

Cross-linguistic variation in the expression of incrementation and related functions

Barend Beekhuizen

Guillaume Thomas

Abstract Words like English *more* and German *noch* have incremental uses, which express addition to a previously asserted quantity. Incrementation is commonly colexified with comparison and with continuation, and Thomas (2018), on the basis of a 12-language sample, observed a gap in the attested patterns: comparison and continuation are never colexified without concomitant colexification of incrementation. He derived this gap from a decompositional analysis in which comparison is the semantic primitive out of which the other two functions are built. This paper revisits the generalization with greater typological rigour, using a genealogically and areally stratified sample of 50 languages and a two-pronged methodology: a traditional survey of reference grammars and dictionaries, and an automatic extraction of form–function pairs from a verse-aligned parallel corpus of New Testament translations. We further broaden the empirical base by including additive functions beyond incrementation (repetition, non-scalar and scalar additivity, and disanaphora) and by recording the comparative strategies available in each language. Both studies confirm the colexification gap: no language in the sample colexifies comparison and continuation to the exclusion of incrementation. We also find that standard markers and exceed-predicates never colexify with incrementation, and that incrementation and continuation may be lexicalized in languages with no grammaticalized comparison strategy. These results strongly support the existence of the gap, but only partially support the decompositional analysis proposed to explain it.

Keywords: Additivity, Comparison, Incrementation, Focus Particles

Contents

1	Introduction	3
2	Underspecification in Degree operators	4
3	Language Sampling	7
4	Grammar and Dictionary Survey	8
4.1	Features of interest and search strategy	8
4.1.1	Comparative strategies	8
4.1.2	Colexification	9
4.1.3	Results	10
5	Parallel Corpus Study	13
5.1	Motivation	13

5.2	Methodology	13
5.3	Results	15
5.3.1	Extracting Individual Forms	15
5.3.2	Extracting Colexification Patterns	17
6	Conclusion	19
	References	20
	Appendix 1: Language Sample	23
	Appendix 2: Grammar and Dictionary Survey Results	24
	Appendix 3: Grammar and Dictionary Survey References	27
	Appendix 4: Automatic Extraction Results	33

List of Tables

1	Attested Patterns of Colexification (Thomas 2018)	5
2	Language Sampling	7
3	Elements of Comparative Constructions	8
4	Colexification of Comparison, Incrementation and Continuation	11
5	Colexification of Additive Functions	12
6	Match rates for survey forms and extractions at 0.34 NED threshold.	15
7	Automatic Extraction Results in 4 Languages	16
8	Shared-core forms across colexifications (MCL = 2)	17
9	Shared-core forms (MCL = 2) and related survey forms	18
10	Automatic Extraction Results in Romanian	19
11	Languages included in the sample.	23
12	Grammar and dictionary survey results: comparison-related markers.	24
13	Grammar and dictionary survey results: additive and continuation markers.	25
14	Grammar and dictionary survey results: sources consulted per language.	26
15	Extracted Markers, Part 1 of 3	33
16	Extracted Markers, Part 2 of 3	34
17	Extracted Markers, Part 3 of 3	35

1 Introduction

Words like English *more* and German *noch* have incremental uses, which express addition to a previously asserted quantity:

- (1) John baked 3 cakes. Then he bought one *more*.
- (2) Otto hat *noch* einen Schnaps getrunken.
Otto had NOCH one schnaps drunk
'Otto had another schnaps.'

One can approach incrementation as a form of additivity: additive constructions presuppose that a property holds of some entity and asserts that it holds of one of its alternatives (König 1991, Krifka 1998). The truth- and felicity-conditions of additive constructions can be schematized as in (3):

- (3) ADD($\langle \Phi, \alpha \rangle$)
Presupposition: $\exists[x \neq \alpha \wedge \Phi(x)]$
Assertion: $\Phi(\alpha)$

Incrementation, however, is stronger than the more familiar type of additivity expressed by particles like *also* and *too*. Intuitively, incremental additivity conveys that some quantities are “adding up to a larger whole” (König 1991). What this added requirement amounts to in practice is a matter of debate (see a.o. Greenberg 2010, Thomas 2010, Umbach 2012).

An interesting property of incrementation is that it is commonly colexified with comparison (for instance with English *more*) and continuation (as with German *noch*). Following François (2008), we say that a language colexifies “two functionally distinct senses if, and only if, it can associate them with the same lexical form.” In other words, colexification serves as a cover term for different forms of multiplicity of senses, such as polysemy, ambiguity, vagueness or underspecification, which become impractical to distinguish in cross-linguistic studies involving a large enough number of languages (Haspelmath 2003).

Thomas (2018) investigates colexification of incrementation with comparison and continuation in a sample of 12 languages and observes that although different patterns of colexification of these three functions are attested, no language appears to colexify comparison and continuation without concomitant colexification of incrementation. To explain this gap, Thomas proposes a decompositional analysis of the three functions that derives incrementation and continuation from comparison.

The present manuscript revisits colexification patterns involving incrementation with greater typological rigour. While Thomas’s (2018) study is suggestive, it is based on a small sample that suffers from severe over-representation of Western European languages. We remedy this issue by creating a sample of 50 languages stratified genealogically and by area. We investigate colexification patterns in this sample using a two-pronged approach: first, through a traditional survey of grammars and dictionaries, and second, by automatically extracting colexification patterns from a parallel corpus. We also extend the scope of Thomas’s (2018) study by taking into account typological variation in comparison strategies and by including other additive functions besides incrementation in our study.

Our results support the hypothesis of a colexification gap: comparison and continuation are never colexified without incrementation in our sample. We also find that certain types of comparative morphemes do not participate in colexification of comparison with incrementation and continuation,

and that the latter two functions are attested even in the absence of grammaticalized strategies of comparison. We discuss the significance of these findings for formal analyses of the colexification gap.

The paper is structured as follows. Section 2 reviews Thomas's (2018) analysis of colexification patterns involving incrementation and discusses limitations of his study. Section 3 introduces our language sample. Section 4 presents our grammar and dictionary survey and section 5 presents our parallel corpus study. Section 6 concludes.

2 Underspecification in Degree operators

Colexification of incrementation with continuation is well documented in German (Doherty 1973, König 1977, Eckardt 2006, Umbach 2012), Modern Hebrew (Greenberg 2012), French, Italian and Mandarin Chinese (Tovena & Donazzan 2008). Colexification of incrementation with comparison received less attention and was first described, to the best of our knowledge, by Greenberg (2009) and Thomas (2010) in English and Guarani, respectively. Building on this work, Thomas (2018) investigates patterns of colexification in 12 languages:

- (4) Thomas's (2018) language sample:
- a. Romance: French, Italian, Brazilian Portuguese, Romanian, Spanish
 - b. Germanic: English, German
 - c. Semitic: Modern Hebrew
 - d. Austroasiatic: Vietnamese
 - e. Tupi: Mbyá Guarani
 - f. Slavic: Russian
 - g. Finno-Ugric: Hungarian

Six possible patterns of colexification of comparison (COMP), continuation (CONT) and incrementation (INCR) are considered: COMP-INCR (exclusive colexification of comparison and incrementation), CONT-INCR, COMP-CONT, COMP-CONT-INCR (colexification of all three functions) and $\emptyset$ (no colexification).

Thomas finds that, in positive environments (i.e., when the exponent is not in the scope of negation), all patterns of colexification of the three functions are attested in the sample, with the exception of COMP-CONT (colexification of comparison and continuation without concomitant colexification of additivity). Table 1 summarizes his findings.¹

Note that table 1 paints a simplified picture of colexification patterns in the sample. Firstly, multiple colexification patterns are attested in some languages. To illustrate, French (*de plus*) colexifies comparison and additivity, while *encore* colexifies additivity and continuation. Secondly, some lexical items have a broader pattern of colexification in the scope of negation. This is the case for instance with Portuguese *mais*, which colexifies comparison and additivity in positive environment, but may also be used to express continuation under negation (cf. *Não está mais chovendo*, 'It's not raining anymore.'). Even so, no colexification of comparison and continuation without additivity is found in the sample.

¹ In table 1, identical letter identifiers across functions indicate evidence of colexification of the functions in the language.

Languages	Comparison	Incrementation	Continuation
<i>Romanian</i>	A	A	A
<i>English</i> <i>French</i> <i>M. Guaraní</i> <i>B. Portuguese</i> <i>Spanish</i>	A	A	B
<i>German</i> <i>M. Hebrew</i> <i>Hungarian</i> <i>Italian</i> <i>Russian</i>	A	B	B
<i>Vietnamese</i>	A	B	C
unattested	A	B	A

Table 1: Attested Patterns of Colexification (Thomas 2018)

To explain the colexification gap, [Thomas \(2018\)](#) proposes a decompositional analysis of comparison, additivity and continuation. Expressions like *more* and *noch* spell out combinations of three abstract features labelled COMP, ADD (for additivity) and CONT. The simplest expressible function, comparison, is encapsulated in the feature COMP. Incrementation results from the combination of COMP and ADD, while continuation results from the combination of all three features:

- (5) [Thomas's \(2018\)](#) decompositional analysis:
- Comparison: {COMP}
 - Incrementation: {COMP, ADD}
 - Continuation: {COMP, ADD, CONT}

Colexification is argued to arise by a combination of underspecification and blocking in a realizational theory of morphology ([Halle & Marantz 1993](#)). As shown in example (6), the lexical item *more* in English is only required to spell out the feature {COMP}, which allows it in principle to spell out any super-set of this feature, including incrementation and continuation. The use of *more* to spell out continuation is blocked by the lexical item *still*, which is more specific. By contrast, no competitor blocks speakers from spelling out incrementation as *more*:

- (6) a. *more* $\leftrightarrow$ {COMP}
b. *still* $\leftrightarrow$ {COMP, ADD, CONT}

[Thomas's \(2018\)](#) analysis predicts that underspecification never results in colexification of comparison and continuation without incrementation: if a lexical item is specific enough to block the realization of incrementation by a less specific form, this lexical item is *ipso facto* a preferable realization of continuation.

Thomas' decomposition of the three functions captures their basic truth- and felicity conditions and fits within Schwarzschild's (2012) scale segment theory of degree semantics. More recently, Li (2023) proposes an alternative analysis of the three functions that eschews scale segment semantics and captures uses of incrementation and continuation that Thomas (2018) overlooked. Nevertheless, Li's analysis maintains the overall logic of Thomas' feature-based decomposition, which posits that comparison is the most basic function from which the other two are derived.

While Thomas's (2018) analysis and Li's (2023) alternative proposal offer a detailed formal analysis of colexification patterns involving incrementation, they suffer from a number of limitations that we aim to address in the present manuscript. Firstly, the sample used by Thomas (2018) is not representative of the world's linguistic diversity. In addition to its small size, it is strongly biased towards Indo-European and specifically Western European languages. We address this issue with a larger diversity sample of 50 languages.

Another limitation is that it ignores colexification with a number of potentially relevant functions. This is most obviously the case with additive functions at large. As we argued in the introduction, incrementation is a form of additivity. Therefore, one should ask whether colexification of comparison and continuation with incrementation also extends to other additive and related functions, such as simple non-scalar additivity (e.g., *also*), repetition (*again*) and scalar additivity (*even*). One should also ask whether the gap in colexification patterns that Thomas (2018) identified is still attested once this broader inventory of additive functions is taken into account.

A third weakness of Thomas's (2018) analysis is its unexamined assumption that comparison is the simplest of the three functions under discussion. We might expect that combinations of morphosyntactic features that are simpler are more likely to be lexicalized or grammaticalized than more complex ones. If this is the case, Thomas's (2018) analysis predicts that comparison should be expressible through lexical or grammatical means whenever incrementation or continuation are. We test this hypothesis in our sample.

Finally, although Thomas's (2018) analysis gave a central role to comparison, it did not take into account the significant typological variation in its manifestations. Languages use different morphosyntactic strategies to express comparison (Stassen 1985, Kennedy 2005, Beck et al. 2009: see a.o). Is this variation in any way associated with the colexification of comparison with additivity and continuation? For instance, are *exceed*-type comparatives (which express comparison through the use of predicates that mean 'exceed' or 'surpass') ever colexified with additivity or continuation?

In sum, we identified four limitations in Thomas's (2018) study, which we aim to address in the present manuscript:

- (7) Limitations of Thomas's (2018) cross-linguistic study:
 - a. Small and biased sample
 - b. No attention to additivity beyond incrementation
 - c. Unexamined assumption of unmarkedness of comparison
 - d. No attention to typological variation in comparison strategies

In the next two sections, we introduce our language sample and we discuss our survey of grammars and dictionary.

3 Language Sampling

We constructed a variety sample with a predetermined size of 50 languages, using the Genus-Macroarea sampling method described by [Miestamo et al. \(2016\)](#). Our sample is drawn from the list of 2662 languages included in the World Atlas of Language Structures Online CLDF dataset ([Dryer & Haspelmath 2022](#)), which are classified by genus, family and macroarea. We use genera as our unit of genealogical stratification. The notion of genus was proposed by [Dryer \(1989\)](#) as a level of classification whose time depth is comparable across the world, unlike top families, which may differ significantly in time depth. Genera are intended to be comparable to subfamilies of Indo-European like Germanic and Romance.

Our sampling procedure guarantees that genealogically more diverse macroareas are represented by a greater number of languages in the sample. Specifically, for every macroarea m , we calculate the proportion p_m of genera that m contributes to the total number of genera in the dataset. We then sample $p_m * 50$ languages from m , using different genera and if possible different families. Table 2 shows the number of languages sampled by macroarea. The list of sampled languages is provided in Appendix 1.

Macroarea	Genera	Proportion	# in sample
Africa	125	0.20	10
Australia	40	0.06	3
Eurasia	89	0.14	7
North America	108	0.17	8
Papunesia	167	0.27	13
South America	112	0.18	9
Total	624	1.00	50

Table 2: Language Sampling

Within each genera, we opted for convenience sampling, selecting languages for which good grammatical descriptions and dictionaries were available. Whenever possible, we also selected languages for which we had access to a native-speaker expert.

