

1 A cross-linguistic study of emphatic negative coordination

2

3 Iker Salaberri

4 University of the Basque Country (UPV/EHU)

5

6 The purpose of this paper is to provide a detailed cross-linguistic analysis of so-called
7 emphatic negative coordination (ENC). This kind of clause linkage is illustrated by
8 *neither* and *nor* in *She neither could nor would speak lightly of the accident*. On the
9 basis of a 250-language sample, the paper lays out a new typology of ENC meant to gain
10 novel insights. It is shown that languages can combine ENC types, and that contact and
11 borrowing are relevant triggers for the emergence of this sort of clause linkage. The
12 article also reveals that there is considerable variety in the etymological sources and
13 grammaticalization paths of ENC markers.

14

15 **Keywords:** negation, clause linkage, typology, language contact, grammaticalization

16

17 1. Introduction

18

19 Despite the recent, increased interest in the grammatical properties of negation and
20 clause linkage, there is to date no extensive typological study of so-called emphatic
21 negative coordination (ENC),¹ which combines both functions: contributions to this topic
22 have only been made in passing within broader studies on negation (Payne 1985a,
23 Bernini & Ramat 1996, Miestamo et al. 2015, De Swart 2020), coordination
24 (Haspelmath 2007, Mauri 2008) and the relationship between negation and clause
25 linkage (Bond 2011).²

26 As a result, this phenomenon is poorly understood: it is unclear whether ENC is

1 This function has different labels depending on which of its properties is highlighted. ‘Rejection’ (Payne 1985b) indicates that this is neither a kind of conjunction nor a kind of disjunction, ‘connective negation’ (Van der Auwera 2021) emphasizes that it serves to connect and negate, and ‘emphatic negative coordination’ (Haspelmath 2007) points out its emphatic nature. For reasons explained in Sections 2 and 4 the latter term will be used throughout.

2 Crystal (2008: 323) defines negation as “a process or construction in grammatical and semantic analysis which typically expresses the contradiction of some or all of a sentence’s meaning”. In turn, Dik (1997: 189) characterizes coordination as “a construction consisting of two or more members which are functionally equivalent, bound together at the same level of structure by means of a linking device”.

27 specific to European languages, or whether it is more widespread (Haspelmath 2007:
28 17), the connection of ENC to functional motivations has not been explored in depth, and
29 the diachronic developments undergone by ENC are not well-understood either
30 (Haspelmath 2007: 17–19).

31 In view of the former, this study aims to answer the following questions: (1) how
32 widespread is ENC in the languages of the world? (2) which restrictions (if any) govern
33 the use of ENC? (3) what is the areal distribution of ENC, and how does this function
34 spread under contact? (4) do ENC markers tend to be distinct from markers of standard
35 negation (Miestamo 2005: 42)? (5) do ENC markers tend to be etymologically related to
36 markers of standard negation and, if not, what are their etymological sources?

37 This study is structured as follows: Section 2 is dedicated to delimiting the object of
38 study of this paper, followed in Section 3 by a layout of the methodology and sampling
39 procedure. Section 4 provides an overview of the generalizations that can be made on
40 the basis of the data, and the grammaticalization paths for ENC markers are discussed in
41 Section 5. Finally, Section 6 presents the conclusions.

42

43 **2. The domain of inquiry**

44

45 This paper starts out from the premise that negative coordination and, by extension,
46 emphatic negative coordination are distinct functions whose structural expression can be
47 studied across languages. The following (1-2) are comparative concepts meant to enable
48 cross-linguistic comparison of negative coordination and emphatic negative
49 coordination respectively:

50

51 (1) Negative coordination

52 Negative coordination is a kind of clause linkage

53 (i) whereby two or more functionally equivalent members are linked together at
54 the same level of structure by an overt or covert linking device,

55 (ii) and whereby all linked members are under the scope of negation, whether of
56 the same or of distinct negators.

57 (2) Emphatic negative coordination

58 Emphatic negative coordination is a kind of clause linkage

- 59 (i) whereby two or more functionally equivalent members are linked together at
 60 the same level of structure by an overt or covert linking device,
 61 (ii) whereby all linked members are under the scope of negation, whether of the
 62 same or of distinct negators,
 63 (iii) and whereby it is emphasized that the coordinated members are part of a
 64 coordination structure and are thus considered separately.³

65

66 These comparative concepts (1-2) are expressed by means of language-specific
 67 constructions, morphemes and categories. Some examples of emphatic negative
 68 coordination are given in (3a-c):

69

70 HUP (Naduhup > Eastern Naduhup, Brazil, Colombia; Epps 2008: 736)

- 71 (3) a. *Yíkán né cəkʷ'ət ʔid-nih-ĩp*
 72 over.there EMPH.NEG Tucano speak-NEG-DEP
 73 *né potugěc wiʔ-nih-ĩp*
 74 EMPH.NEG Portuguese understand-NEG-DEP
 75 ‘There, neither speaking Tucano nor understanding Portuguese (there I
 76 arrived)’⁴

77 BASQUE (Language isolate, France, Spain; Hualde & Ortiz de Urbina 2003: 563)

- 78 b. *Ez zituen ez hitz-ak leun-du ez*
 79 NEG AUX NEG word-DEF.PL.ABS soften-PFV NEG
 80 *gorroto-ak estal-i*
 81 hatred-DEF.PL.ABS hide-PFV
 82 ‘(S)he neither softened her/his words nor hid her/his hatred’

83 SIRAYAIC (Austronesian > East Formosan, Taiwan PC; Adelaar 2011: 219)

- 84 c. *Araraw-aw ta ay-ayam ka säwäbix tu*
 85 look.at-SBJV.UO NOM RDP-animal LK spread.out LOC
 86 *vũlũm, k'='äsi-ahaw dma-diri, k'='äsi-ahaw*
 87 cloud LK=NEG-either AO3.RDP-sow LK=NEG-either
 88 *ma-ayam, k'='äsi-ahaw ma-'lidtu kuvaw*
 89 AO1-reap LK=NEG-either AO4-collect LOC barn

3 See Section 4.1 for arguments in favor of emphasis as an inherent property of ENC.

4 Interlinear glosses follow the *Leipzig glossing rules* (Bickel et al. 2015).

90 ‘Look at the birds of the air: they neither sow nor reap nor gather into
91 barns’

92

93 In (3a) from Hup, the functionally equivalent members *cɔkw’ət ʔidnihĩp* ‘not speaking
94 Tucano’ and *potugēc wiʔnihĩp* ‘not understanding Portuguese’ are coordinated by means
95 of the negative marker *né*, which also conveys emphasis. Negation is expressed by both
96 instances of *né* as well as by negative marking on each coordinated verb (*-nih*). This is
97 different from negative coordination, which usually ensues via juxtaposition in
98 combination with negative marking on the verb (Epps 2008: 334), as illustrated by (4).

