ник#

34.36 #нич#

33.35 но

sme-x-bible.txt

136.44 ii

98.00 #tage#

94.79 oktage

93.48 #ktage#

93.05 #oktag#

66.41 ge

52.59 ii_ge

41.81 eai

25.96 #ege#

22.53 #hto#

lit-x-bible-ecumenical.txt

16.53 #niek#

9.91 #nieka#

8.63 #niekas#

8.63 niekas

2.09 #nep#

-1.32 #jok#

-1.40 #septyni#

-1.66 #negalè#

-1.75 #etur#

-2.25 #okio#

Negation (vague)

Bulu (bum) < Bantu < Atlantic-Congo

rus-x-bible-synodal.txt

245.73 и_не

96.71 ни

92.97 #нет#

86.73 #кто#

83.94 нет

71.31 ,_не

51.41 #без#

39.84 #ник#

34.36 #нич#

33.35 но

sme-x-bible.txt

136.44 ii

98.00 #tage#

94.79 oktage

93.48 #ktage#

93.05 #oktag#

66.41 ge

52.59 ii_ge

41.81 eai

25.96 #ege#

22.53 #hto#

lit-x-bible-ecumenical.txt

16.53 #niek#

9.91 #nieka#

8.63 #niekas#

8.63 niekas

2.09 #nep#

-1.32 #jok#

-1.40 #septyni#

-1.66 #negalè#

-1.75 #etur#

-2.25 #okio#

(Abomo-Maurin 2006: 181): tèkè(è) ‘non, ce n’est pas cela, sans’

Negation (vague)

Hiligaynon (Austronesian)

112.73 nga_wala 112.32 wala_sing

Again, a range of negative polarity elements returned, e.g., ‘noone’, ‘be.NEG’, *-less* (Icelandic *-laus-*, general negators)

Hiligaynon-English Dictionary (Kaufmann 1934), <http://www.bohol.ph/kved.php?sw=wala&where=hw>

walâ, No, not, none, no one, not any, nothing; there is not, does not exist, there has not been (existed); to be not, have not, be a nonentity, to lack, be deprived of, be not there, to disappear, be gone.

walâ sing, Nothing, nought, naught, not any, not anything, none, nobody, no man, no one, not a single person, not a soul; there is (was) not, nothing, etc.; to lack, be without, have not, nothing, etc.

Negation (standard)

Vietnamese **không**: top-candidates = standard negators (Russian **не** with very high scores, over 1000, Latvian and Lithuanian **#ne#**, Sami **ii**), also 'but'

NB: also **không có** 'NEG + have' (Rus. synodal **#имее#**, **#имею#**, Lith. **#netur#**)

Welsh: dedicated or NEG?

the preposition *heb*, apparently caritive in its core function

But: negative perfects ('I haven't slept' = 'I am without sleeping') and other negative constructions

also returns NEG.EX / NEG.POSS and other negators as its candidates; likely to be used with predicate roots (cf. Sámi verbal abessive *-keahttá*)

Originally dedicated > other negative contexts - how to check???

Problems

- Allomorphy
 - different allomorphs can be extracted separately, but they are not identified as the representatives of the same element
- Discontinuous marking
 - consecutive elements can be extracted as bigrams, but not discontinuous
- Collocational attraction
 - some lexemes that often co-occur with caritive markers in the corpus can get extracted as candidates
- Homonymy (homography)
- Non-dedicated markers which get special morphology in typical caritive contexts, what to choose and how to classify?
 - Tagalog walang = wala=ng ‘NEG.EX=LINK’; NEG.EX = wala
- Also in general: the quality of the corpus (verse alignment, quality of translations and variability, lack of normalization, some mess with diachronic version, etc.)

Bonus: Case studies

Case study 1: Creoles and conlangs

- Interpretability of the results
 - comparing to the lexifiers: do markers get borrowed?
 - distributions of caritive markers in the language and its lexifier
- Higher similarity usually expected
 - non-specialized strategies (e.g., NEG, NEG.EX/NEG.POSS), if dedicated marking - then rather borrowed?
 - more transparent form-meaning correspondences, cf. the analysis of transparency by Leufkens (2013) - “a consistent one-to-one relation between meaning and form”

2 conlangs (Esperanto, Klingon), 24 creoles and pidgins

Cf. (Stolz, Stroh & Urdze: 103-104) - 11 creoles, the focus on dedicated borrowed markers