(8) Sampled languages by Macro-Area:

- a. **Africa:** Kabyle (Berber), Somali (Lowland East Cushitic), Amharic (Semitic), Hausa (West Chadic), Swahili (Bantu), Yoruba (Defoid), Igbo (Igboid), Wolof (Wolof), Kenga (Bongo-Bagirmi), Shilluk (Western Nilotic)
- b. **Australia:** Burarra (Burarran), Kuku-Yalanji (Northern Pama-Nyungan), Djambarrpuyngu (Western Pama-Nyungan)
- c. **Eurasia:** Turkish (Turkic), Vietnamese (Vietic), Standard German (Germanic), Hindi (Indic), Basque (Basque), Hakka Chinese (Chinese), Hungarian (Ugric)
- d. **North-America:** Severn Ojibwe (Algonquian), Bribri (Western Isthmic Chibchan), Greenlandic Inuktitut (Eskimo), Carrier (Athapaskan), Cherokee (Southern Iroquoian), Central (Quiché Mayan), Central Mazahua (Otomian), Western (Huasteca Nahuatl)

- e. **Papunesia:** Tagalog (Greater Central Philippine), Bali (Malayo-Sumbawan), Hawaiian (Oceanic), Kuot (Border), Moskona (East Bird's Head-Sentani), Qaqet (Baining), Barai (Koiarian), Bargam (Bargam), Iatmul (Ndu), Biangai (Kunimaipen), Guhu-Samane (Guhu-Samane), Amele (Mabuso), Kamula (Kamula-Elevala)
- f. **South-America:** Carib (Cariban), Kaingang (Je Meridional), Cofán (Cofán), Ticuna (Ticuna), Caquinte (Souther Mainpueran), Chacobo (Panoan), Cuzco Quechua (Quechuan), Cavineña (Tacanan), Mbyá Guaraní (Maweti-Guarani)

4 Grammar and Dictionary Survey

4.1 Features of interest and search strategy

We extracted information on comparison, additivity and continuation from reference grammars and dictionaries for every language of our sample. This subsection describes the features that were extracted and the methodology we followed.

4.1.1 Comparative strategies

There is significant cross-linguistic variation in the expression of comparison (Stassen 1985, Kennedy 2005, Beck et al. 2009). Following Stassen (1985) and Kennedy (2005), we recognize the following elements of comparative constructions, which do not appear in all comparative strategies:

<i>Kim</i>	<i>is</i>	<i>more</i>	<i>impatient</i>	<i>than</i>	<i>Lee</i>
Target of		Comparative	Comparative	Standard	Standard of
Comparison		Marker	Predicate	Marker	Comparison

Table 3: Elements of Comparative Constructions

When surveying comparative strategies attested in languages of our sample, we record the occurrence of morphemes used as comparative markers and standard markers. In addition, we also record the use of three comparative strategies that do not make use of comparative markers: bare comparatives, *exceed*-comparatives and conjoined comparatives.

Bare comparatives (Schwarzschild 2010) lack a comparative marker but contain all other elements of comparative constructions identified in table 3, including a comparative predicate and a standard marker. Bare comparatives are illustrated in several languages of our sample, including in Shilluk (Nilo-Saharan), which we use to illustrate the construction:

- (9) *Shilluk Bare Comparative* (Nilo-Saharan; Remijsen & Ayoker 2020: ch. 4, §3.3):
 atɪŋɪpɪŋ mɛt kɪ bɔɔl-ɪ kwɔn
 milk.porridge tasty PRP front.side-PRT.SG porridge
 'Milk-based porridge is tastier than porridge made with water.'

Shilluk also makes use of *exceed-comparatives*, in which the standard of comparison is the object of a transitive predicate meaning 'exceed' or 'surpass'. The comparative predicate is realized as a satellite of the exceed-predicate; in Shilluk, it is realized as a nominalized prepositional adjunct of that predicate:

- (10) *Shilluk Exceed-Comparative* (Nilo-Saharan; Remijsen & Ayoker 2020: ch. 4, §3.3):

atɪŋɪpɪŋ lʊot a kwan ki met-ɔ
 milk.porridge surpass:NT FOC porridge PRP tasty-NOM
 ‘Milk-based porridge is tastier than porridge made with water.’

In *conjoined comparatives*, comparison is expressed by conjoining two clauses that contrast the target and standard of comparison along some dimension. This strategy is attested for instance in Bargam (Trans-New Guinea), where the two clauses are related by parataxis:

- (11) *Bargam* (Trans-New Guinea; Hepner 2006: §6.3.3)

Ya sisaʔ-ta. Ni mat siʔim-ta.
 I long-MOD you man short-MOD
 ‘I am tall. You are a short man. (I’m taller than you)’

We classified comparative strategies in languages of our sample along five boolean variables: presence of a comparative marker, presence of a standard marker, presence of bare comparatives, presence of exceed-comparatives and presence of conjoined comparatives. Note that these variables are not mutually exclusive. Indeed, several constructions may compete for the expression of comparison in a language, as illustrated already with Shilluk bare and exceed-comparatives.

We come back to this classification later in this section, where we discuss the findings of our grammar and dictionary survey and the association of comparison strategies and colexification type.

4.1.2 Colexification

Previous studies of colexification focused on comparison, incrementation and continuation. We extend this set of functions to include other forms of additivity beside incrementation: repetition (*again*), simple non-scalar additivity (*also*, *too*) and scalar additivity (*even*). We also included disanaphora (e.g., *other*), a function closely related to additivity.

Non-scalar additives like *also* and *too* are pure additives that express the schema in (3), repeated as (12) for convenience. Scalar additives like *even* have been argued to further induce an ordering of the asserted proposition with respect to its alternatives. The classical view of this ordering is that the asserted proposition is understood to be less likely than its alternatives (Horn 1969, Karttunen & Peters 1979): *Even Muriel voted for Hubert* presupposes that someone other than Muriel voted for Hubert and also conveys that Muriel was the least likely person to vote for Hubert.²

- (12) ADD($\langle \Phi, \alpha \rangle$)
 Presupposition: $\exists[x \neq \alpha \wedge \Phi(x)]$
 Assertion: $\Phi(\alpha)$

Repetitives like *again* are additive expressions that operate in the domain of events and add a temporal ordering to their alternatives: *The temperature will fall again*, asserts that there will be an event of the temperature falling, and presupposes that there is a different event of the same kind that precedes it. Repetition has been observed to colexify with incrementation in several Romance languages, as il-

² Note that this characterization of the scalar component of *even* and related particles has been challenged in subsequent research. See Zhang (2022) for a recent discussion.

lustrated by French *encore*, Italian *ancora* and Romanian *mai* (Tovena & Donazzan 2008, Donazzan & Mardale 2010).

Finally, we also included disanaphoric expressions like *other* in our survey. Manutento (2020) defines them as “deep anaphors indicating a non-identity relationship between the DP that contains them and their antecedent.” While disanaphors are typically ignored in discussions of additive constructions, we include them in our survey because they are conceptually related to additivity and because of attested cases of colexification with comparison and incrementation: in San Sebastián del Monte Mixtec, the enclitic =*ga* colexifies disanaphora and comparison (Manutento 2020), and it can be argued that English *another* and Italian *un altro* colexify disanaphora and incrementation in English and Italian (Thomas 2011, Cinque 2015).

For each function of interest, we looked for relevant topics in the index and table of contents of reference grammars in our sample. We also searched the grammars and dictionaries for occurrences of the following English keywords or their metalanguage translations:

- (13) List of search keywords by function:
- a. Comparison: *more, than, exceed, surpass*
 - b. Incremental additivity: *more, another*
 - c. Repetitive additivity: *again*
 - d. Simple non-scalar additivity: *also, too*
 - e. Scalar additivity: *even*
 - f. Continuation: *still*
 - g. Disanaphora: *other, another*

In dictionaries, we not only looked for occurrences of these keywords as dictionary entries but also performed a reversed search whenever possible, and searched for occurrences of the keywords inside dictionary definitions.

4.1.3 Results

The detailed results of our survey are presented in Appendix 2 in tabulated form. A list of references consulted for each languages is provided in Appendix 3. We start with patterns of colexifications of comparison, incrementation and continuation, presented in table 4. The three colexification patterns identified by Thomas (2018) are attested in the sample, although three way colexification of comparison, additivity and continuation is only attested with one form, Turkish *daha*. It is debatable whether *daha* truly colexifies all three functions, since its status as a comparative marker is debated in the literature on Turkish (see references in Hofstetter 2012: who analyzes *daha* as a norm-related intensifier). Colexification of comparison and continuation without concomitant colexification of incrementation is not attested in the sample.

Instead of restricting our attention to incrementation, we may consider colexification patterns involving additivity as a whole. That is to say, we now lump all flavours of additivity as a single function and record its colexification with comparison and continuation. Doing so only adds two possible cases of colexification: the Wolof suffix *-ati* colexifies repetitive additivity with continuation (Munro & Gaye 1997: p. 8), (Nouguier Voisin 2002: p. 85) and the Amele particle *ca* colexifies comparison with a func-

INCR-CONT		COMP-INCR		COMP-INCR-CONT	
Bali	<i>buin</i>	Cofán	<i>ti'tshe</i>	Turkish (?)	<i>daha</i>
Carib	<i>noro</i>	Hawaiian	<i>(a) keu</i>		
Chacobo	<i>jari</i>	Hindi	<i>zyaadaa,</i> <i>adhik,</i> <i>aur</i>		
German	<i>noch</i>	Mbyá Guaraní	<i>-ve</i>		
Hungarian	<i>még</i>	Western Huasteca Nahuatl	<i>mas</i>		
Kaingang	<i>ver</i>	Tagalog	<i>mas</i>		
Central Mazahua	<i>xe</i>				
Severn	<i>kiyaapic</i>				
Ojibwe					
Tagalog	<i>pa</i>				
Turkish (?)	<i>daha</i>				

Table 4: Colexification of Comparison, Incrementation and Continuation

tion that Roberts (1987) glosses as additivity, although Amele examples suggest that non-comparative uses of *ca* express conjunction rather than presuppositional additivity:

- (14) *Amele* (Mabuso; Roberts 1987: p. 135)
 Jo i ben (qa) jo eu ben ca
 house this big but house that big ADD
 ‘This house is big but that house is bigger.’
- (15) *Amele* (Mabuso; Roberts 1987: p. 32)
 Age na ca ceta ca eeta ca je-gi-na?
 3P taro ADD yam ADD what ADD eat-3P-PRES
 ‘They are eating taro, yam and what?’

Within additive functions, multiple forms of colexification are attested, as shown in in table 5. Colexification between incrementation and repetition is especially common, which might be an artifact of the difficulty in teasing the two functions apart (cf. *it rained again* vs. *it rained once more*).

In sum, these results support Thomas’s (2018) hypothesis of a colexification gap involving comparison and continuation.

If we now focus more closely on variation in comparative strategies, we observe that standard markers (e.g., *than*) never colexify with incrementation in the sample, even in languages with bare comparatives (Amharic, Biangai, Carib, Hindi, Kabyle, Cuzco Quechua, Shilluk, Somali, Turkish). This is consistent with Thomas’s (2018) analysis, according to which the comparative feature that colexifies with incrementation and continuation is properly associated with comparative markers rather than with standard markers. This assumption is rooted in Schwarzschild’s (2012) semantics of comparison and its application to the decompositional analysis of comparison, incrementation and continuation. Schwarzschild argues that comparison asserts the existence of a rising segment of a scale, such as a height scale in the example *Alex is taller than Ryan*. The function of standard and target markers is to

Incremental + Repetitive		Incremental + Simple Additive		Repetitive + Simple Additive		Simple Additive + Scalar Additive		Incremental + Repetitive + Simple Additive	
Amele	<i>haun</i>	Hindi	<i>aur</i>	Cavineña	<i>-nuka</i>	Basque	<i>ere, are</i>	Kabyle	<i>da</i>
Bali	<i>buin</i>	Igbo	<i>kwu</i>	Kamula	<i>ma</i>	Moskona	<i>tin</i>		
Burarra	<i>baja</i>			Qaqet	<i>saqi</i>				
Djambarrpuyngu	<i>daha</i>								
Hakka Chinese	<i>tsai4</i>								
Hausa	<i>ḱara</i>								
Hawaiian	<i>hou</i>								
Kenga	<i>dààlà</i>								
Moskona	<i>tas</i>								
Shilluk	<i>kete</i>								
Vietnamese	<i>nũa</i>								

Table 5: Colexification of Additive Functions

introduce the endpoints of this segment, while comparative markers specify that it is a rising segment (i.e., one that increases along the scale from the standard to the target). It is this element of meaning that Thomas (2018) argues is shared with incrementation and continuation and enables colexification.

Colexification of incrementation with *exceed*-comparison is not attested either in the sample. That is to say, *exceed*-predicates do not colexify with incrementation. Note however that some languages in our sample express both comparison and incrementation with verbs, albeit different ones. In Hausa (Afroasiatic; Jaggar 2001, Newman 2000; 2007), comparison is expressed with the verb *fi* ('exceed, surpass') and incrementation is expressed with the verb *ḱara* ('increase, add to, repeat'). In Wolof (Niger-Congo; Munro & Gaye 1997), comparison is expressed with the verb *gën* ('be more than'), while incrementation is expressed with the verbs *dolli* ('give more to, add more'), *yokk* ('add more') or *feccali* ('to fill up: especially, to add more to fill, to have some more food'):

- (16) Hausa (West Chadic; Jaggar 2001: p. 474)
 Audù yā fi Mūsā wāyō
 Audu 3M.PF exceed Musa cleverness
 'Audu is cleverer than Musa.'
- (17) Hausa (West Chadic; Jaggar 2001: p. 197)
 bà zā à kārà cín wani àbù ba sai...
 'then they will not eat anything else until...'
- (18) Hausa (West Chadic; Newman 2000: p. 64)
 Wàzīrī Aku zāi kārà bā shì misālai
 'Vizier Parrot will give him more examples'
 (lit. ... will increase give him examples)

It is somewhat surprising that incrementation and comparison are never colexified as verbs in the sample, since both types of verbs express increase along a dimension, and *exceed*-comparatives are rather frequent, being attested in 17 of our 50 languages.