99

100 HUP (Naduhup > Eastern Naduhup, Brazil, Colombia; Epps 2008: 334)

101 (4) *Hĩp, cɔc, wǎn, mɔm j’ám pǎ-ǎhǎ?*
102 grater hoe knife axe DST.CNTR NEGEX-TAG2
103 ‘There used to be no graters, hoes, knives, (or) axes’

104

105 In (4), unlike in (3a), there is no presence of the emphatic negative marker *né* and, in
106 fact, no overt coordinator at all. Negation is marked only once on the verb by means of
107 the prefix *pǎ-*, and all coordinands are under the scope of this negative marker.
108 Therefore, one could argue that in Hup negative coordination (4) and emphatic negative
109 coordination (3a) are distinctly realized.

110 In (3b) from Basque, the second and third occurrences of the standard negator *ez*
111 each take scope over the coordinated verb phrases *hitzak leundu* ‘soften words’ and
112 *gorrotoak estali* ‘hide hatred’, respectively. However, negation is already expressed by
113 the first instance of *ez*, which takes scope over the whole clause. Consequently, one may
114 suggest that the second and third occurrences of *ez* serve to emphasize that the
115 coordinated members are part of a coordination structure and are thus considered
116 separately. This suggestion is further supported by the fact that two instances of *ez*
117 (instead of three) are enough to express negative coordination (5a-b):

118

119 BASQUE (Language isolate, France, Spain; Hualde & Ortiz de Urbina 2003: 509, 893)

120 (5) a. *Ez diot Mikel-i eman liburu-a, eta ez*
121 NEG AUX Mikel-DAT give.PFV book-DEF.ABS and NEG

122 *Aitorr-i*
 123 Aitor-DAT
 124 ‘I did not give the book to Mikel, and not to Aitor’
 125 b. *Ez=pai-takite batzu-ek nola eskiriba, eta ez*
 126 NEG=SUB-know some-ERG.PL how write and NEG
 127 *nola irakur*
 128 how write
 129 ‘Because some do not know how to write it, and neither how to read it’
 130

131 Again, this suggests that negative coordination (5a-b) and emphatic negative
 132 coordination (3b) are different functions in Basque and consequently have distinct
 133 means of expression. In (3c) from Sirayaic the verb phrases *dmadiri* ‘they sow’,
 134 *maayam* ‘they reap’ and *ma’lid* ‘they collect’ are each linked together and negated by
 135 one coordinator *k’āsiahaw*, which also emphasizes that the linked members are part of a
 136 coordination structure. This is in contrast with negative coordination (6), which is
 137 realized by means of the negator *āsi* and the additive morpheme *=āpa* that attaches to
 138 the last coordinated member:

139
 140 SIRAYAIC (Austronesian > East Formosan, Taiwan PC; Adelaar 2011: 155)
 141 (6) *Tawrahey=kamu tu āsi ka-vana-n ki na sa-sulat*
 142 err=2PL.NOM LOC NEG V1-know-UO DF PART RDP-write
 143 *ki lix=āpa ki Alid*
 144 DF power=ADD DF God
 145 ‘You are mistaken, knowing neither the scriptures nor the power of God’
 146

147 Thus in (6), unlike in (3c), negation is marked only once (*āsi*) and the same is true for
 148 coordination (*=āpa*). No emphasis seems to be placed on the fact that the coordinands
 149 *ki na sasulat* ‘the scriptures’ and *ki lixāpa ki Alid* ‘the power of God’ belong to a
 150 coordination structure and are considered separately. The fact that Sirayaic has two
 151 distinct means to express negative coordination (6) and emphatic negative coordination
 152 (3c) indicates that these are different functions.⁵

⁵ The authenticity of *k’āsiahaw* as a marker of ENC is, however, doubtful: Sirayaic is attested in Dutch translations and, therefore, a loan translation of Dutch *noch* ‘neither, nor’ cannot be excluded. Details

153 The examples discussed so far thus illustrate that at least some languages tend to
 154 have different formal means in order to encode negative coordination (4, 5a-b, 6) as
 155 opposed to emphatic negative coordination (3a-c), and that this contrast might imply a
 156 difference in meaning. Instances (3a-c) also show that the comparative concept of ENC
 157 materializes variably: depending on the language, this can occur via emphatic negators
 158 (Hup), standard negators (Basque) and negative conjunctions (Sirayaic). In short, then,
 159 Hup, Basque and Sirayaic present dedicated, language-specific means used to express
 160 ENC.

161 However, there is an issue with this view: having different formal means to express
 162 negative coordination vs. emphatic negative coordination is not enough to determine
 163 that a given language has dedicated, language-specific ENC markers. Indonesian, for
 164 example, encodes the latter kind of clause linkage by means of the conjunctive
 165 coordinators *baik* ‘both’ and *maupun* ‘and’ under the scope of a single negator *tidak*
 166 (7a). By contrast, negative coordination is indicated by means of a single coordinator
 167 *dan* ‘and’ under the scope of negator *tidak* (7b):

168

169 INDONESIAN (Austronesian > Malayo-Polynesian, Indonesia; Sneddon 1996: 339, 348)

- 170 (7) a. *Baik ke-pandaian maupun kecantikan tidak berguna*
 171 both ability and beauty NEG useful
 172 *untuk mencapai kebahagiaan*
 173 for achieve happiness
 174 ‘Neither ability nor beauty is useful for achieving happiness (lit. both
 175 ability and beauty are not useful for achieving happiness)’
- 176 b. *Majikan tidak bisa sembarangan memperkerjakan*
 177 employer NEG can at.random engage
 178 *dan membayar pembantu-nya*
 179 and pay servant-3
 180 ‘An employer cannot employ and pay his servants just as he pleases’

181

182 The clause-linkage markers in (7a) and (7b) differ in shape, but (7a) is formally no
 183 different from instances of correlative conjunctive coordination (8) other than the fact

concerning the relationship between language contact and ENC are discussed in Section 4.2 below.

184 that (7a) is negated and (8) is not:

185

186 INDONESIAN (Austronesian > Malayo-Polynesian, Indonesia; Sneddon 1996: 339, 348)

187 (8) *Baik di kota maupun di desa sepak bola digemari*

188 both in town and in village kick ball be.enjoyed

189 *orang*

190 person

191 ‘Soccer is enjoyed both in the town and in the village’

192

193 Therefore, in Indonesian ENC (7a) and negative coordination (7b) are distinctly realized,

194 but ENC equates to a negated version of correlative conjunctive coordination (8).

195 Accordingly, Indonesian cannot be claimed to have a dedicated means for expressing

196 ENC.

197 Many languages align with Indonesian in that they formally distinguish negative

198 coordination from ENC yet do not have a dedicated means for the latter. Very often this

199 clause-linkage function ensues via disjunctive coordinators under the scope of negation,

200 as in Warlpiri (9a), and conjunctive coordinators under the scope of negation as in Ewe

201 (9b):⁶

202

203 WARLPIRI (Pama-Nyungan > Desert Nyungic, Australia; Bowler 2014: 139)

204 (9) a. *Cecilia manu Gloria kula=pala yanu*

205 Cecilia either Gloria NEG=3DU.SBJV go.PST

206 *Lajamanu-kurra*

207 Lajamanu-ALL

208 ‘Neither Cecilia nor Gloria went to Lajamanu (lit. either Cecilia or Gloria

209 did not go to Lajamanu)’

210 EWE (Atlantic-Congo > Volta-Congo, Ghana, Togo; Rongier 2004: 176–177)

211 b. *Tɔ-nyè kple nɔ-nyè me-lè dɔ wɔm ò*

212 father-1SG and mother-1SG NEG-PROG work make NEG

213 ‘Neither my father nor my mother are working (lit. my father and my

⁶ Conjunction and disjunction are here regarded as subtypes of coordination. Conjunction implies that the proposition holds true only if all parts of the proposition, i.e., all conjuncts are true (Bussmann 1998: 231). In turn, disjunction indicates that either one, or more than one, or all of the parts of the proposition are true (Brown & Miller 2013: 137).