iso code	language	source lg	six markers	sim 6	without	sim eng	sans	sim fre	sin	sim spa		
epo	Esperanto	artificial	sen #sendifekta#	218.24 133.12	sen #sendifekta#	261.99 73.81	sen #difekt#	210.06 125.46	#sen# #dife#	290.10 115.25		
tlh	Klingon	artificial	Hutlh blemish	649.76 143.63	Hutlh blemish	765.76 82.69	Hutlh	632.50	Hutlh	609.47	'to lack'	collocation
acf	Saint Lucian Creole French	French			san	40.01	san pyès	98.35 16.62	san	91.03	NB: no positive results for 6 langs	
bis	Bislama	English	no_mekem	7.94	no_mekem	8.62	no gat	30.36	no_gat no_mekem	40.41 11.37		
bjz	Belize Kriol English	English	widowt	60.24	widowt	50.79	widowt no	99.27 38.05	widowt no	110.40 25.63		
cbk	Chavacano	Spanish	sin	10.01	sin	17.11	sin	47.66	sin	58.83		
crs	Seselwa Creole French	French	defo _san	134.55 71.92	san defo	340.05 75.79	san	940.68	san okenn	952.16 76.37		
djk	Aukan	English, Dutch	sondee_fu	2.75	sondee_fu _sondee	6.24 3.54	sondee sondee_fu	109.49 35.57	sondee sondee_fu	105.33 28.76		
gul	Sea Island Creole English	English	dout got_no	7.25 1.29	got_no	2.45	hab_no aim_hab got_no	24.98 17.44 7.99	ain_hab hab_no	28.06 27.97		
hat	Haitian	French	enfimite san_okenn _san	131.35 76.61 46.68	enfimite san_okenn _san	71.36 56.94 40.83	san	527.10	san	569.76		
hmo	Hiri Motu	Motu, Tok Pisin	bero	77.59	bero	51.17	bero #tubua#	92.60 108.49	#ubua# paraoa	64.30 42.62		
hwc	Hawai'i Creole English	English			an_no	3.97	an_no notting neva	16.19 12.39 11.43	an_no neva notting	12.73 24.71 12.93	NB: no positive results for 6 langs	
icr	San Andres Creole English	English	bitout gat_non	45.54 2.48	bitout gat_non	37.24 8.63	bitout non gat_non	81.77 32.31 24.21	bitout non gat_non	94.84 28.30 18.43		
jam	Jamaican Creole English	English	an_no widout	11.36 4.87	an_no widout	16.58 widout	an_no notn neva	26.11 17.44 3.97	an_no neva	28.89 10.29		

kri	Krio	English			klin	0.20	we_no	16.55		13.30	NB: no positive results for 6 langs
							no_get	3.79		4.48	
ktu	Kituba	Kikongo	ya_konda	19.84	ya_konda	19.22	konda	113.88	konda	83.64	
			kondaka	kondaka	kondaka	2.91			ya_konda	37.40	
lin	Lingala	Bantu	ezangi	133.84	ezangi	105.69	ezangi	234.22	#zang#	402.65	
			mbeba	117.16	kozanga	76.70	mbeba	137.08	ezangi	212.31	
mbf	Baba Malay	Malay,Indone	dngan_t'ada	98.27	dngan_t'ada	89.24	dngan_t'ada	206.60	dngan_t'ada	172.69	Malay dengan tiada
mfe	Morisyen	French	san	72.33	san	72.12	san	208.36	san	212.04	
pap	Papiamento	Portuguese	_sin	7.35	sin	146.67	sin	263.33	sin	262.42	
pcm	Nigerian Pidgin	English	nor_get	34.12	nor_get	39.37	nor_get	60.37	nor_get	62.07	
			witout	11.62	witout	10.20					
pis	Pijin	English	no_garem	9.56	no_garem	40.45	no_garem	135.78	no_garem	153.99	not have
rop	Kriol	English	nomo_garram	110.07	nomo_garram	71.20	nomo_garram	185.60	nomo_garram	207.76	
sag	Sango	Ngbandi	daa_pepe	6.64	daa_pepe	9.36			ni_pëpe	10.86	
srm	Saramaccan	English, Portuguese			söndö	23.72	söndö	78.13	söndö	83.77	NB: no positive results for 6 langs
							föutu	7.31			
srn	Sranan Tongo	Dutch	sondro	109.04	sondro	126.67	sondro	345.29	sondro	342.32	sondro = without
			mankeri	104.49	mankeri	64.63	mankeri	111.76	mankeri	101.99	mankeri = harm, wound, lack, be missing
			abi_nowan	91.46	abi_nowan	51.67	abi_nowan	102.25	nowan	81.24	
tcs	Torres Strait Creole	English, Other	meke_sin	1.50	meke_sin	0.45	nogad	44.99	nogad	41.38	
			dempla_no	1.39							
tpi	Tok Pisin	English	sua_samting	145.69	sua	52.26	sua	107.47	no_gat	222.92	
			sua	103.40	skin_bilong	34.01	skin_bilong	63.70	skin_bilong	56.45	sua=wound
			bagarapim_ski	125.06	i_bagarapim	32.71	i_bagarapim	59.10			bagarapim=spoil
					skin_bilong	34.01	no_gat	28.55			gat=have