Finally, we observe that incrementation and continuation are attested in languages that lack lexicalized or grammaticalized comparison strategies. This is the case in Qaquet, a Baining language of East New Britain. Hellwig (2019: p. 145) observes that “Qaquet adjectives do not have comparative or superlative forms, and there is no dedicated comparative construction. Instead, comparative or superlative interpretations are left to inference.” Incrementation and continuation are nevertheless lexicalized in the language, as illustrated by the following examples:

- (19) Qaquet (Baining; Hellwig 2019: p. 389)
 as amaɽgil-em
 still ART3small-SG.RCD
 ‘(she) is still a small (child).’
- (20) Qaquet (Baining; Hellwig 2019: p. 393)
 miika nguɽnarlip nguɽsrlup
 more 1SG.SUBJ.NPST3like:PURP 1SG.SBJ.NPST3drink.CONT
 ‘I want to drink more.’

This observation is somewhat problematic for Thomas’s (2018) compositional analysis of incrementation and continuation. Thomas proposes that these two functions are derived by adding features to comparison, which is taken to be semantically and morpho-syntactically simpler than incrementation and continuation. Everything else being equal, we may infer from this proposal that the grammaticalization or lexicalization of the more complex functions in a language should entail that of the simpler ones. Qaquet may of course be an isolated example, but it raises the question whether comparison is indeed the simpler notion.

5 Parallel Corpus Study

5.1 Motivation

This study complements our grammar and dictionary analysis in two respects. First, it is a completeness check that may capture strategies that were missed in the manual survey. Secondly, and more importantly, it opens a window on colexification in language use: a survey of reference grammars and dictionaries may record rare cases of colexification that are hardly attested in discourse. While resorting to a parallel corpus limits our choice of texts to a non-representative convenience sample (verses from the New Testament), we believe that the insights to be drawn from this study outweighs the limitations imposed by the corpus.

5.2 Methodology

Our corpus consists of translations of the New Testament aligned by verses. The general strategy we follow is to extract morphemes that express a function of interest from a sample of New Testament verses in English, and to automatically extract morphemes or words expressing the same function in matching verses from our 50-language sample.

We extracted 1843 English Bible verses containing an occurrence of the morphemes *another*, *more*, *-er* and *still*. Three annotators then categorized each pair of verse and target morpheme as express-

ing Comparison, Continuation, Incrementation, as well as other salient functions.³ Annotators were also asked to indicate whether the target morpheme in each verse occurred in the scope of negation. Cohen-Kappa inter-annotator agreement was substantial for all pairs (0.66, 0.68, 0.76). In a second step, we resolved disagreements between annotators by defaulting to the majority annotation. In cases where genuine ambiguity was suspected, we preserved it by creating combined categories (e.g., Comparison/Incrementation). For the purposes of the present study, we then filtered the data set to New Testament verses whose target morpheme expressed Comparison, Incrementation, or Continuation. We also excluded verses where target morphemes occurred in the scope of negation, and verses whose function was ambiguous. This resulted in a set of 712 seed verses (Comparison: 325, Incrementation: 170, Continuation: 217).

Next, we automatically extracted morphemes that are likely to express Comparison, Incrementation, or Continuation in our 50-language sample, building on Liu et al.'s (2023) concept-alignment algorithm. For this purpose, we represent a function F (e.g., Comparison) as the subset of verses annotated with it in the set of English seed verses. We then extract the most likely markers of F in other languages by looking for word strings or substrings that are significantly associated with F . More precisely, for each target language, we retrieved all of its verses as B . From B , we extract the set T of word strings and word substrings that occur in B . Then, for each function F under consideration, we search for the string t in T that has the most significant association with F . Remember that F is represented as a set of verses. Therefore, testing the association between a term t and a function F in any non-English set of verses B in the parallel corpus amounts to testing the association between two indicator variables over verses of B : whether a verse $v \in B$ contains t and whether a verse $v \in B$ is included in F .

For each function F , we select the term t with the lowest p -value on a one-tailed Fisher's exact test. We then remove the verses containing t and repeat this selection on the remainder of the verses until the set of verses remaining falls below some proportion P of the verses in F . P is a parameter of the extraction procedure that we selected on a tuning set: 30 verses out of 712 were held out and manually annotated in German, Mbyá Guaraní, Romanian and Vietnamese.⁴ P was set to 50% using a trade-off between adjusted Rand scores and number of markers correctly extracted.⁵

Having extracted for each language a set of likely markers of comparison, incrementation and continuation, we can then identify colexification patterns in the sample: if a string is extracted for two or more functions, it is said to colexify those functions. Note that since extraction is done by function and targets substrings as well as whole word strings, looking for colexification by strict string identity is likely to yield many false negatives. To illustrate, the procedure correctly extracts *-ve* for the comparative function in Mbyá Guaraní, but it also extracts *-tave* for the additive function (an undersegmentation, lumping *-ve* with the syllable *ta*, which happens to frequently, but accidentally, co-occur with *-ve* in additive verses in our sample. Hence, if for two extracted strings the shorter string is a substring of the other string and the substring meets or exceeds a given (parametrized) minimum length (the "Minimum Core Length", or MCL), we treat the two strings as identical. Looking for colexification by

³ Distinction, Reciprocity, Support, Concession, and Concession. Annotators also had the option to annotate a pair of verse and target morpheme as Unsure or Other.

⁴ Note that Romanian is not part of our sample but was nonetheless used for tuning.

⁵ Adjusted Rand scores measure the similarity between the verse clusterings obtained from the extraction procedure and from the manual annotation, regardless of the number of markers correctly extracted by the algorithm. We select a proportion P that optimizes both the Rand score and the number of markers correctly extracted.

substring containment correctly identifies *-ve* as a comparative *and* an additive marker when MCL=2. To illustrate, setting MCL to 3 would rule out identifying comparative *-ve* and additive *-tave* as a case of colexification, since the core substring *-ve* has a length less than 3. Rather than finetuning MCL, we consider colexification patterns obtained from values of MCL ranging from 1 to 5.

5.3 Results

5.3.1 Extracting Individual Forms

In total, 497 form/function pairs were extracted across the 50 languages. The full list is provided in Appendix 4. Assessing which of them match forms identified in our grammar and dictionary survey is challenging for several reasons. Firstly, there is spelling variation across the relevant sources. Secondly, the extraction procedure is not aware of morphological boundaries and is prone to extract parts of morphemes, or to lump morphemes with adjacent segments, as discussed in the previous section. To address these issues, we compare the morphemes identified in our survey with the results of our automatic extraction using length-normalized minimal edit distance (NED). At a threshold of 0.34, 48% of survey forms have an extracted match, and 37% of automatically extracted forms have a match in the survey:

	Survey forms with a match	Extractions with a match
COMP	48/119 (0.40)	46/158 (0.29)
CONT	41/65 (0.63)	44/170 (0.26)
INCR	177/368 (0.48)	94/169 (0.56)
All	266/552 (0.48)	184/497 (0.37)

Table 6: Match rates for survey forms and extractions at 0.34 NED threshold.

Let us zoom in on a few selected languages to illustrate important features and limitations of the extraction procedure, before we consider hypothetical cases of colexification. Table 7 shows all extracted targets for Standard German, Mbyá Guaraní, Hungarian and Vietnamese, together with the frequency of each form and the proportion of seed verses it covers. The special character \$ indicates word boundaries. To illustrate, the German form \$noch\$ was extracted 86 times as a marker of continuation (CONT), and it covers 43% of seed verses associated with this function in Standard German. As an incremental marker (INCR), it was extracted 45 times and covers 30% of seed verses.

We note that the top forms by frequency for each function are mostly true positives. Some extracted forms have functions that are closely associated with their seed rather than identical to it, such as German *als* \$, which is a standard marker rather than a comparative marker, and Vietnamese *\$lại* \$, which expresses repetition rather than incrementation. Some of the forms are plainly false positives, such as Hungarian *míg* \$ (*while, until*), whose extraction might be due to the fact that many continuation markers occur in temporal adverbial clauses (see example (22) below).

An inherent limitation of the extraction procedure is the fact that not all verses tagged with a certain function in a seed English Bible will express it in other versions of the Bible. Consider for instance the following English and Hungarian versions of Acts 9:39. While the English seed verse contains a continuation marker, the Hungarian verse does not express continuation. If the seed function is

Language	Seed	Form	Frequency	Proportion	Attested in Survey	Function(s)
deu	CONT	\$noch\$	86	0.43	Yes	CONT, INCR
deu	CONT	\$ges	36	0.18	No	Extraction Noise
deu	COMP	als\$	114	0.41	Yes	Standard Marker
deu	COMP	\$eu	51	0.18	No	Extraction Noise
deu	INCR	\$noch\$	45	0.3	Yes	CONT, INCR
deu	INCR	\$wieder\$	20	0.13	Yes	Repetitive
deu	INCR	\$ander	19	0.12	Yes	Disanaphora
gun	CONT	\$teri	79	0.39	Yes	CONT
gun	CONT	\$aỹ	21	0.1	No	<i>now</i>
gun	CONT	mbyt	11	0.05	No	Extraction Noise
gun	COMP	ve\$	205	0.73	Yes	COMP, INCR
gun	INCR	\$ju\$	70	0.46	Yes	Repetitive
gun	INCR	tave	13	0.09	Yes (ve)	COMP, INCR
hun	CONT	an\$	66	0.33	No	Extraction Noise
hun	CONT	\$még\$	36	0.18	Yes	CONT, INCR
hun	CONT	míg\$	17	0.08	No	<i>while, until</i>
hun	CONT	\$mai\$	7	0.03	No	<i>today</i>
hun	COMP	bb	145	0.52	Yes	COMP
hun	INCR	\$ismét\$	33	0.22	Yes	Repetition
hun	INCR	\$más	27	0.18	Yes	Disanaphora
hun	INCR	\$még\$	15	0.1	Yes	CONT, INCR
hun	INCR	\$újra\$	5	0.03	Yes	Repetition
vie	CONT	\$còn\$	82	0.41	Yes	CONT
vie	CONT	\$đang\$	28	0.14	No	Progressive/Durative
vie	COMP	\$hơn	123	0.44	Yes	COMP (<i>exceed</i>)
vie	COMP	\$chị\$	40	0.14	No	Kinship Term
vie	INCR	\$lại\$	44	0.29	Yes	Repetitive
vie	INCR	\$nữa\$	31	0.21	Yes	INCR, Repetitive
vie	INCR	\$khác	17	0.11	Yes	Disanaphora

Table 7: Automatic Extraction Results in 4 Languages

then significantly associated with an unrelated construction in the target Bible, like temporal clauses introduced by *míg*, the procedure may extract false positives based on spurious associations:

- (21) Acts 9:39, *New Simplified Bible* (2009)
 (...) the coats and clothes that dorcás had made while she was still alive .
- (22) Acts 9:39, *Ujszövetseg: elet, igazsag es vilagossag* (2003)
 a ruhá-k-at és öltözet-ek-et, amely-ek-et Dorkász csinál-t, míg vel-ük
 DEF garment-PL-ACC and clothing-PL-ACC which-PL-ACC Dorkás make-PST while 3PL-INS
 együtt vol-t
 together be-PST
 ‘(...) the garments and clothes which Dorkás made while she was with them.’

Another limitation is that seed functions may be expressed as parts of collocations that exhibit more morphological variation in the seed marker than in its collocates. This is the case in German, where the comparative marker is realized either as a bound suffix *-er* or a free morpheme *mehr*, while the standard marker is uniformly realized as *als*. This will strengthen the association of comparison with *als* relative to either allomorph of the comparative marker, thereby favouring the extraction of *als* over *-er* and *mehr*.

5.3.2 Extracting Colexification Patterns

Having extracted candidate forms in all 50 languages, we look for hypothetical cases of colexification using a string containment strategy. For each pair of extracted forms expressing different functions in a language, we ask whether one of them is a subset of the other. If so, and if the shortest form reaches a Minimum Core Length, we hypothesize that these forms may colexify the two functions. We extract hypothetical colexification patterns for MCL values ranging from 1 to 5. Table 8 shows the result at MCL = 2:

Language	Colexification	Shared Core	Forms
amh	INCR+CONT	ገኛ	ገኛ [INCR N=16]; ገኛ [CONT N=37]
byx	INCR+CONT	as	askerlka\$ [INCR N=46]; \$as\$ [CONT N=47]; atiaskerl [CONT N=38]
bzd	INCR+COMP	kí	kí\$ [INCR N=19]; kí [COMP N=36]
cao	INCR+CONT	ari	ari\$ [INCR N=31]; \$jari\$ [CONT N=50]
car	INCR+CONT	nororo	\$nororo\$ [INCR N=22]; \$nororo\$ [CONT N=69]
chr	INCR+CONT	asi	\$asi\$ [INCR N=11]; \$asi\$ [CONT N=68]
deu	INCR+CONT	noch	\$noch\$ [INCR N=45]; \$noch\$ [CONT N=86]
gun	INCR+COMP	ve	tave [INCR N=13]; ve\$ [COMP N=205]
hak	INCR+CONT	hàn	\$hàn-yû\$ [INCR N=8]; hàn\$ [CONT N=54]; \$hàn- [CONT N=37]
hau	INCR+COMP	ara	\$ara\$ [INCR N=19]; \$ara\$ [COMP N=16]
haw	INCR+COMP	nu	nui [INCR N=26]; \$nu\$ [COMP N=36]
hun	INCR+CONT	még	\$még\$ [INCR N=15]; \$még\$ [CONT N=36]
kgp	INCR+CONT	ver	\$ver\$ [INCR N=14]; \$ver\$ [CONT N=67]
maz	INCR+CONT+COMP	xe	\$xe\$ [INCR N=47]; \$xe\$ [CONT N=55]; \$xenda\$ [COMP N=107]
mlp	INCR+CONT	ta	\$ta\$ [INCR N=80]; \$tuta\$ [CONT N=19]
nhw	INCR+COMP	más	más\$ [INCR N=27]; \$más\$ [COMP N=138]
ojis	INCR+CONT	q̣	q̣ [INCR N=26]; \$q̣^ [CONT N=45]
quz	INCR+COMP	tawan	tawan [INCR N=29]; \$astawan [COMP N=20]
som	INCR+COMP	badan	\$badan\$ [INCR N=17]; \$badan\$ [COMP N=42]
swh	INCR+COMP	id	id [INCR N=29]; \$zaidi\$ [COMP N=48]
tgl	INCR+CONT	pa	\$pang\$ [INCR N=11]; \$pa\$ [CONT N=35]
tur	INCR+COMP	aha	\$daha\$ [INCR N=28]; aha\$ [COMP N=99]

Table 8: Shared-core forms across colexifications (MCL = 2)

We do not observe any hypothetical colexification of Comparison and Continuation to the exclusion of Incrementation, even at MCL = 1. This supports the results of our grammar and dictionary survey and of Thomas's (2018) comparative study. Zooming in on our results at MCL = 2, table 9 compares shared cores of extracted forms to related forms from our grammar and dictionary survey. 17 out of 22 shared cores match forms attested in our survey, and 11 of the 22 hypothetical colexifications are attested in our survey. Conversely, out of the 19 cases of colexification involving comparison, incrementation and continuation that we identified in table 4, 10 are attested in our parallel corpus study: *nororo* (Carib, INCR-CONT), *jari* (Chacobo, INCR-CONT), *noch* (German, INCR-CONT), *ve* (Mbyá Guaraní, COMP-INCR), *még* (Hungarian, INCR-CONT), *ver* (Kaingang, INCR-CONT), *xe* (Cen-

tral Mazahua, INCR-CONT), *kiyaapic* (Severn Ojibwe, INCR-CONT), *pa* (Tagalog, INCR-CONT), *daha* (Turkish, COMP-INCR-CONT). For Turkish *daha*, our corpus study only supports the colexification of comparison with incrementation.