214 mother are not working)’

215

216 The Warlpiri example (9a), where the disjunctive coordinator *manu* ‘either, or’
217 combines with negative marking (*kula=*) on the verb, literally translates as ‘either
218 Cecilia or Gloria did not go to Lajamanu’. Thus in Warlpiri the formal means to express
219 ENC corresponds to negated disjunctive coordination. Ewe is a similar case: in example
220 (9b) the coordinator *kple* co-occurs with the complex negative marker *me- ... ò*. A
221 verbatim translation of the sentence would be ‘my father and my mother are not
222 working’, i.e., in Ewe ENC is equal to negated conjunctive coordination.

223 In short, then, languages such as Indonesian, Warlpiri and Ewe do not make a formal
224 distinction between negative coordination and ENC. For the purpose of delimitation, the
225 focus of this study is on languages (i) which make a formal distinction between negative
226 coordination vs. ENC and (ii) where there is no formal overlap between ENC and other
227 clause-linkage functions, such as conjunctive and disjunctive coordination. As stated
228 above, Hup, Basque and Sirayaic fall into this group, whereas Indonesian, Warlpiri and
229 Ewe do not. In accordance with this conception of ENC, the expression ‘dedicated,
230 language-specific ENC markers’ will be used henceforth to refer to languages where
231 correlative conjunctions encode, *when co-occurring together*, no more and no fewer of
232 the functions mentioned in (2) above. Whether ENC markers coexpress other functions
233 (such as standard negation or scalar focus) *when occurring individually* is a different
234 matter whose implications are discussed in Section 5.

235 A second issue when defining ENC concerns the scope of negators and coordinators
236 involved in clause linkage. The focus of this investigation is on situations where
237 emphasis, negation and coordination are all expressed in the same clause as at least one
238 of the coordinands. Consequently, sentences such as (10) from Tariana, in which
239 negation is marked externally to the clause, have been excluded from the analysis:

240

241 TARIANA (Arawakan > Japura-Colombia, Brazil, Colombia; Aikhenvald 2003: 403)

242 (10) *Di-ra,* *di-kama,* *na-na* *di-kwisa*
243 3SG.NF-drink 3SG.NF-be.drunk 3PL-OBJ 3SG.NF-scold
244 *di-we* *ma-ni-kade-na*
245 3SG.NF-become.CAUS NEG-do-NEG-REMP.VIS

246 'He did not drink, get drunk, start scolding them (lit. he drank, got drunk,
247 started scolding people – this was NOT the case)'

248

249 In (10) the forms *dira* 'he drank', *dikama* 'he got drunk' and *nana dikwisa diwe* 'he
250 started scolding people' are all under the scope of a single complex negator *ma-* ...
251 *-kade*, which is marked on the final verb. Nonetheless, it is doubtful whether these
252 forms are part of the same clause, since they are separated by pauses—which are
253 indicated by commas—and because they can take constituents of their own, such as
254 *nana* '3PL.OBJ' (Aikhenvald 2003: 403). Accordingly, examples like (10) have been left
255 out of the discussion.

256 Moreover, it should be noted that the emphasis on functionally equivalent members
257 being linked together at the same level of structure (cf. (2i) above) excludes other kinds
258 of clause linkage, including subordination and cosubordination, from the definition of
259 ENC. Accordingly, sentences such as (11), where a matrix negative verb ('*ikai*') takes a
260 clausal complement (*ke alu 'a siale*), i.e., where a so-called 'higher negative verb'
261 (Payne 1985a: 207–208, Miestamo 2005: 84–85) is involved, fall outside the scope of
262 this study:

263

264 TONGA (TONGA ISLANDS) (Austronesian > Malayo-Polynesian, Tonga Islands;
265 Churchward 1953: 56)

266 (11) *Na'e 'ikai ke alu 'a siale*
267 PST NEG SBJV go ABS Siale
268 'Siale did not go'

269

270 Finally, the form of coordinators may change depending on the illocutionary force of the
271 utterance (Haspelmath 2007: 3-4). For example, in Mandarin Chinese the disjunctive
272 coordinator *háishi* is used in questions (12a) and *huòzhe* in statements (12b):

273

274 MANDARIN CHINESE (Sino-Tibetan > Sinitic, China; Li & Thompson 1981: 654)

275 (12) a. *Nǐ yào wǒ bāng nǐ háishi yào zìjǐ zuò*
276 2SG want 1SG help 2SG or want self do
277 'Do you want me to help you, or do you want to do it yourself?'

278 b. *Wǒmen zài zhèli chī huòzhe chī fàndiàn*
 279 1PL at here eat or eat restaurant
 280 *dōu xíng*
 281 all OK
 282 ‘We can either eat here or eat out’

283

284 For the purpose of delimitation, instances of non-declarative coordination such as (12a)
 285 have been excluded from the analysis.

286 In summary, only those clause-linkage functions that have the properties described in
 287 (2) and which fall inside the scope of this study have been considered to instantiate
 288 dedicated, language-specific ENC markers. More details on the typological variability
 289 concerning the expression of ENC are discussed in Section 4.

290

291 **3. Language sample: description and analysis**

292

293 **3.1 Choice of sampling method**

294

295 Language sampling is an integral part within the methodology of linguistic typology,
 296 which relies on empirical research of cross-linguistic variation. Depending on the
 297 research question(s) of any given typological study, different methods of language
 298 sampling might be appropriate. As noted by Rijkhoff et al. (1993: 171), there are
 299 essentially two ways to approach this matter: if the aim is to find out potential cross-
 300 linguistic frequencies of features and correlations between them, the languages of the
 301 sample should be genetically, areally and typologically independent of each other. The
 302 reason for this is that one can make statistically valid generalizations only on the basis
 303 of independent units, in this case languages. Sampling that relies on languages
 304 independent of one another has been labeled probability sampling (Rijkhoff et al. 1993:
 305 171, Miestamo et al. 2016: 233). Methods used to construct probability samples have
 306 been put forward in a number of contributions to the topic (Perkins 1989, Dryer 1989).

307 By contrast, studies aiming to determine all possible realizations of a specific
 308 grammatical feature should strive to represent as much variety as possible. Accordingly,
 309 for this kind of research representativeness is more important than the genetic, areal and

310 typological independence of the languages being analyzed. Sampling which aims at
311 representing variety has been labeled variety sampling (Rijkhoff et al. 1993: 171).
312 Furthermore, searching for potential cross-linguistic frequencies of a specific linguistic
313 phenomenon and for possible connections between this phenomenon and other parts of
314 grammar—an aim which favors probability sampling—implies that the feature being
315 studied is fully grasped. However, as mentioned in Section 1, ENC lacks as of yet
316 complete understanding. This means that variety sampling, which relies on
317 representativeness, is better suited for the purposes of this study.