Case study 1: Creoles and artificial languages

- Again, better interpretability with seed grams from the Romance languages (Spanish/French)
 - data less skewed to the specific collocations;
 - rather the Spanish type than the English one ('without' and *-less* rather covered by the same marker)? NB: **Belize Kriol English** - both *widowt* and *no*
- Interpretability of the markers by their form
 - possible to identify the sources in most cases: *widout* [bɜzj, jam], *dout* [gul] *bitout* [icr], *without* [pcm] (< English *without*), *san* [crs, hat, mfe] (< French *sans*), *sin* [cbk, pap] (< Spanish *sin*), *sondee* [auk] and *sondro* [srm] (< Dutch *zonder*). In Baba Malay, *dngan t'ada* clearly corresponds to the Malay caritive *dengan tiada*, which is very interesting, as it is a complex marker (COM + NEG.EX). Kituba (*konda*) has a clear correspondence in its lexifier Kikongo (*kondwa*);
 - dedicated markers (borrowed);
 - negators, 'to have': *nor get* [pcm], *no gat* [bis, tpi], *no gət* [kri], *no garem* [pis], *got / hab no* [gul], *abi nowan* [srn], *nomo garram* [rop] = NEG + COM (COM ~ POSS ~ EX)

Case study 2: Turkic

- Three major types of relevant patterns:
 - caritive (privative) affixes **-slz** (most likely within the attributive domain)
 - negative existentials, cf. Turkish **yok**
 - negative converbs, cf. Old Turkic **-ma-Di[n]** (Johansson 1995: 337), Turkish **-mAdAn**

Cf. Turkish:

46004008: **biz olmadan** = 1PL be:NEG.CV ‘without us’

65001012: **yağmursuz** = rain:CAR ‘waterless’, **meyvesiz** = fruit:CAR
‘fruitless’

57001014: **senin rızanı almadan** = 2SG.POSS content take:NEG.CV
‘without your consent’

Case study 2: Turkic

- Three major types of relevant patterns:
 - caritive (privative) affixes **-slz** (most likely within the attributive domain)
 - negative existentials, cf. Turkish **yok**
 - negative converbs, cf. Old Turkic **-ma-Di[n]** (Johansson 1995: 337), Turkish **-mAdAn**

21 language in the corpus

<i>Subgroup</i>	<i>Language</i>	<i>1st preferred</i>		<i>2nd preferred</i>	
Oghuz	Azerbaijani	CAR	#siz#, #suz#	NEG.CV	#mada#, #mada#
	Gagauz	CAR	#суз#, #сыз#		
	Turkish	CAR / NEG.CV	#siz#, #suz# #madan#, #meden#		
	Turkmen	CAR	#syz#	NEG.CV	#mān#
West Kipchak	Crimean Tatar	CAR	#сыз#	NEG.CV	#мейип#, #майып#
	Karachay-Balkar	NEG.CV	#май#, #мей#	CAR	#суз#, #сыз#
	Kumyk	CAR	#сыз#, #сюз#	NEG.CV	#магъан#, #мей#
South Kipchak	Kara-Kalpak	CAR	#сыз#, #нсиз#		
	Kazakh	CAR	#нсиз#, #сыз#		
	Nogai	CAR	#сыз#	NEG.CV	#май#, #мей#
East Kipchak	Kirghiz	CAR	#сиз#, #сыз#	NEG.EX	жок
	Altai	NEG.EX	жок		
North Kipchak	Bashkir	NEG.CV	#майенса#, #майынса#	CAR	#һыз#
	Tatar	NEG.CV	#миҗа#, #мыйча#		
Bulgar	Chuvash	CAR	#сёр#, #сър#	NEG.CV	#месёр#, #асёр#, #масёр#
Turkestan Turkic	Uighur	NEG.CV	#май# - #mey# #май# - #mey#	CAR	#сиз# - #siz#
	Uzbek	CAR	#nsiz#		
South Siberian Turkic	Khakas	NEG.EX	чох		
	Shor	-	-		
	Tuvan	NEG.EX	#чок#, чок		
North Siberian Turkic	Sakha	NEG.EX	суох		

References

- Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. Introduction to the special issue of *Sprachtypologie und Universalienforschung (STUF)* 60(2). 95-99.
- Dahl, Ö. 2020. Indo-European perfects in typological perspective. In Robert Crellin, Thomas Jügel (eds.) *Perfects in Indo-European languages and beyond*. John Benjamins, 636–667.
- Dahl, Ö., Wälchli, B. 2016. Perfects and iamitives: two gram types in one grammatical space. *Letras de Hoje* 51(3), 325–348.
- Mayer, Th., Cysouw, M. 2014. Creating a massively parallel Bible corpus. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 3158–3163 (http://www.lrec-conf.org/proceedings/lrec2014/pdf/220_Paper.pdf)

References

- Leufkens, S. 2013. The transparency of creoles. *Journal of Pidgin and Creole Languages* 28(2), 323–362.
- Stolz, T., C. Stroh & A. Urdze. 2007. WITH(OUT): On the markedness relations between comitatives/instrumentals and abessives. *Word* 58(2). 63-122.
- Vafaeian, G. 2018. Progressives in use and contact: A descriptive, areal and typological study with special focus on selected Iranian languages. Ph.D. thesis, Stockholm university.
- Wälchli, B. & Cysouw, M. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3), 671-710.
- Wälchli, B. 2010. Similarity semantics and building probabilistic semantic maps from parallel texts. *Linguistic Discovery* 8(1). Available at:
<https://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/1/article/356>