Note that the lack of corpus support for some of the colexifications identified in our grammar and dictionary survey is not necessarily a failure of the extraction method, but might reflect a fact of language use: it is possible that many of the colexifications attested in grammars and dictionaries are under-represented in actual language use.

Language	Colexification	Shared Core	Related Survey Form
amh	INCR+CONT	ገና (gäna)	ändägänä (repetitive), ahunəm gänä (CONT)
byx	INCR+CONT	as	?
bzd	INCR+COMP	kí	kĩ (INCR)
cao	INCR+CONT	ari	(j)ari (CONT, non-scalar additive)
car	INCR+CONT	norō	norō (INCR, CONT)
chr	INCR+CONT	asi	?
deu	INCR+CONT	noch	noch (INCR, CONT)
gun	INCR+COMP	ve	ve (COMP, INCR)
hak	INCR+CONT	hàn	han2 (CONT)
hau	INCR+COMP	ara	āra (INCR, repetitive)
haw	INCR+COMP	nu	?
hun	INCR+CONT	még	még (INCR), még mindig (CONT)
kgp	INCR+CONT	ver	ver (INCR, CONT)
maz	INCR+CONT+COMP	xe	xe (INCR, CONT)
mlp	INCR+CONT	ta	ta (disanaphoric, repetitive)
nhw	INCR+COMP	más	mas (COMP, INCR)
ojs	INCR+CONT	ᑭᑦᑦᑦ (keyaa)	kiyaapic (INCR, CONT)
quz	INCR+COMP	tawan	astawan (COMP), huqtawan (INCR)
som	INCR+COMP	badan	?
swh	INCR+COMP	id	?
tgl	INCR+CONT	pa	pa (INCR, CONT)
tur	INCR+COMP	aha	daha (COMP, INCR, CONT)

Table 9: Shared-core forms (MCL = 2) and related survey forms

A good example of the factors that may block colexification in language is provided by Romanian. Although not included in our sample, we nevertheless used the language for tuning and can run our extraction pipeline on Romanian new testaments to extract markers of Comparison, Incrementation and Continuation. The results, summarized in table 10, are striking:

Thomas (2018) identifies *mai* as a three-way marker of Comparison, Incrementation and Continuation. Only the colexification of Comparison and Incrementation is identified in our sample. One factor that may explain this is the fact that Continuation is largely expressed by *încă* in the corpus. In other words, the presence of a more specific marker may preempt the expression of a function with a less specific marker in language use, although the colexification surfaces in targeted elicitation and grammatical descriptions. Examination of individual verses reveals that *mai* is actually attested as a Continuation marker, but its frequency in this function is too low to support extraction:

Language	Seed	Form	Frequency	Proportion
ron	COMP	\$mai\$	155	0.56
ron	CONT	\$cî\$	61	0.3
ron	CONT	\$încă\$	35	0.17
ron	CONT	cum\$	25	0.12
ron	INCR	\$mai\$	38	0.25
ron	INCR	\$iarăş\$	18	0.12
ron	INCR	\$alt\$	13	0.09
ron	INCR	\$nou\$	11	0.07

Table 10: Automatic Extraction Results in Romanian

- (23) 2 Peter 1:13, *Biblia Cornilescu* (1921)
dar socotesc că este drept cât voi mai fi în cortul acesta să
but think-1SG that be-3SG right as.long.as will-1SG still be-INF in tent-DEF.M this SBJV
vă țin treji aducîndu-vă aminte
2PL.ACC keep-1SG awake.PL bring-GER-2PL.DAT to.mind
‘(Lit.) But I think it right, as long as I am still in this tent, to keep you awake, reminding you.’

6 Conclusion

In this study, we set out to address some of the limitations of [Thomas’s \(2018\)](#) investigation of patterns of colexification involving comparison, incrementation and continuation, which we repeat here from (7)

- (24) Limitations of Thomas’s (2018) cross-linguistic study:
- Small and biased sample
 - No attention to additivity beyond incrementation
 - Unexamined assumption of unmarkedness of comparison
 - No attention to typological variation in comparison strategies

We addressed the first of these limitations with a sample of 50 languages that represents macro-area in proportion to their language diversity by genera. We then combined a traditional grammar and dictionary survey of languages in the sample with a computational analysis of a parallel corpus of New Testament translations.

Our main finding is that Comparison and Continuation are never colexified without Incrementation, in keeping with [Thomas’s \(2018\)](#) conclusion. However, we also observed that the presence of markers Incrementation or Continuation in a language does not entail the presence of markers of comparison, or of other grammatical means of expression comparison. This calls into question [Thomas’s \(2018\)](#) argument that Comparison is semantically unmarked relative to Incrementation and Continuation, under the assumption that the lexicalization or grammaticalization of marked functions should entail that of less marked functions. More precisely, we might predict that if Comparison is a semantic building block of Incrementation and Continuation, the lexicalization or gram-

maticalization of these latter two functions should entail that of Comparison. Our study falsifies this prediction.

We also broadened Thomas's (2018) survey to consider other forms of Additivity and morpho-syntactic variation in the expression of Comparison. We observed that Continuation and Comparison do colexify with non-incremental forms of Additivity. We also observed that *exceed* predicates and standard markers do not colexify with either Additivity or Continuation. The absence of colexification with standard markers is consistent with Thomas's (2018) decompositional analysis of Comparison and related functions, which posits that the common semantic core of these functions is the assertion that some scale segment is rising from its source to its goal. If standard markers encode the source of the segment rather than its direction, Thomas's (2018) analysis predicts that they lack semantic motivation for colexification with Incrementation and Continuation. The absence of colexification of *exceed* comparative predicates remains a puzzle, especially given the find that both Comparison and Incrementation may be realized as predicates, as is the case in Hausa.

In sum, our study provides strong confirmation the colexification gap identified by Thomas (2018), but only partially supports his decompositional analysis of this gap, which posits that Comparison is the semantic primitive from which Incrementation and Continuation are built.

References

- Beck, Sigrid & Krasikova, Sveta & Fleischer, Daniel & Gergel, Remus & Hofstetter, Stefan & Savelsberg, Christiane & Vanderelst, John & Villalta, Elisabeth. 2009. Crosslinguistic variation in comparison constructions. *Linguistic Variation Yearbook* 9.
- Cinque, Guglielmo. 2015. A note on 'other'. In Brandner, Ellen & Cypionka, Anna & Freitag, Constantin & Trotzke, Andreas (eds.), *Webschrift for Josef Bayer*, 22–27. Konstanz: Universität Konstanz. <https://ling.sprachwiss.uni-konstanz.de/pages/WebschriftBayer/2015/>. Last accessed on 8/10/2025.
- Doherty, Monika. 1973. Noch and schon and their presuppositions. In Kiefer, Ferenc & Ruwet, Nicolas (eds.), *Generative Grammar in Europe*, 154–177. Dordrecht: Reidel.
- Donazzan, Marta & Mardale, Alexandru. 2010. Additive and aspectual adverbs: towards an analysis of romanian 'mai'. *Revue roumaine de linguistique* 10. 247–269. <https://shs.hal.science/halshs-00663487v1>. Last accessed on 8/10/2025.
- Dryer, Matthew & Haspelmath, Martin. 2022. The world atlas of language structures online. <https://doi.org/10.5281/zenodo.7385533>. Last accessed on 8/10/2025. <https://doi.org/10.5281/zenodo.7385533>
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13(2). 257–292.
- Eckardt, Regine. 2006. Was noch? Navigating in Question Answer Discourse. In Späth, Andreas (ed.), *Interface and interface conditions*, 77–96. Berlin and Boston: Mouton de Gruyter.
- 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), 163–215. Amsterdam and Philadelphia: John Benjamins.
- Greenberg, Yael. 2009. Additivity in the domain of eventuality. In *LOLA* 10. 35–44.

- Greenberg, Yael. 2010. Additivity in the domain of eventualities (or: Oliver twist's more). In *Sinn und Bedeutung* 14, 151–167.
- Greenberg, Yael. 2012. Event based additivity in english and modern hebrew. In Laca, Brenda & Hofherr, Patricia Cabredo (eds.), *Verbal plurality and distributivity*, 127–158. Berlin and Boston: De Gruyter.
- Halle, Morris & Marantz, Alec. 1993. Distributed morphology and the pieces of inflection. In Hale, Ken & Keyser, Samuel Jay (eds.), *The view from building 20*, 111–176. Cambridge, MA: The MIT Press.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Tomasello, Michael (ed.), *The new psychology of language*, vol. 2, 211–243. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hellwig, Birgit. 2019. *A grammar of Qaqet*. Berlin and Boston: De Gruyter Mouton.
- Hepner, Mark. 2006. *Bargam grammar sketch*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Hofstetter, Stefan. 2012. Turkish comparatives: the status and semantic contribution of the adverb *daha* and beyond. In *15th international conference on turkish linguistics (ICTL)* 15, 231–242.
- Horn, Laurence R. 1969. A presuppositional analysis of 'only' and 'even'. In *Fifth regional meeting of the Chicago Linguistic Society (CLS)* 5, 98–107. Chicago, IL: Chicago Linguistic Society.
- Jaggar, Philip J. 2001. *Hausa*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Karttunen, Lauri & Peters, Stanley. 1979. Conventional implicature. In Oh, Choon-Kyu & Dinneen, David A. (eds.), *Syntax and semantics 11: Presupposition*, vol. 11, 1–56. New York: Academic Press.
- Kennedy, Christopher. 2005. Variation in the expression of comparison. Talk given at Cornell University, Ithaca, NY, 11 November 2005.
- Krifka, Manfred. 1998. Additive particles under stress. In *8th Semantics and Linguistic Theory (SALT)* 8, 111–128.
- König, Ekkehard. 1977. Temporal and non-temporal uses of *noch* and *schon*. *Linguistics and Philosophy* 1(2), 173–198.
- König, Ekkehard. 1991. *The meaning of focus particles*. London and New York: Routledge.
- Li, Ang. 2023. Alternative comparison in underspecified degree operators. In *33rd Semantics and Linguistic Theory (SALT)* 33, 480–499.
- Liu, Yihong & Ye, Haotian & Weissweiler, Leonie & Wicke, Philipp & Pei, Renhao & Zangenfeind, Robert & Schütze, Hinrich. 2023. A crosslingual investigation of conceptualization in 1335 languages. Last accessed on 8/10/2025. <https://arxiv.org/abs/2305.08475>.
- Manutento, Iara. 2020. *'other' and 'more' in san sebastián del monte mixtec and beyond*. Los Angeles, CA: University of California, Los Angeles dissertation.
- Miestamo, Matti & Bakker, Dik & Arppe, Antti. 2016. Sampling for variety. *Linguistic Typology* 20(2), 233–296. <https://doi.org/doi:10.1515/lingty-2016-0006>
- Munro, Pamela & Gaye, Dieynaba. 1997. *Ay Baati Wolof: a Wolof dictionary*, vol. 19 (UCLA Occasional Papers in Linguistics). Los Angeles, CA: Department of Linguistics, University of California, Los Angeles.
- Newman, Paul. 2000. *The hausa language: an encyclopedic reference grammar*. New Haven, CT: Yale University Press.
- Newman, Paul. 2007. *A hausa-english dictionary*. New Haven, CT: Yale University Press.