318 So far two approaches to variety sampling have been defined: the so-called Diversity
319 Value (Rijkhoff et al. 1993, Rijkhoff & Bakker 1998, Bakker 2011) and Genus-
320 Macroarea (Miestamo 2005, Miestamo et al. 2016) methods. The first of these relies on
321 genealogical grouping of the languages selected for the sample. Depending on the
322 internal complexity of each genealogical grouping a diversity value is calculated, which
323 is then used to determine how many languages of each genealogical grouping should be
324 included in the sample. More specifically, this internal diversity measure is computed on
325 the basis of the nodes (the points of intersection internal to language phyla) at the
326 intermediate levels between the top node, that is to say, the name of the language
327 phylum, and the terminal nodes at the bottom end of the genetic language tree, i.e., the
328 individual languages within the phylum. The internal diversity measure or diversity
329 value (DV) of each phylum is calculated based on the width and depth of a genetic
330 language tree. The width of a genetic language tree is equal to the number of nodes at
331 any given level, and its depth is equivalent to the number of nodes between the name of
332 the language phylum and the individual languages. These two values (width and depth)
333 of each phylum are then converted into the DV by calculating the average number of
334 nodes per intermediate level.

335 When computing the DV, high level nodes are given more importance than low level
336 ones, since the distinguishing power (the genetic distance between languages) is greater
337 in higher than in lower nodes. As an illustrative example, the Eskimo-Aleut phylum is
338 defined by Grimes (1997) as consisting of four nodes in depth (for instance, Naukan
339 Yupik (1) < Siberian Yupik (2) < Yupik (3) < Eskimo (4)) and five in width at the fourth
340 level of depth (Naukan Yupik (1), Sirenik Yupik (2), Alaskan Yupik (3), Inuit (4) and
341 Aleut (5)) (Rijkhoff & Bakker 1998: 269–270). The second level of depth in this

342 phylum, which consists of three nodes in width (Yupik (1), Inuit (2) and Aleut (3)),
343 should thus be given more importance than the fourth level of depth because the genetic
344 distance between Yupik and Aleut is greater than, for instance, Naukan Yupik and
345 Siberian Yupik. The DV of each language phylum is then used to calculate the number of
346 languages of that phylum that should be included in a sample, subject to the sample's
347 size. Moreover, each genealogical grouping should be represented by at least one
348 language. Pidgins and creoles taken together count as one genealogical grouping,
349 whereas language isolates each represent one phylum.

350 As opposed to the Diversity Value method for variety sampling, which relies
351 exclusively on phyla, in the Genus-Macroarea method two levels of stratification are
352 created: a genealogical one at the genus level and an areal one at the macroarea level.
353 Both concepts "genus" and "macroarea" are adopted from Dryer (1989, 1992). The first
354 of these refers to a level of genealogical classification with a maximum time depth of
355 3,500-4,000 years. Genera are conceived as a comparable entity across the world's
356 languages, since this is in most cases the maximal level of grouping whose genealogical
357 relationship is uncontroversial (Miestamo et al. 2016: 239). Examples of genera are,
358 within Indo-European, Iranian, Slavic and Germanic, as well as Berber, Chadic and
359 Semitic within Afro-Asiatic. Similarly to the Diversity Value method, language isolates
360 each constitute one genus, whereas contact languages are usually not considered
361 (Miestamo et al. 2016: 250).

362 Macroareas refer, in turn, to continent-level linguistic areas that are independent of
363 each other and which comprise typologically relatively similar languages. This
364 resemblance is due either to contact or remote genealogical affinity. Dryer (1989: 268–
365 269, 1992: 133–135) distinguishes six macroareas: Africa, Eurasia, Australia-New
366 Guinea, North America, South America and Southeast Asia-Oceania. Despite the fact
367 that macroareas mostly follow geographical divisions, sometimes the boundaries are
368 drawn by genealogical groupings. These exceptions are due to the fact that a given
369 genealogical group can be spread over two continents, while all member languages of
370 that group have typological features that are characteristic of one macroarea. Thus the
371 Chibchan languages of Mesoamerica fall within South America, and the Afro-Asiatic
372 languages of Asia within Africa. In other cases one genealogical grouping is split into
373 two macroareas: the Munda genus is included in Eurasia due to its typological

374 similarities with other Eurasian languages, whereas the rest of the Austro-Asiatic family
375 is included in the Southeast Asia-Oceania macroarea due to its structural affinity with
376 other languages of that area.⁷

377 Each of the approaches to variety sampling discussed here has its advantages and
378 disadvantages. Thus the Diversity Value method has been criticized for not
379 encompassing any areal stratification (Miestamo et al. 2016: 245). As pointed out by
380 Bakker (2011: 118), however, areal data can be introduced into the method. In fact,
381 many of the genealogical classifications it can be based on —such as Grimes (1997)—
382 already include areal information, e.g., the “Australian”, “East Papuan” and “North
383 Caucasian” phyla. Another point of critique concerns the fact that computation of DVs
384 depends to some extent on the genealogical classification that is chosen. Consequently,
385 the proportion of each phylum in same-size samples varies depending on the sources.
386 Nevertheless, these differences cannot be considered significant, since samples based on
387 different sources as discussed by Rijkhoff et al. (1993) and Rijkhoff & Bakker (1998)
388 have been argued to yield similar results (Croft 2003: 21, though see Miestamo et al.
389 2016: 245). Finally, the genealogical groupings laid out by the Diversity Value method
390 have been deemed unreliable due to their representing different time depths (Miestamo
391 et al. 2016: 246). Despite these downsides, the Diversity Value approach to variety
392 sampling represents the diversity of the world’s genealogical groupings, it is fully
393 explicit and formalized, and it enables reproducible sampling that is comparable across
394 different studies.

395 In turn, the Genus-Macroarea method accounts for both areal and genealogical
396 stratification, it intends to solve the problem of reliability of genealogical classifications
397 by drawing on genera instead of phyla, and it is likewise explicit and formalized.
398 However, as pointed out by Miestamo et al. (2016: 259) the delineation of genera is
399 problematic, since there is no upper limit to the number of languages per genus. Also,
400 the method based on genera has been argued to overrepresent some phyla, whereas it
401 underrepresents others (Rijkhoff & Bakker 1998: 300–301). Thus one of the main
402 purported advantages of the Genus-Macroarea over the Diversity Value method —the
403 reliability of genera in comparison to phyla— is undermined by the difficulty to define

7 For details on how the number of languages to be included in a sample is computed based on the concepts “genus” and “macroarea” of the Genus-Macroarea method, see Miestamo et al. (2016: 251–258).

404 the limits of this kind of genealogical grouping. Moreover, drawing on genera excludes
405 contact languages due to their minor time depth. Consequently, part of the world's
406 linguistic diversity is disregarded. Finally, computer simulations have shown that the
407 Diversity Value and Genus-Macroarea methods perform similarly in capturing the
408 variety of the world's languages (Miestamo et al. 2016: 260–270). Due to these reasons
409 the Diversity Value approach to variety sampling has been chosen for this study.

410

411 **3.2 Sample generation and bias resolution**

412

413 Concerning the steps taken in the generation of the sample, first of all the ready-made
414 calculations in Rijkhoff & Bakker (1998: 306–310) were adopted as a basis. As
415 mentioned above, these authors do not offer a preconceived sample, but they do give
416 numbers of languages to be included in the different phyla. Second, the individual
417 languages of each phylum were selected according to Rijkhoff & Bakker's (1998)
418 calculation for a 250-language sample. This calculation is based on a classification of
419 the world's languages by *Ethnologue* (Grimes 1997).⁸ In so doing, the availability of
420 grammatical descriptions and dictionaries had to be given priority, since the relevant
421 data are hardly ever discussed outside these kinds of texts.

422 Therefore, at the lower levels of genealogical grouping, a balancing act was
423 performed between adhering to the methods of selecting individual languages according
424 to Rijkhoff & Bakker's (1998) method, on the one hand, and the availability of
425 comprehensive descriptions for the languages in question, on the other. For example,
426 these authors argue (Rijkhoff & Bakker 1998: 308) that a 250-language sample should
427 include one Katukinan language. So far the Katukinan phylum has been identified as
428 having two members: Katawixi and Katukína-Kanamari (Adelaar 2007). This implies
429 that, theoretically, both languages are suitable candidates to be included in the sample.
430 However, only Katukína-Kanamari has been thoroughly described (Groth 1985, Dos
431 Anjos 2011, Ishy de Magalhães 2018). Therefore, by necessity Katukína-Kanamari was
432 selected over Katawixi.

433 Another example of involves the Geelvink Bay phylum, which according to Rijkhoff

⁸ The reader should notice that a much more recent version of *Ethnologue* is currently available (Eberhard et al. (eds.) 2020). In keeping with Rijkhoff & Bakker's (1998) sampling method, however, the classification by Grimes (1997) was implemented here.

434 & Bakker’s (1998: 308) proposal should include two representatives. So far ten
 435 languages have been identified as belonging to this phylum: Barapasi, Bauzi, Burate,
 436 Demisa, Kofei, Nisa-Anasi, Sauri, Tefaro, Tunggare and Woria (Voorhoeve 1975).
 437 Therefore, in principle ten candidates are available to be included in the sample.
 438 However, to date only one of these has been thoroughly described, namely Bauzi
 439 (Briley 1997). Therefore, Bauzi was selected as the sole representative of the Geelvink
 440 Bay phylum. The second slot available to this phylum had to be assigned to another
 441 phylum, preferably to the closest one geographically, in this case Hatam, a West Papuan
 442 language. Consequently, the Geelvink Bay phylum is underrepresented by one language
 443 at the expense of the West Papuan phylum, which is overrepresented by one language
 444 according to Rijkhoff & Bakker’s (1998) standards for a Grimes (1997)-based 250-
 445 language sample. This was, at any rate, the only case in which this procedure was
 446 followed. All in all, one might argue that the variety sample created for this study is
 447 similar to what Miestamo et al. (2016: 250) label a “core sample”.

448 Giving preference to the availability of grammatical descriptions and dictionaries
 449 runs the risk of yielding a bibliographically biased sample. Moreover, since some
 450 macroareas are better studied than others, bibliographic biases tend to introduce an areal
 451 bias as well (Miestamo et al. 2016: 251). Accordingly, a potential areal and
 452 bibliographic bias can be mitigated by underrepresenting well-studied areas or,
 453 alternatively, by overrepresenting poorly investigated areas. According to Hammarström
 454 (2009), Eurasian languages tend to be overrepresented in some typological studies at the
 455 expense of Papuan and South American languages. In order to cancel out bias effects,
 456 the latter two macroareas have been here overrepresented by 5 languages each at the
 457 expense of the Eurasian area, which has been underrepresented by 10 languages. The
 458 number of languages and the proportion of each macroarea in the sample are illustrated
 459 by Table 1.

460

461 **Table 1.** Number and proportion of sample languages by macroarea

Macroarea	Languages	Proportion of sample languages	Families	Proportion of sample families
Africa	42	16.8%	21	13.8%
Australia-Papua New Guinea	47	18.8%	37	24.3%
Eurasia	31	12.4%	16	10.5%

North America- Mesoamerica	41	16.4%	29	19.1%
South America	47	18.8%	39	25.7%
Southeast Asia- Oceania	42	16.8%	10	6.6%
Total	250	100%	152	100%

462

463 A second measure against a bibliographically biased sample involves random selection
464 of subfamilies. When applying the Diversity Value method to sampling not every
465 subfamily can be represented by a language due to the fact that the number of available
466 languages is lower than the number of subfamilies (Rijkhoff & Bakker 1998: 276–277).
467 This can lead researchers to select only well-described languages. Therefore, randomly
468 distributing languages over subfamilies can avoid creating a bibliographic bias. One
469 way to achieve chance distribution is to assign a number to each subfamily and let a
470 computer program generate random numbers, which is the approach taken here.

471 For instance, a Grimes (1997)-based 250-language sample created on the basis of the
472 Diversity Value method requires the Uralic phylum to be represented by two languages.
473 However, this phylum is subdivided by Grimes (1997) into seven branches (Finnic,
474 Khantyc, Mari, Mordvin, Permian, Sami and Samoyed), which means that five
475 subdivisions must be left out. By means of random number assignment the Khantyc
476 and Permian subfamilies were selected. In the same vein, one language each was
477 assigned to Central Salish and Interior Salish among the three Salishan subbranches
478 (Central, Interior and Tsamosa).

479 Despite these measures and concerning the subject matter, grammatical descriptions
480 often fail to specify restrictions on ENC, which are believed to vary from language to
481 language. This limitation should be borne in mind when better than at the time of
482 evaluating the validity of the results, which are presented in Section 4.