- Nouguier Voisin, Sylvie. 2002. *Relations entre fonctions syntaxiques et fonctions sémantiques en wolof*. Lyon, France: Université Lumière Lyon 2 dissertation.
- Remijsen, Bert & Ayoker, Otto Gwado. 2020. *A grammar of shilluk* (Special Publication 14). Language Documentation & Conservation. <https://hdl.handle.net/10125/56637>. Last accessed on 8/10/2025.
- Roberts, John R. 1987. *Amele*. London: Croom Helm.
- Schwarzschild, Roger. 2010. Comparative markers and standard markers. In *MIT workshop on comparatives*, vol. 69. 87–105. MIT Working Papers in Linguistics.
- Schwarzschild, Roger. 2012. Directed scale segments. In *22nd Semantics and Linguistic Theory (SALT)* 22. 65–82.
- Stassen, Leon. 1985. *Comparison and universal grammar*. Oxford: Basil Blackwell.
- Thomas, Guillaume. 2010. Incremental ‘more’. In *20th Semantics and Linguistic Theory (SALT)* 20. 233–250.
- Thomas, Guillaume. 2011. Another additive particle. In *21st Semantics and Linguistic Theory (SALT)* 21. 634–651.
- Thomas, Guillaume. 2018. Underspecification in degree operators. *Journal of Semantics* 35(1). 43–93.
- Tovena, Lucia & Donazzan, Marta. 2008. On ways of repeating. *Recherches Linguistiques de Vincennes* 37. 85–112.
- Umbach, Carla. 2012. Strategies of additivity: German additive noch compared to auch. *Lingua* 122(15). 1843–1863.
- Zhang, Linmin. 2022. The presupposition of ‘even’. In *32nd Semantics and Linguistic Theory (SALT)* 32, 249–269. Washington, DC: Linguistic Society of America.

Appendix 1: Language Sample

ISO	Language	Macroarea	Family	Genus
kab	Kabyle	Africa	Afro-Asiatic	Berber
som	Somali	Africa	Afro-Asiatic	Lowland East Cushitic
amh	Amharic	Africa	Afro-Asiatic	Semitic
hau	Hausa	Africa	Afro-Asiatic	West Chadic
swl	Swahili	Africa	Niger-Congo	Bantu
yor	Yoruba	Africa	Niger-Congo	Defoid
ibo	Igbo	Africa	Niger-Congo	Igboid
wol	Wolof	Africa	Niger-Congo	Wolof
kyq	Kenga	Africa	Nilo-Saharan	Bongo-Bagirmi
shk	Shilluk	Africa	Nilo-Saharan	Western Nilotic
bvr	Burarra	Australia	Mangrida	Burarran
gvn	Kuku-Yalanji	Australia	Pama-Nyungan	Northern Pama-Nyungan
djr	Djambarrpuyngu	Australia	Australian	Western Pama-Nyungan
tur	Turkish	Eurasia	Altaic	Turkic
vie	Vietnamese	Eurasia	Austro-Asiatic	Vietic
deu	Standard German	Eurasia	Indo-European	Germanic
hin	Hindi	Eurasia	Indo-European	Indic
eus	Basque	Eurasia	Language Isolate	Basque
hak	Hakka Chinese	Eurasia	Sino-Tibetan	Chinese
hun	Hungarian	Eurasia	Uralic	Ugric
ojs	Severn Ojibwe	North America	Algic	Algonquian
bzd	Bribri	North America	Chibchan	Western Isthmic Chibchan
kal	Greenlandic Inuktitut	North America	Eskimo-Aleut	Eskimo
crx	Carrier	North America	Eyak-Athabaskan	Athapaskan
chr	Cherokee	North America	Iroquoian	Southern Iroquoian
quc	Central Quiché (K'iche')	North America	Mayan	Mayan
maz	Central Mazahua	North America	Otomanguan	Otomian
nwh	Western Huasteca Nahuatl	North America	Uto-Aztecan	Aztecan
tgl	Tagalog	Papunesia	Austronesian	Greater Central Philippine
ban	Bali	Papunesia	Austronesian	Malayo-Sumbawan
haw	Hawaiian	Papunesia	Austronesian	Oceanic
kto	Kuot	Papunesia	Language Isolate	Border
mtj	Moskona	Papunesia	East Bird's Head-Sentani	East Bird's Head-Sentani
byx	Qaqet	Papunesia	East New Britain	Baining
bbb	Barai	Papunesia	Koiarian	Koiarian
mlp	Bargam	Papunesia	Nuclear Trans New Guinea	Bargam
ian	Iatmul	Papunesia	Sepik	Ndu
big	Biangai	Papunesia	Kunimaipen	Kunimaipen
ghs	Guhu-Samane	Papunesia	Nuclear Trans New Guinea	Guhu-Samane
aeu	Amele	Papunesia	Nuclear Trans New Guinea	Mabuso
xla	Kamula	Papunesia	Kamula-Elevala	Kamula-Elevala
car	Carib	South America	Cariban	Cariban
kgp	Kaingang	South America	Jean	Je Meridional
con	Cofán	South America	Language Isolate	Cofán
tca	Ticuna	South America	Language Isolate	Ticuna
cot	Caquinte	South America	Maipurean	Southern Maipurean
cao	Chacobo	South America	Panoan	Panoan
quz	Cuzco Quechua	South America	Quechuan	Quechuan
cav	Cavineña	South America	Tacanan	Tacanan
gun	Mbyá Guaraní	South America	Tupian	Maweti-Guarani

Table 11: Languages included in the sample.

Appendix 2: Grammar and Dictionary Survey Results

ISO	Language	Genus	Comparative	Standard	Exceed-predicate	Disanaphoric	Incremental
kab	Kabyle	Berber		f, deg	yif/tif	-iden/-iqin/eniden/ enidrin/miden, wayeḍ	da /da en /da enni /da enna- kin /da nakin /di /di en
som	Somali	Lowland East Cushitic	yasaq, yabāit	ká	laqā, bābāllātā	kale	tāčāmmari, čāmmārā
anh	Amharic	Semitic		kā-/tā-	fi, āra (āra)	lela	čara (āra), dāda (da ā)
hau	Hausa	West Chadic			kupka, kuzidi		(tun) si
swi	Swahili	Bantu	zaidi	ya, kuliko	yu	-ingine/-engine	ōzō kwā, -kwa, -kwu
yor	Yoruba	Defoid		lọ	-ka, -lahy, -nahi, karia	omiran (omiran), omi (omi)	feccali, doli, yokk
ibo	Igbo	Wolof			gēn, ēpp, daq, tane	yeneen, beneen, keneen, feneen, leneen, neneen	dālā
wol	Wolof				ciri	m ko	kete, nyenyo, kal
kyq	Kenga	Bongo-Bagirmi		k' b' l-			
shk	Shilluk	Western Nilotic			r' t		
bvr	Burarra	Burarra		-ka		yindu	baja
gyn	Kuku-Yalanji	Northern Pama-Nyungan	jarra-			wiripu	bulu-bulu
djr	Djambarrpungu	Western Pama-Nyungan	mirithirr, miririthi-			baška, diđer	daha
tur	Turkish	Turkic	daha	-DAn/-den	hon, quā	trüdc, cāi khác, người khác	thēm, nĩa
vie	Vietnamese	Vietic				ander	noch
deu	Standard German	Germanic	-er, mehr	als		aur	aur, zyaadaa, adhiik
hin	Hindi	Indic	zyaadaa, adhik, aur	se		bertze/beste	bertze/beste
huk	Basque	Basque	-ago, gox6	baino		phet5 ng4 kai4, phet5 yong4	lang4 ngo4, lang4 tshi4,
hak	Hakka Chinese	Chinese				kai4, m1 chung2 (yong4) kai4	tsai4
hun	Hungarian	Ugric	-bb, több	mint		más, egyeb. más	még
ojs	Seyern Ojibwe	Algonquian	nawac, pitosh	ahpilhc		miina	kiyaapic
bzd	Bribri	Western Isthmic Chibchan	-niru-	tsōā/ tsātā		skā	kí
kal	Greenlandic Inuktitut	Eskimo	uutli, uukōoti, -kēē t, -tiya,	ABL		alla	
crx	Carrier	Athapaskan	-nas	sihnv/ sihnno			
chr	Cherokee	Southern Iroquoian	-kha				
quc	Central Quiché (K'iche')	Mayan	na, más	chikiwäch, chuwäch		chik	sikwu
maz	Central Mazahua	Otomian	mású			'naja	chik
nhw	Western Huasteca Nahuatl		mas	tlen		-nok/-yok/-ok	xe
tdl	Tagalog	Greater Central Philippine	mas, lalong	kaysa, sa, kaysa sa	higit	iba	pa mas
ban	Bali	Malayo-Sumbawan	-an, lebh	teken	me-lebh-in	tosan	buin, malih
hau	Hawaiian	Oceanic	a e, keu	ma mua	'oi	kekahi, ha'i, 'oko'a	hou, keu
kto	Kuot	Border	eseter	fen		tas, cergem	tas
mtj	Moskona	East Bird's Head-Sentani	-ki			be, bedua, bino, gue	mitka
byx	Qaet	Baining				ta	
bbp	Barai	Kolarian				na-	mawinta?
mlp	Bargam	Bargam				temu/temui/temuke	haun
ian	Iatmul	Ndu		-kak	kulak	hale, opi	noro
big	Blangai	Kunimaipian	-ni	-yegí	wol	'kopore	ũ'ũ, ver
ghs	Guhu-Samane	Guhu-Samane	ca	kopo	tlemta	fúesú	ti'she
aey	Anele	Mabuso		ve kŷ		ipa	
xla	Kamula	Kamula-Elevala				w rtsa (huētsa), t (tēxē)	
car	Carib	Cariban	kāfór, mē			peya, huq kaq, -waq	huqtawan
kpg	Kaingang	Je Meridional	ti'she			mboae, amboae, jeyv, ju	-ve, -ve ju
con	Cofán	Cofán	-xera /-xēra /-xira, -māē		yexera		
tca	Tcuna	Tcuna					
cot	Caquite	Souther Mainpuerian					
cao	Chacobo	Panoan		-manta	wino (huino)		
quz	Cuzco Quechua	Quechuan	aswan, astawan		yally		
cav	Cavineña	Tacanan	dyaque?	gui			
gun	Mbyá Guaraní	Maweti-Guarani	-ve				

Table 12: Grammar and dictionary survey results: comparison-related markers.

ISO	Language	Repetitive	Non-scalar additive	Scalar additive	Continuation	Colexifying forms
kab	Kabyle	da /da en/da enni/da enna-kin/ da nakin/di /di en	da /da en/da enni/da enna-kin/ da nakin/di /di en, ula	ula	mazal	
som	Somali	mar kale, kol kale	dagno	ank a, sank a	weli ahunam, ahunam gana,	
amh	Amharic				zaremm	
hau	Hausa	kara (āra)	kua/kwa, mā	har, mā, ko	hā yānzū	
swh	Swahili	tēna	piā	na-	-ngali, bado	
yor	Yoruba	lẹkẹjẹ (lẹkẹjẹ), tun (tún)	pẹlẹ-pẹlẹ	bótlẹjẹpẹ, bótlẹjẹpẹ	sibẹ-sibẹ	
ibo	Igbo	-ghari, ọzọ	-kwa, -kwu, azii	sax	-riri/-liri	
wol	Wolof	-ati, -aat	itam, tam		-ati, des, dese	
kyq	Kenga	daāla	kici		utú/t', r k, ̀r̀	
shk	Shilluk	b r, c ka, kete			nuti	
bvr	Burarra	banwa, mola	rrapa		burdak	
gyn	Kuku-Yalanji	baja	ngarra, yalariku		ku	
djr	Djambarrpuyngu	bulu, biyapul	yurr, yurniha		badak	
tur	Turkish	daha	-da	bile	daha, halá	
vie	Vietnamese	nũa, lạy	cđng	thăm chí	luôn, còn, hãy, vãn	
deu	Standard German	wieder	auch	sogar	noch	
hin	Hindi	phir	bhii, aur	bhii	abhi	
eus	Basque	berriro	ere/are	ere/are	oraindik	
hak	Hakka Chinese	yü4, tsai4	ya1, tu1, in5, lien2	lien2	han2	
hun	Hungarian	ismét, újra, megint, ismételten, újbol	is	raadásul, söt	eddig, még mindig	
ois	Severn Ojibwe	miina, kihtwaam, napi-	kaye, lewiin/kenin/kewinawa/ kekinawint/kekinawaa/kekin		kiyaapic	
bzd	Bribri	ská	afles		-iã	
kal	Greenlandic Inuktitut	kingmut, qqig	aamma, sinna -ttaaq	allaat, -luunniit	suli	
crx	Carrier	k a, ne#	gh , weh/wah	etah		
chr	Cherokee	li-/w-, nvkwale, ja-	xukuje'		-si, siinv, ni-	
quc	Central Quiché (K'iche')	jumul chik	nde, xo		k'a maja	
maz	Central Mazahua	ndo	nokkia, no, -nok/-yok/-ok		xe	
nhw	Western Huasteca Nahuatl	sampa, sempa	man, kahit (na)		noha	
tg1	Tagalog	mul/uli	masih, taler	kadi rasa	pa	
ban	Bali	buin	ho'i, kekahi		enu, buin	
haw	Hawaiian	hou	get		koe	
kto	Kuot	ina	tin		ka	
mtj	Moskona	tas	saqi	tin	ros	
byx	Qaqet	saqi	fudua, dua		as	
bbx	Barai	una, to, kuke	-no		uria	
mlp	Bargam	ta, ton	da, okwi, -o		te-mo	
ian	Iatmul	nak li'vi'	-rau/-tau		wata	
big	Biangai	ngerike	paha		ngoluk	
ghs	Guhu-Samane	haun	ha, ca		agobara	
aez	Amele	ma/madipo/mamiske/matle	ma			
xla	Kamula	rapa/raa, -poty, konaa, konapa	erapa/enapa/enaa	ala	COMP-ADDT: ca	
car	Carib	mān	ge, gé, ke gé, mari	paro	ADDT-CONT: noro	
kgp	Kaingang	khase	-khe	mari	ADDT-CONT: ver	
con	Cofán	wéna	-ta		COMP-ADDT: ti'tshe	
tca	Ticuna	-ah	aisa/aisati			
cot	Caquinte	t k (tëqué)	ri, ari, háska (jasca)			
cao	Chacobo	yapamanta, huqmanta,	hinallataq, -pas/-pis,			
quz	Cuzco Quechua	waqmanta, -pa	-puwan/-piwan			
cav	Cavineña	-nuka	tumebae, -nuka			
gun	Mbyá Guaraní	ju, jey	avei, guive	eteve	-jari	COMP-ADDT: ve
					teri	

Table 13: Grammar and dictionary survey results: additive and continuation markers.