483

484 **3.3 Sample outline and analysis**

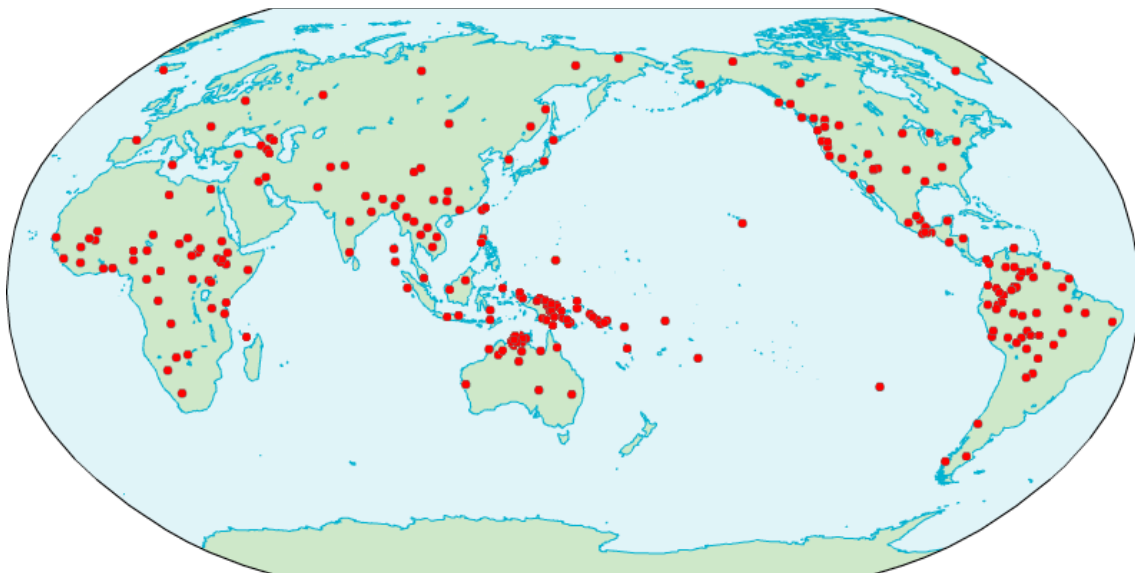
485

486 As mentioned above, the aim of this study is to determine all possible realizations of
487 ENC by drawing on the Diversity Value approach to variety sampling (Rijkhoff &
488 Bakker 1998). Accordingly, this study is based on a sample of 250 languages, which
489 represent a total of 96 (out of 124, 77.4%) linguistic phyla. Because the data have been

490 drawn from dictionaries and grammatical descriptions, no relevant information could be
491 found for Amto-Musan, Andoque, Arutani, Burmeso, Busa, Caddoan, Cayubaba,
492 Itonama, Karkar-Yuri, Kibiri, Left May, Máku, Mascoian, Pankararú, Pauwi, Puelche,
493 Quileute, Salivan, Ticuna, Tol, Tonkawa, Trumai, Tuxá, Uru-Chipaya, Vilela,
494 Warembori, Yámana and Yale, as these are mostly fragmentarily attested and poorly
495 documented languages whose descriptions do not offer sufficient information
496 concerning the topic at hand. The list of sample languages is included in the Appendix.⁹
497 Figure 1 shows the geographic distribution of the languages in the sample:

498

499 **Figure 1.** Geographic distribution of the languages under study



501 Even though the main goal of this paper is qualitative rather than quantitative, the
502 sample data nonetheless allow for statistical analysis. In what remains of this section an
503 explanation is provided of the statistical analysis that has been performed on the raw
504 data.

505 One of the key factors when statistically analyzing variables is their type or nature,
506 namely, variables can exhibit ‘continuous’ or ‘discrete’ behavior; the latter are also
507 frequently referred to as ‘nominal’ or ‘categorical’ parameters in statistical literature. As
508 a result of this distinction, there are generally three types of parameter-pairs that can be
509 addressed when, for instance, trying to compute variable correlation: those between

⁹ The following anonymous link provides access to a list of references to the Appendix: https://zenodo.org/record/4636282#.YFxfLXko_IU. The names of the languages in the sample and information on their genetic affiliation have been adopted from *Glottolog* (Hammarström et al. 2020).

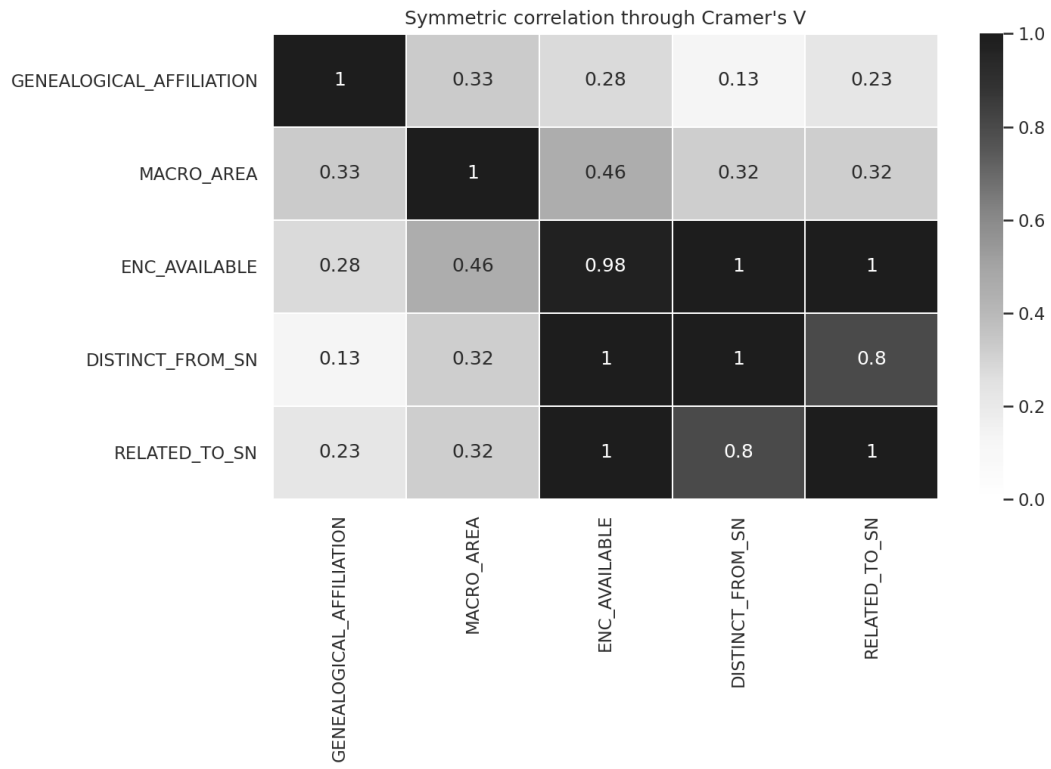
510 categorical variables, those between discrete variables and those with mixed types. It is
511 essential to point out that each of the contexts requires a tailored set of statistical
512 techniques. For measuring the correlation of continuous variables, Pearson's correlation
513 (Pearson 1895) is typically used, and for mixed types, on the other hand, logistic
514 regression (Wright 1895) or an adaptation of Pearson's Correlation by the name of point
515 biserial correlation (Tate 1954) can be applied.

516 In what concerns this research, however, all of the considered variables (see further
517 below) exhibit a nominal behavior; therefore, correlation has been quantified by means
518 of Cramer's V (Cramer 1946), which is a correlation technique based on Pearson's chi-
519 squared statistic and also sometimes referred to as Cramer's phi. In addition, given the
520 symmetric nature of Cramer's V and the limited size of the database, Theil's U (also
521 known as the uncertainty coefficient) (Theil 1966) has been computed over the set of
522 considered parameters in order to get a clearer and non-symmetric view of the
523 parameter correlations without 'losing' any instances to symmetry.

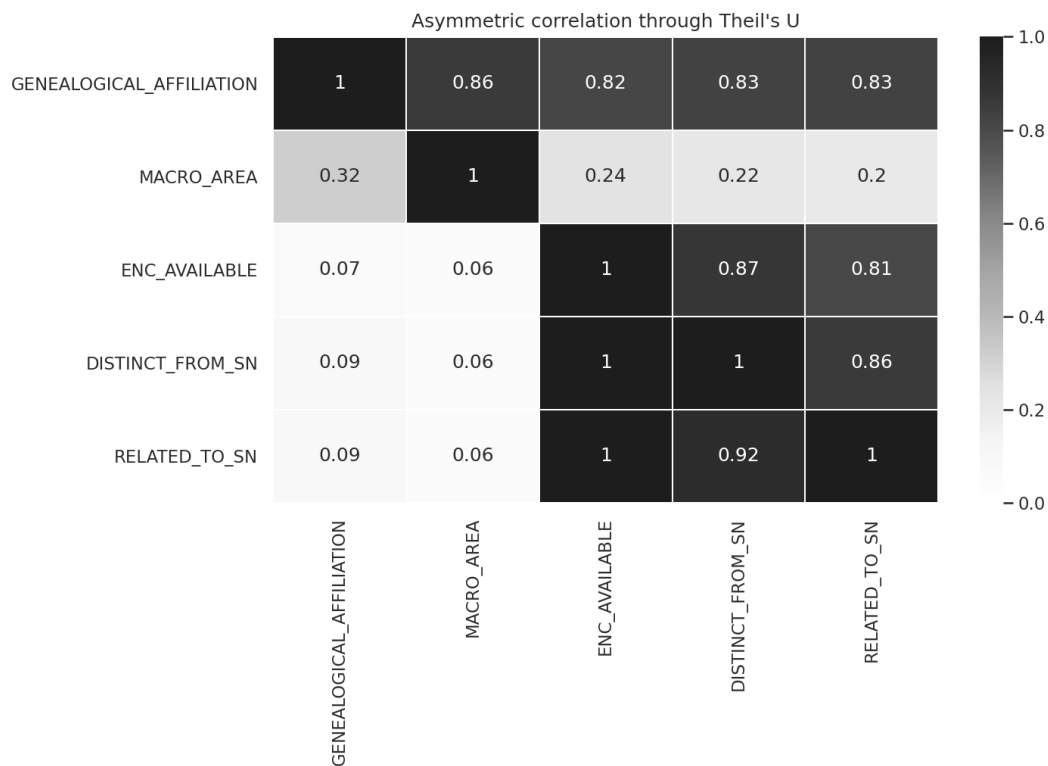
524 It is also worth mentioning that there are two major ways in which correlations
525 between discrete variables can be calculated, by so-called distance metrics such as the
526 Manhattan and the Canberra distances (Black 2006) and through contingency table
527 analytics such as the ones implemented in Cramer's V and Theil's U. One of the biggest
528 drawbacks of distance metric techniques is their strong sensitivity to input scale
529 adaptations, making it hard to correctly compare correlation factors across several
530 iterations of corpus extensions. In addition, distance metrics are said not to be easily
531 comparable when correlating variable pairs which can take different numbers of
532 categories. Consequently, correlations have been calculated here by means of
533 contingency tables. Figure 2 represents the results of Cramer's V analysis of the corpus.
534 Figure 3 represents the results of correlation analysis based on Theil's U or uncertainty
535 coefficient.

536

537 **Figure 2.** Results of Cramer's V correlation analysis on the dependent variables in question



539 **Figure 3.** Results of Theil's U correlation analysis on the dependent variables in question



541

542 Cramer's V is said to be symmetric because the correlation values represented in the Y-
 543 axis and the X-axis in Figure 2 are identical. In turn, this does not apply to Theil's U,
 544 which is why it is referred to as asymmetric. Symmetric and asymmetric correlation are
 545 two different perspectives on the same data. Accordingly, both methods of statistical
 546 analysis can be seen as complementing each other. The degree of correlation is
 547 represented on a scale from 0 (= no correlation whatsoever) to 1 (= absolute
 548 correlation). The intersections between dependent variables are set by default at 1,
 549 except for the variable ENC_AVAILABLE in Cramer's V, which does not reach this
 550 number due to error.

551 The abbreviations in Figures 2 and 3 are short terms for the dependent variables
 552 under investigation. Many of them have been discretized in keeping with their subtypes.
 553 Accordingly, the following are their denotations:

554

555 - GENEALOGICAL_AFFILIATION = the genealogical affiliation of each sample language,
 556 i.e., any one of the 152 families listed in the Appendix;

557 - MACRO_AREA = the macroarea each of the sample languages falls into, i.e., Africa (1),

558 Australia-Papua New Guinea (2), Eurasia (3), North America-Mesoamerica (4), South
559 America (5) and Southeast Asia-Oceania (6);

560 - ENC_AVAILABLE = whether or not dedicated, language-specific ENC markers are
561 available to each language, according to the definition of ENC laid out in Section 2;

562 - DISTINCT_FROM_SN = whether or not individual ENC markers are formally distinct
563 from the language's marker of standard negation;

564 - RELATED_TO_SN = whether or not individual ENC markers are etymologically related
565 to the language's marker of standard negation, to the extent that this can be determined.

566

567 The degree of correlation between these dependent variables is discussed in Section 4.2
568 below.¹⁰

569

570 **4. Typological overview of emphatic negative coordination**

571

572 **4.1. Cross-linguistic tendencies of emphatic negative coordination**

573

574 According to the definition laid out in Section 2, 37/250 (14.8%) of the languages in the
575 sample have dedicated, language-specific ENC markers. Among these 37 languages four
576 types of ENC markers stand out. First of all, ENC can be expressed by means of
577 bisyndetic/correlative markers, without the addition or support of any other element.
578 This type is here referred to as α_1 , and it is present 20/37 (54.1%) of the sample
579 languages with dedicated, language-specific ENC markers, which makes it the most
580 frequent ENC type.

581 In Brahui for example, the functionally equivalent members *bahā kēk* 'he sells
582 (them) for money' and *xudānā pēnaṭ ētik* 'he gives them for God's sake' are coordinated
583 in (13a) by the correlative elements *na ... na ...*, which also convey negation and
584 emphasis. This is different from negative coordination, which is encoded by the
585 negative suffixes *-ta, -pa* (13b).

586

587 BRAHUI (Dravidian > North Dravidian, Afghanistan, Pakistan; Andronov 2001: 108,
588 Barjasteh 2018: 105)

10 The raw data and other details concerning statistical analysis can be accessed through the following link: <https://github.com/IkerSalaberri/catcorrel>.

- 589 (13) a. *Asē ambal-as palēze das-ē maga*
590 one fellow-INDF plantation sow-3SG.PST but
591 *kūṭiḡ-ō-galav-āte na bah-ā kēk*
592 watermelon-and-melon-PL ENCM money-GEN sell.3SG.PRS
593 *na xudā-nā pēn-aṭ ētik*
594 ENCM God-GEN sake-for give.3SG.PRS
595 ‘One fellow has sowed a plantation of melons and watermelons, but
596 he neither sells them for money nor gives them for God’s sake’
- 597 b. *Pābō-ta ki bādšā jōṛ aff, pēš*
598 well-NEG that king ready be outdoors
599 *tammi-pa-k*
600 go.out-NEG-3SG.PRS
601 ‘The king is not well and cannot go out of doors’
602

603 In (13b) there is no overt conjunction coordinating the clauses *Pābōta ki bādšā jōṛ aff*
604 ‘the king is not well’ and *pēš tammipak* ‘he cannot go outdoors’, i.e., these are
605 juxtaposed clauses, and this is the common means of conjunction in Brahui. According
606 to Andronov (1980: 91), conjunctions are absent as a separate grammatical category
607 from early attestations of Dravidian languages, and their use is largely optional in
608 present-day Brahui. There are no attestations in the sources of *na ... na ...* being
609 accompanied by conjunctive coordinators such as *ō* and *a* ‘and’. Therefore, it is quite
610 safe to assume that the correlative ENC markers are used individually as a rule.