ISO	Language	References
kab	Kabyle	Dallet (1982); Quitout (1997)
som	Somali	Adam (1999); de Larjasse (1997); Farah (1995); Saeed (1999)
anh	Anharic	Leslau (1995); Stassen (1985)
hau	Hausa	Howeidy (1953); Jaggar (2001); Newman (2000); Newman (2007); Robinson (1913)
swh	Swahili	Ashton (1947); Mpiranya (2015)
yor	Yoruba	Beck et al. (2009); Samuel Olanrewaju (p.c.)
ibo	Igbo	Echeruo (2001); Emenajo (1987); Emenajo (2015); Nkwoakwu (n.d.); Stassen (1985)
wol	Wolof	Munro & Gave (1997); Nougier Voisin (2002)
kyq	Kenga	Neukom (1995); Palayer (2004); Vandame (1968)
shk	Shilluk	Kohnen (1994); Remijsen & Ayoker (2020)
bvr	Burarra	AUSIL (n.d.); Janguny (n.d.)
gyn	Kuku-Yalanji	Hershberger & Hershberger (1986); Patz (2002)
djr	Djambarrpuyngu	Wilkinson (1991)
tur	Turkish	Beck et al. (2009); Hofstetter (2012); Kornfilt (1997); Cambridge Dictionary (n.d.a); Tureng (n.d.)
vie	Vietnamese	Thompson (1965); Cambridge Dictionary (n.d.b)
deu	Standard German	
hin	Hindi	Bhatt & Takahashi (2007); Koul (2008); Bahri (n.d.)
eus	Basque	de Rijk (2008); Hualde & Ortiz de Urbina (2003)
hak	Hakka Chinese	Chappell & Lamarre (2005); Hashimoto (1973)
hun	Hungarian	Beck et al. (2009); Dictionary.com (n.d.)
ojs	Severn Ojibwe	Todd (1970); Ojibwe People's Dictionary (n.d.); Kwayaciwin Education Resource Centre (2014)
bzd	Bribri	Jara Murillo (2018)
kal	Greenlandic Inuktitut	Fortescue (1984); Sadock (2003); Ordbog.gl (n.d.)
crx	Carrier	Hargus (2007)
chr	Cherokee	Montgomery-Anderson (2008)
quc	Central Quiché (K'iche')	Christenson (2004); Henne (1980); Tzj (n.d.)
maz	Central Mazahua	Amador Hernández (1976); Benitez Reyna (2017); Stewart (1966)
nhw	Western Huasteca Nahuatl	Beller & Beller (1979); Beller N. & Cowan de Beller (1985)
tgl	Tagalog	Schachter & Otanes (1972); Trussel2.com (n.d.); Tagalog.com (n.d.)
ban	Bali	Kesiman & Malherbe (2009); Umiyati et al. (2023); Satyawati (2015); Trussel2.com (n.d.); BASAbali Wiki (n.d.)
haw	Hawaiian	Elbert & Pukui (1979); Hawaiian-grammar.org (n.d.)
kto	Kuot	Chung & Chung (1996); Lindström (2002)
miŋ	Moskona	Gravelle (2010)
byx	Qadjet	Hellwig (2019)
bbb	Barai	Olson (1981); Oraro et al. (2015)
mlp	Bargam	Hepner (2006)
ian	Iatmul	Jendraschek (2012)
big	Biangai	Dubert & Dubert (1978)
ghs	Gulu-Samane	Gamudze et al. (2013); Richert & Hoopusu (2002)
aeŷ	Amele	Roberts (1987, 2016)
xla	Kamula	Routamaa (1994)
car	Carib	Courtz (2008); Stassen (1985)
kgp	Kaingang	Gonçalves (2011); Navarro (2024); Wiesemann (2011)
con	Cofán	ALDP (n.d.); Fischer & Hengeveld (2023)
tca	Ticuna	Anderson (1962); Bertet (2020)
cot	Caquite	O'Hagan (2020); Swift (1988)
cao	Chacobo	Tallman (2018); Zingg (1998)
quz	Cuzco Quechua	Cushman (2001); Faller (2002); Itier (1997); Ladrón de Guevara de Cuadros (1998); Parker (1964); Stassen (1985); von Tschudi (1884)
cav	Cavneña	Camp & Liccardi (1989); Guillaume (2008)
gun	Mbyá Guaraní	Dooley (2016)

Table 14: Grammar and dictionary survey results: sources consulted per language.

Appendix 3: Grammar and Dictionary Survey References

- Adam, S. M. (ed.). 1999. *New student dictionary–dikshaneeriga cusub ee ardayda: English-somali*. London: HAAN Associates Publishing.
- ALDP. n.d. Cofán—a'ingae LingView. <https://cofan-aldp.github.io/LingView>. <https://cofan-aldp.github.io/LingView>. Last accessed on 8/10/2025.
- Amador Hernández, Mariscela. 1976. *Gramática del mazahua de san antonio pueblo nuevo*. México, D.F.: Escuela Nacional de Antropología e Historia Tesis profesional (licenciatura en antropología, especialidad en lingüística).
- Anderson, Doris G. 1962. *Conversational ticuna*. Yarinacocha, Peru: Summer Institute of Linguistics, University of Oklahoma.
- Ashton, E. O. 1947. *Swahili grammar (including intonation)*. London: Longmans, Green and Co. 2nd edn.
- Australian Society for Indigenous Languages (AuSIL). n.d. Burarra dictionary. <https://ausil.org.au/Dictionary/Burarra>. Last accessed on 8/10/2025.
- Bahri. n.d. Bahri english-hindi dictionary. <https://dsal.uchicago.edu/dictionaries/bahri/>. Last accessed on 8/10/2025.
- BASAbali Wiki. n.d. Dictionary – basabali wiki. <https://dictionary.basabali.org>. Last accessed on 8/10/2025.
- Beck, Sigrid & Krasikova, Sveta & Fleischer, Daniel & Gergel, Remus & Hofstetter, Stefan & Savelsberg, Christiane & Vanderelst, John & Villalta, Elisabeth. 2009. Crosslinguistic variation in comparison constructions. *Linguistic Variation Yearbook* 9.
- Beller, Richard & Beller, Patricia. 1979. Huasteca nahuatl. In Langacker, Ronald W. (ed.), *Studies in uto-aztecan grammar, volume 2: Modern aztec grammatical sketches* (SIL Publications in Linguistics 56), 199–306. Dallas, TX: Summer Institute of Linguistics and The University of Texas at Arlington.
- Beller N., Ricardo & Cowan de Beller, Patricia. 1985. *Curso del náhuatl moderno: Náhuatl de la huasteca, vol. 2*. México, D.F.: Instituto Lingüístico de Verano.
- Benítez Reyna, Rufino. 2017. *Vocabulario práctico bilingüe mazahua-español* (Colección Vocabularios en Lenguas Indígenas 1). México, D.F.: Comisión Nacional para el Desarrollo de los Pueblos Indígenas (CDI).
- Bertet, Denis. 2020. *Aspects of tikuna grammar (san martín de amacayacu variety, colombia): Phonology, nominal phrase, predicative phrase*. Lyon: Université Lumière Lyon 2 dissertation.
- Bhatt, Rajesh & Takahashi, Shoichi. 2007. Direct comparisons: Resurrecting the direct analysis of phrasal comparatives. In *17th Semantics and Linguistic Theory (SALT)* 17, 19–36. Ithaca, NY: CLC Publications.
- Cambridge Dictionary. n.d.a. Cambridge dictionary. <https://dictionary.cambridge.org>. Last accessed on 8/10/2025.
- Cambridge Dictionary. n.d.b. Cambridge dictionary. <https://dictionary.cambridge.org>. Last accessed on 8/10/2025.
- Camp, Elizabeth L. & Liccardi, Millicent R. 1989. *Diccionario cavineña-castellano, castellano-cavineña, con bosquejo de la gramática cavineña*. Dallas, TX: Summer Institute of Linguistics.
- Chappell, Hilary & Lamarre, Christine. 2005. *A grammar and lexicon of Hakka: Historical materials from the Basel Mission Library*. Paris: CRLAO, École des Hautes Études en Sciences Sociales.

- Christenson, Allen J. 2004. K'iche'-english dictionary and guide to pronunciation of the k'iche'-maya alphabet. Unpublished manuscript, Brigham Young University.
- Chung, Chul-Hwa & Chung, Kyung-Ja. 1996. Kuot grammar essentials. In *Two non-austronesian grammars from the islands*, vol. 42 (Data Papers on Papua New Guinea Languages), 1–75. Ukarumpa: Summer Institute of Linguistics.
- Courtz, Hendrik. 2008. *A Carib grammar and dictionary*. Toronto, ON: Magoria Books.
- Cusihuamán G., Antonio. 2001. *Gramática quechua: Cuzco-collao* (Serie Saber Andino 1). Cuzco, Perú: Centro de Estudios Regionales Andinos “Bartolomé de Las Casas” (CBC) 2nd edn.
- Dallet, Jean-Marie. 1982. *Dictionnaire kabyle-français: Parler des at mangellat, algérie* (Études Ethno-Linguistiques Maghreb-Sahara). Paris: SELAF (Société d'études linguistiques et anthropologiques de France).
- de Larajasse, Evangelist. 1897. *Somali-english and english-somali dictionary*. London: Kegan Paul, Trench, Trübner & Co., Ltd.
- de Rijk, Rudolf P. G. 2008. *Standard basque: A progressive grammar, volume 1: The grammar*. Cambridge, Massachusetts and London, England: The MIT Press.
- Dictionary.com. n.d. Dictionary.com. <https://www.dict.com>. Last accessed on 8/10/2025.
- Dooley, Robert A. 2016. Léxico guarani, dialeto mbyá: Guarani-português, com informações Úteis para o ensino médio, a aprendizagem e a pesquisa linguística. SIL Brasil / Associação Internacional de Linguística – SIL Brasil.
- Dubert, Raymond E. & Dubert, Marjorie. 1978. Biangai: Lower levels of grammar. Ukarumpa: Summer Institute of Linguistics.
- Echeruo, Michael J. C. 2001. *Igbo-english dictionary: A comprehensive dictionary of the igbo language with an english-igbo index*. Ikeja, Lagos State, Nigeria: Longman Nigeria Plc.
- Elbert, Samuel H. & Pukui, Mary Kawena. 1979. *Hawaiian grammar*. Honolulu: University Press of Hawaii.
- Emenanjo, E. Nolue. 1987. *Elements of modern igbo grammar: A descriptive approach*. Ibadan: University Press Limited.
- Emenanjo, E. Nolue. 2015. *A grammar of contemporary igbo: Constituents, features and processes*. Port Harcourt, Nigeria: M & J Grand Orbit Communications Ltd.
- Faller, Martina T. 2002. *Semantics and pragmatics of evidentials in cuzco quechua*. Stanford, CA: Stanford University dissertation.
- Farah, Abdirahman A. ("Barwaaqo"). 1995. *Abwaan cusub oo af-soomaali iyo af-ingiriisiya: A modern somali-english dictionary*. Ottawa 1st edn.
- Fischer, Rafael & Hengeveld, Kees. 2023. A'ingae (cofán). In Epps, Patience & Michael, Lev (eds.), *Amazonian languages: An international handbook, volume 1: Language isolates i: Aikanã to kandozi-chapra*, 65–123. Berlin: De Gruyter Mouton.
- Fortescue, Michael. 1984. *West greenlandic* (Croom Helm Descriptive Grammars). London and Sydney: Croom Helm.
- Gamudze, Titus & Hadzata, Tom & King, Phil. 2013. Tentative grammar description for the guhu samane language spoken in morobe province. PILAT / SIL Papua New Guinea, Ukarumpa.
- Gonçalves, Solange Aparecida. 2011. *Tempo, aspecto e modo em contextos discursivos no kaingang sul (jê)*. Campinas, SP: Universidade Estadual de Campinas (UNICAMP), Instituto de Estudos da Linguagem Tese de doutorado.

- Gravelle, Gloria J. 2010. *A grammar of moskona: An east bird's head language of west papua, indonesia*. Vrije Universiteit Amsterdam PhD dissertation.
- Guillaume, Antoine. 2008. *A grammar of cavineña* (Mouton Grammar Library 44). Berlin and New York: Mouton de Gruyter.
- Hargus, Sharon. 2007. *Witsuwit'en grammar: Phonetics, phonology, morphology* (First Nations Languages). Vancouver and Toronto: UBC Press.
- Hashimoto, Mantaro J. 1973. *The hakka dialect: A linguistic study of its phonology, syntax and lexicon*. Cambridge: Cambridge University Press.
- Hawaiian-grammar.org. n.d. hawaiian-grammar.org. <http://www.hawaiian-grammar.org>. Last accessed on 8/10/2025.
- Hellwig, Birgit. 2019. *A grammar of qaqet*, vol. 79 (Mouton Grammar Library). Berlin, Boston: De Gruyter Mouton.
- Henne, David L. 1980. *Diccionario quiché-español*. Guatemala, Centro América: Instituto Lingüístico de Verano.
- Hepner, Mark. 2006. Bargam grammar sketch. S.I.L.
- Hershberger, Henry D. & Hershberger, Ruth. 1986. *Kuku-yalanji dictionary* (Work Papers of SIL-AAB, Series B, Volume 7). Darwin, Australia: Summer Institute of Linguistics, Australian Aborigines Branch.
- Hofstetter, Stefan. 2012. Turkish comparatives: the status and semantic contribution of the adverb *daha* and beyond. In Kincses-Nagy, Éva & Biacsi, Mónika (eds.), *The szeged conference: Proceedings of the 15th international conference on turkish linguistics held on august 20–22, 2010 in szeged*, vol. 49 (Studia uralo-altaica), 231–242. Szeged: SZTE Egyetemi Kiadványok (University of Szeged). <https://acta.bibl.u-szeged.hu/16704/>. Last accessed on 8/10/2025.
- Howeidy, A. 1953. *Concise hausa grammar*. London: George Ronald 2nd edn.
- Hualde, José Ignacio & Ortiz de Urbina, Jon (eds.). 2003. *A grammar of basque*, vol. 26 (Mouton Grammar Library). Berlin and New York: Mouton de Gruyter.
- Itier, César. 1997. *Parlons quechua: La langue du cuzco* (Collection Parlons). Paris: L'Harmattan.
- Jaggar, Philip J. 2001. *Hausa*, vol. 7 (London Oriental and African Language Library). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Janguny. n.d. Janguny. Burarra/Gun-nartpa language and dictionary resource site. <https://janguny.com>. Last accessed on 8/10/2025.
- Jara Murillo, Carla Victoria. 2018. *Gramática de la lengua bribri*. San José, Costa Rica: E-Digital ED.
- Jendraschek, Gerd. 2012. *A grammar of iatmul*: Universität Regensburg, Fakultät für Sprach-, Literatur- und Kulturwissenschaften Habilitationsschrift.
- Kesiman, Made Windu Antara & Malherbe, Michel. 2009. *Parlons balinais* (Parlons...). Paris: L'Harmattan.
- Kohnen, B. 1994. *Dizionario shilluk* (Bibliotheca Africana 7). Trieste: Università degli Studi di Trieste.
- Kornfilt, Jaklin. 1997. *Turkish* (Descriptive Grammars). London and New York: Routledge. Series editor: Bernard Comrie.
- Koul, Omkar N. 2008. *Modern hindi grammar*. Springfield, VA: Dunwoody Press.
- Kwayaciiwin Education Resource Centre. 2014. *Anihshiniimowin oji-cree dictionary (severn river and winisk river)*. Sioux Lookout, Ontario: Kwayaciiwin Education Resource Centre draft edn.

- Ladrón de Guevara de Cuadros, Laura. 1998. *Diccionario quechua: Inglés-quechua-español, español-quechua-inglés, quechua-inglés-español*. Lima, Perú: Editorial Brasa.
- Leslau, Wolf. 1995. *Reference grammar of amharic*. Wiesbaden: Harrassowitz Verlag.
- Lindström, Eva. 2002. *Topics in the grammar of kuot: A non-austronesian language of new ireland, papua new guinea*. Stockholm: Stockholm University, Department of Linguistics PhD dissertation.
- Montgomery-Anderson, Brad. 2008. *A reference grammar of oklahoma cherokee*. Lawrence, KS: University of Kansas Ph.D. dissertation.
- Mpiranya, Fidèle. 2015. *Swahili grammar and workbook*. London and New York: Routledge.
- Munro, Pamela & Gaye, Dieynaba. 1997. *Ay baati wolof: A wolof dictionary* (UCLA Occasional Papers in Linguistics 19). Los Angeles: Department of Linguistics, University of California, Los Angeles revised edn.
- Navarro, M. 2024. Nominal countability in Kaingang and beyond. *Glossa: a journal of general linguistics* 9(1). <https://doi.org/10.16995/glossa.9334>. Last accessed on 8/10/2025.
- Neukom, Lukas. 1995. *Description grammaticale du kenga (langue nilo-saharienne du Tchad)*. Köln: Rüdiger Köppe Verlag.
- Newman, Paul. 2000. *The hausa language: An encyclopedic reference grammar* (Yale Language Series). New Haven & London: Yale University Press.
- Newman, Paul. 2007. *A hausa-english dictionary*. New Haven & London: Yale University Press.
- Nkowaokwu. n.d. Nkowaokwu: The igbo dictionary. <https://nkowaokwu.com>. Last accessed on 8/10/2025.
- Nouguier Voisin, Sylvie. 2002. *Relations entre fonctions syntaxiques et fonctions sémantiques en wolof*. Lyon: Université Lumière Lyon 2 dissertation.
- O'Hagan, Zachary. 2020. *Focus in caquinte*. Berkeley, CA: University of California, Berkeley dissertation.
- Ojibwe People's Dictionary. n.d. The ojibwe people's dictionary. <https://ojibwe.lib.umn.edu>. Last accessed on 8/10/2025.
- Olson, Michael Leon. 1981. *Barai clause junctures: Toward a functional theory of interclausal relations*: The Australian National University PhD thesis.
- Oraro, Benjamin & Unini, Dunston & Bradshaw, Rob. 2015. Tentative grammar description for the barai [bbb] language spoken in oro province. PILAT / SIL Papua New Guinea, Ukarumpa.
- Ordbog.gl. n.d. Ordbog.gl. <http://ordbog.gl>. Last accessed on 8/10/2025.
- Palayer, Pierre. 2004. *Dictionnaire kenga (tchad)* (Collection Afrique et Langage 6). Louvain-Paris: Peeters.
- Parker, Gary & others. 1964. *English-Quechua dictionary: Cuzco, Ayacucho, Cochabamba*. Ithaca, NY: Cornell University. Report no. BR-5-1231-14.
- Patz, Elisabeth. 2002. *A grammar of the kuku yalanji language of north queensland* (Pacific Linguistics 527). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Quitout, Michel. 1997. *Grammaire berbère (rifain, tamazight, chleuh, kabyle)*. Paris: Éditions L'Harmattan.
- Remijnsen, Bert & Ayoker, Otto Gwado. 2020. A descriptive analysis of adjectives in shilluk. In *A descriptive grammar of shilluk* (Language Documentation & Conservation Special Publication 14),

- chap. 4. Honolulu: Language Documentation & Conservation, University of Hawai'i Press.
- Richert, Ernest L. & Hoopusu, Ttopoqogo. 2002. *Noo supu: A triglot dictionary (guhu-samane language)*. Ukarumpa, Eastern Highlands Province, Papua New Guinea: Summer Institute of Linguistics. Editors: Doris Bjorkman and Mumure Ttopoqogo.
- Roberts, John R. 1987. *Amele* (Croom Helm Descriptive Grammars). London, New York, Sydney: Croom Helm.
- Roberts, John R. 2016. Amele RRG grammatical sketch. SIL International.
- Robinson, Charles Henry. 1913. *Dictionary of the hausa language, vol. i: Hausa-english*. Cambridge: Cambridge University Press 3rd edn.
- Routamaa, Judy. 1994. Kamula grammar essentials. Summer Institute of Linguistics.
- Sadock, Jerrold M. 2003. *A grammar of kalaallisut (west greenlandic inuttut)* (Languages of the World/Materials 162). Munich: LINCOM Europa.
- Saeed, John I. 1999. *Somali*, vol. 10 (London Oriental and African Language Library). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Satyawati, Made Sri. 2015. Grammatical analysis of balinese adjectives. *International Journal of Linguistics* 7(3). 109–128. <https://doi.org/10.5296/ijl.v7i3.7706>
- Schachter, Paul & Otnes, Fe T. 1972. *Tagalog reference grammar*. Berkeley, Los Angeles, London: University of California Press. Prepared under the auspices of the Philippine Center for Language Study.
- Stassen, Leon. 1985. *Comparison and universal grammar*. Oxford: Basil Blackwell.
- Stewart, Donald. 1966. Gramática de mazahua. Unpublished manuscript (ms. 1966), Instituto Lingüístico de Verano (SIL International).
- Swift, Kenneth E. 1988. *Morfología del caquinte (arawak preandino)* (Serie Lingüística Peruana 25). Lima, Perú: Ministerio de Educación / Instituto Lingüístico de Verano.
- Tagalog.com. n.d. tagalog.com. <https://www.tagalog.com>. Last accessed on 8/10/2025.
- Tallman, Adam J. 2018. *A grammar of chácobo, a southern pano language of the northern bolivian amazon*. Austin, TX: University of Texas at Austin dissertation.
- Thompson, Laurence C. 1965. *A vietnamese grammar*. Seattle: University of Washington Press.
- Todd, Evelyn Mary. 1970. *A grammar of the ojibwa language: The severn dialect*. Chapel Hill, NC: University of North Carolina at Chapel Hill Ph.D. dissertation.
- Trussel2.com. n.d. trussel2.com. <https://www.trussel2.com>. Last accessed on 8/10/2025.
- Tureng. n.d. Tureng dictionary. <https://tureng.com>. Last accessed on 8/10/2025.
- Tzij. n.d. Tzij: K'iche' maya language resources. <https://tzij.coerll.utexas.edu/>. Last accessed on 8/10/2025.
- Umiyati, Mirsa & Darmawan, Ida Bagus Yoga Danu & Sujaya, Nyoman. 2023. The grammatical behaviour of balinese adjectives on phrases and clauses. In *Proceedings of the international seminar on contemporary linguistics (iscl 2022)*, vol. 734 (Advances in Social Science, Education and Humanities Research), 421–437. Atlantis Press.
- Vandame, Charles. 1968. *Grammaire kenga* (Études Linguistiques 2). Lyon: Afrique et Langage.
- von Tschudi, Johann Jakob. 1884. *Organismus der khetšua-sprache*. Leipzig: F.A. Brockhaus.
- Wiesemann, Ursula Gojtéj. 2011. *Dicionário kaingang – português / português – kaingang*. Curitiba, PR: Editora Evangélica Esperança 2nd edn.

- Wilkinson, Melanie P. 1991. *Djambarrpuyngu: A yolŋu variety of northern Australia*. Sydney: Department of Linguistics, University of Sydney dissertation.
- Zingg, Philippe. 1998. *Diccionario chácobo-castellano, castellano-chácobo con bosquejo de la gramática chácobo y con apuntes culturales*. La Paz: Ministerio de Desarrollo Sostenible y Planificación Viceministro de Asuntos Indígenas y Pueblos Originarios.

Appendix 4: Automatic Extraction Results

ISO	Cat.	Form	Freq.	Prop.	ISO	Cat.	Form	Freq.	Prop.
ae	COMP	cnu	87	0.33	hin	INCR	\$fir\$	53	0.35
		\$wool	85	0.32			\$bahut\$	14	0.09
ae	CONT	\$bi	90	0.46			\$svam	10	0.07
		\$wele\$	47	0.24	hun	COMP	bb	145	0.52
		\$hau	52	0.35			an\$	66	0.33
ae	INCR	agec	13	0.09	hun	CONT	\$még\$	36	0.18
		\$fem\$	7	0.05			mig\$	17	0.08
		hadume	4	0.03			\$mai\$	7	0.03
		\$gunacil	3	0.02			\$ismét\$	33	0.22
		À	149	0.55	hun	INCR	\$más	27	0.18
amh	COMP	\$ÉΔΦ\$	57	0.21			\$még\$	15	0.10
		\$fσZΠ	25	0.09			\$újra\$	5	0.03
		ΠΔΠ	8	0.03	ian	COMP	\$nima	95	0.35
		\$Γς	37	0.18			\$kwula	49	0.18
		γΛ	20	0.10	ian	CONT	\$l	106	0.54
amh	CONT	\$Λήη	18	0.09			ala\$	72	0.37
		\$Λξ\$	11	0.05			\$sa	38	0.25
		\$γΛ\$	6	0.03	ian	INCR	\$njambi\$	20	0.13
		ትዩ	6	0.03			\$b	18	0.12
		\$ΛήηUፓ\$	3	0.01			\$wogwendi\$	7	0.05
		\$ደግ	35	0.23			\$nk	113	0.41
amh	INCR	\$ΛΛ	23	0.15	ibo	COMP	kari	104	0.37
		ዙ\$	17	0.11			\$ókè\$	15	0.05
		ገς	16	0.11			nq\$	61	0.30
		aj	76	0.27	ibo	CONT	be\$	47	0.23
		\$bandingan	21	0.08			\$ub	28	0.14
		sayan\$	19	0.07			digide\$	7	0.03
		\$luihan\$	16	0.06	ibo	INCR	\$ozo\$	71	0.47
ban	COMP	\$langkung	14	0.05			\$m'g	12	0.08
		\$becika	12	0.04			wa	133	0.48
		\$leb	9	0.03	kab	COMP	\$akter	51	0.18
		\$padaang\$	7	0.03			ugar	22	0.08
		\$tikel\$	7	0.03			\$axir	16	0.06
		\$utamaaan\$	6	0.02			\$mazal	55	0.27
		kantun\$	50	0.25			sl	27	0.13
ban	CONT	ngkin	21	0.10	kab	CONT	mmu	16	0.08
		durung\$	20	0.10			\$ass-a\$	6	0.03
		\$enu\$	12	0.06			\$lhedd\$	3	0.01
		\$malih\$	58	0.38			\$jeddd\$	3	0.01
ban	INCR	\$buin\$	12	0.08			\$nniden\$	33	0.22
		ios	10	0.07			aye	26	0.17
bbb	COMP	ka\$	197	0.75	kab	INCR	\$wussan\$	10	0.07
bbb	CONT	\$uria	80	0.42			ddam	7	0.05
		vi	76	0.40			\$ieawed\$	4	0.03
bbb	INCR	una\$	91	0.64			neru	100	0.44
big	COMP	\$panu	126	0.46	kal	COMP	aasi	32	0.14
		pur	122	0.44			\$100-ria	6	0.03
		\$ngoluk\$	73	0.36			\$suli\$	33	0.21
big	CONT	erik	27	0.13			\$to	23	0.15
		libu	11	0.05			maan	17	0.11
big	INCR	\$pulaga\$	40	0.26	kal	CONT	utstillugu\$	7	0.04
		erek	38	0.25			allara	6	0.04
bvr	COMP	\$jarra\$	107	0.39			ersinnagi	4	0.03
		uya\$	37	0.13			ornannile	3	0.02
bvr	CONT	\$burdak\$	50	0.25			assissutig	3	0.02
		\$mu-ng	43	0.21			\$oqarluni	12	0.11
		\$ana-munya	12	0.06			merl	12	0.11
bvr	INCR	\$barrwa\$	42	0.28	kal	INCR	qqipp	11	0.10
		\$mola\$	27	0.18			lla\$	10	0.09
		\$an-nerran	17	0.11			\$kingumut\$	7	0.06
		ik	167	0.61			oqaasissaq	4	0.04
byx	COMP	eviit\$	55	0.20	kgp	COMP	\$há\$	96	0.34
		\$malai\$	28	0.10			\$káfor\$	86	0.31
		\$as\$	47	0.24			\$ver\$	67	0.33
byx	CONT	atiaskerl	38	0.19	kgp	CONT	\$ûri\$	13	0.06
		sa\$	37	0.19			\$jógjóg\$	7	0.03
		askerlka\$	46	0.31			\$semana\$	4	0.02
byx	INCR	\$ian	28	0.19	kgp	INCR	\$mā	68	0.45
		\$buup\$	14	0.09			\$ver\$	14	0.09
		ta\$	89	0.33			ase	66	0.24
bzd	COMP	ki	36	0.13	kto	COMP	kurup	64	0.23
		bua'ie	17	0.06			o\$	13	0.05
		tso'i	41	0.21			\$maningan\$	4	0.01
bzd	CONT	cia\$	26	0.13	kto	CONT	\$ka\$	83	0.42
		bak	26	0.13			nama\$	47	0.24
		kam	22	0.11	kto	INCR	\$non\$	64	0.42
		skà\$	47	0.32			\$inagat\$	22	0.15
bzd	INCR	ki\$	19	0.13	kyq	COMP	cir	119	0.43
		\$dene\$	7	0.05	kyq	CONT	\$eyo\$	73	0.26
		\$tsé	7	0.05			utu	110	0.54
		\$oque	123	0.44			\$daala\$	46	0.31
cao	COMP	\$to	91	0.33	kyq	INCR	\$kuuy\$	25	0.17
		\$huino\$	10	0.04			na\$	19	0.13
		\$jari\$	50	0.25			\$xenda\$	107	0.39
		\$jaboqu	24	0.12	maz	COMP	\$me\$	68	0.25
cao	CONT	nox	18	0.09			\$ciento\$	10	0.04
		tia\$	11	0.06			\$xe\$	55	0.28
		ijan	7	0.04	maz	CONT	\$ya\$	47	0.24
		\$huëtsa\$	36	0.24			\$be\$	18	0.09
cao	INCR	ari\$	31	0.21			\$ndänt'ä\$	5	0.03

Table 15: Extracted Markers, Part 1 of 3

ISO	Cat.	Form	Freq.	Prop.	ISO	Cat.	Form	Freq.	Prop.
car	COMP	tèquéniqué	20	0.13	maz	INCR	\$xe\$	47	0.31
		itèquécà	3	0.02			\$ye	36	0.24
		ko'p	104	0.37			nad	124	0.46
		\$ka'tu\$	37	0.13			uriyam	63	0.23
car	CONT	\$noro\$	69	0.34	mlp	CONT	\$os	103	0.51
		\$erome	15	0.07			\$tuta	19	0.09
		uwap	15	0.07			\$ta\$	80	0.54
		tamurumbo\$	4	0.02			dui\$	112	0.40
car	INCR	\$aimytonon	3	0.01	mtj	COMP	\$ekris\$	105	0.38
		rapa\$	73	0.48			jog\$	84	0.43
		\$noro\$	22	0.14			\$ros	74	0.38
		\$dyaque	145	0.53			\$tas\$	91	0.62
cav	COMP	\$jari	74	0.38	nhw	COMP	\$más\$	138	0.50
		\$majuhu	15	0.08			\$ci	14	0.05
		patya	12	0.06			ema	69	0.34
		\$cuanucahu	3	0.02			\$noja\$	68	0.34
cav	INCR	nucacu	44	0.30	nhw	INCR	\$sempa	37	0.24
		ca\$	40	0.27			\$seyoc\$	32	0.21
		bajeje\$	7	0.05			más\$	27	0.18
		\$uri	113	0.41			\$a. < < < \$	87	0.31
chr	COMP	gadv	13	0.05	ojs	COMP	\$<\$	46	0.16
		losvs	10	0.04			\$< < < < \$	45	0.16
		uneqoc	6	0.02			! < < < < \$	6	0.02
		\$asi	68	0.34			\$9 < < < < \$	45	0.22
chr	CONT	iya	57	0.28	ojs	CONT	\$ < < < < \$	37	0.18
		talin	40	0.26			\$ < < < < \$	10	0.05
		adale\$	20	0.13			\$ < < < < \$	7	0.03
		idas	14	0.09			\$ < < < < \$	5	0.02
chr	INCR	\$asi\$	11	0.07	ojs	INCR	\$ < < < < \$	4	0.02
		neqoccele	3	0.02			\$ < < < < \$	57	0.38
		\$ti'tsse	130	0.48			\$ < < < < \$	26	0.17
		\$bove	25	0.09			\$na\$	146	0.53
con	CONT	\$toya\$	98	0.49	quc	COMP	\$cä\$	72	0.36
		\$ja'	25	0.12			etak	40	0.20
		\$ccase\$	42	0.28			chic\$	38	0.25
		\$faesu	34	0.23			\$jumul\$	31	0.21
cot	COMP	maja	104	0.38	quc	INCR	\$xinw	7	0.05
		anaa	89	0.32			\$aswan	127	0.46
		tiraja	46	0.23			\$astawan	20	0.07
		avent	24	0.12			sha	139	0.69
cot	CONT	\$tequerata	20	0.10	quz	CONT	\$jesus	36	0.24
		\$areje	14	0.07			tawan	29	0.19
		\$nochoocat	5	0.03			\$wak	24	0.16
		\$shinejiac	3	0.02			\$mai\$	155	0.56
cot	INCR	tsipa\$	29	0.19	ron	COMP	\$ci	61	0.30
		tsipapae	16	0.11			\$incä\$	35	0.17
		ajari	12	0.08			cum\$	25	0.12
		\$isoto	7	0.05			\$mai\$	38	0.25
cot	INCR	\$mavite\$	7	0.05	ron	INCR	\$iaräs\$	18	0.12
		manajir	4	0.03			\$alt\$	13	0.09
		nonintavet	4	0.03			\$nou\$	11	0.07
		nus\$	126	0.45			di\$	62	0.22
crx	COMP	ltsuk	12	0.04	shk	COMP	\$kiböl	54	0.19
		\$6o\$	4	0.01			\$lod	26	0.09
		\$sawhuz\$	73	0.36			\$miy	11	0.04
		za	44	0.22			nudd\$	49	0.24
crx	CONT	utso\$	16	0.08	shk	CONT	\$nu	37	0.18
		\$doo\$	48	0.32			nëen	12	0.06
		on\$	31	0.21			píd	6	0.03
		als\$	114	0.41			\$köbb	22	0.14
deu	COMP	\$eu	51	0.18	shk	INCR	\$ög	21	0.14
		\$noch\$	86	0.43			\$megö\$	17	0.11
		\$ges	36	0.18			\$ärye	13	0.09
		\$noch\$	45	0.30			cigi	11	0.07
deu	INCR	\$wieder\$	20	0.13	som	CONT	\$këti\$	11	0.07
		\$ander	19	0.12			\$laakiim\$	53	0.19
		ithi	104	0.40			\$badan\$	42	0.15
		\$djulkmaram	42	0.16			\$masiix	21	0.08
djr	COMP	\$badak	52	0.27	som	COMP	wanaagsan\$	20	0.07
		lj'	27	0.14			\$xisaab	10	0.04
		\$balanyami	23	0.12			int	64	0.32
		\$bul	65	0.45			\$weli\$	39	0.19
djr	INCR	\$wiripupu	16	0.11	som	INCR	\$kale\$	63	0.41
		\$baino\$	80	0.30			\$badan\$	17	0.11
		iza	60	0.22			hi	89	0.33
		iago	26	0.10			\$kuliko\$	74	0.27
eus	COMP	ehun\$	10	0.04	swh	COMP	\$zaidi\$	48	0.18
		ela	65	0.33			\$bado\$	55	0.27
		aindi	40	0.20			kati	49	0.24
		iro\$	37	0.25			lipoku	22	0.11
eus	INCR	beste\$	32	0.21	swh	CONT	\$kabla\$	12	0.06
		\$ontz	8	0.05			\$tena\$	46	0.30
		\$uztara	3	0.02			id	29	0.19
		\$anipamu	53	0.21			ingine\$	18	0.12
ghs	COMP	\$nikeho	31	0.12	tca	COMP	yexera	114	0.42
		\$suqobi	18	0.07			xcax\$	49	0.18
		itorai\$	17	0.07			memae	17	0.06
		\$nookaama	9	0.04			xm	103	0.52
ghs	INCR	o\$	8	0.03	tca	CONT	yane\$	23	0.12
		\$dzeimake\$	3	0.01			\$to\$	34	0.23

Table 16: Extracted Markers, Part 2 of 3

ISO	Cat.	Form	Freq.	Prop.	ISO	Cat.	Form	Freq.	Prop.
ghs	CONT	baa	76	0.39	tgl	COMP	\$wena	34	0.23
		odza	19	0.10			aanewa\$	10	0.07
		\$mar	9	0.05			\$lalo	76	0.27
		\$qoroqi\$	5	0.03			\$kun	58	0.21
		\$sagana	3	0.02			igit\$	33	0.12
ghs	INCR	okatanita\$	3	0.02	tgl	CONT	\$nan	65	0.32
		\$paha\$	69	0.48			mantalang\$	37	0.18
gun	COMP	\$kina\$	4	0.03	tgl	INCR	\$pa\$	35	0.17
		ve\$	205	0.73			\$muli	39	0.26
gun	CONT	\$teri	79	0.39	tgl	INCR	\$ibang\$	17	0.11
		\$ay	21	0.10			\$marami	14	0.09
gun	INCR	mbyt	11	0.05	tur	COMP	\$pang\$	11	0.07
		\$ju\$	70	0.46			aha\$	99	0.38
gvn	COMP	tave	13	0.09	tur	CONT	\$yar	25	0.10
		aku\$	116	0.42			\$yüz\$	10	0.04
gvn	CONT	\$jarra\$	57	0.21	tur	INCR	\$üstündür\$	8	0.03
		iku	78	0.40			\$ol	74	0.39
gvn	INCR	ter	13	0.07	vie	COMP	ken\$	49	0.26
		-balkawa	9	0.05			\$hâlâ\$	12	0.06
gvn	CONT	larinya\$	5	0.03	vie	INCR	\$daha\$	28	0.20
		\$baja\$	63	0.43			\$yine\$	24	0.17
hak	COMP	\$yindu	34	0.23	wol	COMP	\$baþka\$	12	0.09
		kha	92	0.33			'de	7	0.05
hak	CONT	m-he\$	23	0.08	wol	INCR	çiler\$	5	0.04
		-sut\$	20	0.07			\$hon	123	0.44
hak	INCR	en-kâ\$	9	0.03	wol	CONT	\$çij\$	40	0.14
		\$yang-ko\$	7	0.03			\$côn\$	82	0.41
hak	COMP	pak-phi\$	6	0.02	wol	INCR	\$dang\$	28	0.14
		-lâng	5	0.02			\$laj\$	44	0.29
hak	CONT	hân\$	54	0.27	wol	COMP	\$nûta\$	31	0.21
		\$hân-	37	0.18			\$khác	17	0.11
hak	INCR	-kin\$	31	0.15	wol	CONT	\$gên	73	0.26
		\$yu\$	57	0.38			ol	71	0.26
hak	COMP	\$hân-yû\$	8	0.05	wol	INCR	\$ëpp	27	0.10
		\$pâ-	7	0.05	wol	CONT	\$raw\$	7	0.03
hau	CONT	ot\$	7	0.05			\$tane\$	5	0.02
		\$chhon-tô\$	3	0.02	wol	INCR	\$jógandoo\$	4	0.01
hau	COMP	\$fi\$	71	0.26			\$bi\$	56	0.28
		\$fiye\$	31	0.11	wol	CONT	\$ngi\$	53	0.26
hau	INCR	\$kara\$	16	0.06			\$tey\$	20	0.10
		\$mafi	11	0.04	wol	INCR	\$nekke	11	0.05
hau	CONT	yalwa	6	0.02			eneen\$	37	0.24
		\$ribi\$	4	0.01	xla	COMP	aat\$	28	0.18
hau	INCR	zu\$	40	0.20			\$bare	16	0.11
		\$tun\$	36	0.18	xla	CONT	\$kok	158	0.58
hau	COMP	\$yana\$	33	0.16			tla:mt	42	0.15
		\$sâke	29	0.19	xla	INCR	\$ala\$	88	0.45
hau	CONT	\$kara	19	0.12			-p	17	0.09
		iyu\$	11	0.07	xla	COMP	\$popapapa\$	3	0.02
hau	INCR	\$mala'ika\$	8	0.05			\$matl	27	0.18
		\$ubangida\$	4	0.03	xla	CONT	ab	12	0.08
haw	COMP	naki\$	4	0.03			iwa\$	11	0.07
		\$turo\$	2	0.01	yor	COMP	\$mamisika\$	10	0.07
haw	CONT	ikai	46	0.19			posolotama	8	0.05
		\$oi\$	42	0.17	yor	INCR	\$omhalane\$	4	0.03
haw	INCR	\$nu	36	0.15			\$hauwall	4	0.03
		\$aho\$	14	0.06	yor	CONT	mo-pasatam	3	0.02
hin	COMP	\$keu\$	7	0.03			\$jû	75	0.27
		wa\$	58	0.32	yor	INCR	fe\$	35	0.13
hin	CONT	\$ana\$	58	0.32			\$ju\$	17	0.06
		\$oi\$	3	0.02	yor	COMP	\$pöju\$	14	0.05
hin	INCR	\$hou\$	63	0.46			\$tubo\$	7	0.03
		nui	26	0.19	yor	CONT	igba	78	0.39
hin	COMP	hee\$	103	0.37			\$ns	16	0.08
		\$bad	74	0.27	yor	INCR	iyi\$	15	0.07
hin	CONT	adhaik\$	29	0.10			pe	60	0.39
		ate	54	0.27	yor	COMP	\$miran\$	25	0.16
		\$tha\$	49	0.24			\$tún\$	23	0.15
hin	INCR	\$tak\$	43	0.21					

Table 17: Extracted Markers, Part 3 of 3