611 A similar case can be observed in Huangascar-Topara-Yauyos Quechua, where
612 negative coordination is encoded by the conjunctive enclitic *-pis*, which adheres to each
613 one of the coordinands, in combination with negative marking on the verb by means of
614 negator *-chu* (14a). As opposed to this, there is ENC, which ensues via the correlative
615 elements *ni ... ni ...* (14b):

- 616
617 HUANGASCAR-TOPARA-YAUYOS QUECHUA (Quechuan > Quechua I, Peru; Shimelman
618 2017: 108, 285)
- 619 (14) a. *Ishpa-ni-pis-chu puqu-chi-ni-pis-chu*
620 urinate-1-ADD-NEG ferment-CAUS-1-ADD-NEG

621 'I neither urinate nor ferment (urine)'

622 b. *Ni puñu-y ni miku-y*

623 ENCM sleep-INF ENCM eat-INF

624 'Neither sleeping nor eating'

625

626 In (14a) the functionally equivalent coordinands *ishpa-* 'urinate' and *puqu-* 'ferment'
627 each take one conjunctive enclitic *-pis* and one negative marker *-chu*. By contrast, in
628 (14b) negation, coordination and emphasis are all expressed by the ENC markers without
629 the need of additional conjunctions or negative markers being attached to the
630 coordinands. Therefore, as in Brahui, in Huangascar-Topara-Yauyos Quechua the ENC
631 markers *ni ... ni ...* occur individually.

632 In contrast to type α_1 , in other languages dedicated, language-specific ENC markers
633 cannot occur on their own, but must be accompanied by additional elements. In some
634 languages ENC markers must be accompanied by the standard negator. This type is here
635 referred to as α_2 , and it is present in 5/37 (13.5%) of the sample languages with
636 dedicated, language-specific ENC markers.

637 An example of a language with type- α_2 ENC is Southern Yukaghir. In this language
638 standard negation ensues via addition of the prefix *el-* to the finite verb (15a). Apart
639 from the standard negative marker, there is also another prefix *n'e-*, which is generally
640 used to create negative quantifiers and adverbials (Maslova 2003: 494–495).
641 Furthermore, when used with nouns in prenominal position *n'e-* has a 'not even'
642 meaning (15b). If doubled, *n'e- ... n'e- ...* expresses ENC (15c):

643

644 SOUTHERN YUKAGHIR (Yukaghir > Kolymic, Russia; Maslova 2003: 492, 495–496)

645 (15) a. *Tabun-gele tintan towke el-lej*

646 that-ACC that dog NEG-eat.3SG

647 'That dog did not eat that'

648 b. *Met ahurpe-l juø-de n'e õži*

649 1SG.POSS suffer-ANR see-SS.ITER ENCM water

650 *el-kes 'ī-jemet*

651 NEG-bring-INTR.2PL

652 'You saw how I suffered, you did not even bring me some water'

653 c. *N'e touke-pul el-ojī-ŋi n'e tudel*
 654 ENCM dog-PL NEG-bark-3PL.INTR ENCM 3SG.M
 655 'Neither dogs barked, not it (the bear)'

656

657 Crucially, the correlative ENC marker *n'e- ... n'e- ...*, which in (15c) coordinates the
 658 functionally equivalent members *toukepul* 'dogs' and *tudel* '3SG.M', must co-occur with
 659 the standard negative marker *el-* on the verb. Therefore, Southern Yukaghir contrasts
 660 with languages like Brahui and Huangascar-Topara-Yauyos Quechua, which do not
 661 require additional negative marking on the verb in order to express ENC. This difference
 662 motivates a distinction between types α_1 and α_2 .¹¹

663 Another language with type- α_2 ENC is Huastec. Here there is a negative particle
 664 *ʔiba:*, which is used, among others, to negate some transitive verb phrases (16a). Apart
 665 from *ʔiba:*, there is the negator *ni*, which negates interrogative and locative pronouns as
 666 well as adverbs (16b). When used bisyndetically, *ni ... ni ...* indicates ENC (16c).

667

668 HUASTEC (Mayan > Huastecan Mayan, Mexico; Edmonson 1988: 544–545)

669 (16) a. *ʔi:b ʔin neʔeε*
 670 NEG 1SG go
 671 'I'm not going'

672 b. *Ni hant'ini?*
 673 ENCM how
 674 'In no way (lit. not how)'

675 c. *ʔiba: tin εaʔbiya:mal ni ʔu mi:m ni ʔu*
 676 NEG 3PL visit.PFV ENCM A.1 mother ENCM A.1
 677 *ʔebčal*
 678 sister
 679 'Neither my mother nor my sister has visited me'

680

681 In (16c) the correlative ENC marker *ni ... ni ...*, which coordinates the functionally
 682 equivalent elements *ʔu mi:m* 'my mother' and *ʔu ʔebčal* 'my sister' must be

¹¹ The contrast between type- α_1 and type- α_2 ENC constructions may be seen as one between languages with negative concord and without. For details see, among others, Jeretič (2018), Čéplö & Lucas (2020), Van der Auwera (2021) and Van der Auwera et al. (2021).

684 accompanied by negator *ɔ̃iba*:. Therefore, Huastec is like Southern Yukaghir and unlike
 685 Brahui and Huangascar-Topara-Yauyos Quechua in the sense that correlative ENC
 686 markers must be accompanied by negative marking on the verb.

687 As opposed to languages with type- α 1 and type- α 2 ENC, there are other languages in
 688 which ENC markers are accompanied by coordinators, which can be either conjunctive
 689 or disjunctive. These ENC constructions are here labeled as type α 3, and they exist in
 690 3/37 (8.1%) of the sample languages with dedicated, language-specific ENC markers. In
 691 Turkish, for instance, negative coordination is expressed by means of the standard
 692 negative suffix *-mV* attaching to the verb and negated coordinands being coordinated
 693 with the conjunctive coordinator *de* (17a). By contrast, ENC is conveyed via the
 694 correlative, clause-initial ENC marker *ne ... ne ...*, which can be optionally reinforced
 695 with the same conjunctive coordinator *de* (17b).

696

697 TURKISH (Turkic > Common Turkic, Turkey; Göksel & Kerslake 2005: 272–273,
 698 Kornfilt 2006: 111)

- 699 (17) a. *Hem onları fazla meşgul et-me-miş ol-ur-uz*
 700 also 3PL much occupied make-NEG-PFV be-AOR-1PL
 701 *hem de ziyaret-ler-in-e git-me-miş ol-ma-yız*
 702 also and visit-PL-NMLZ-DAT go-NEG-PFV be-NEG-1PL
 703 ‘We won’t have taken up too much of their time, and on the other hand
 704 we won’t have neglected to visit them’
- 705 b. *Ne Hasan iş-e git-ti, ne (de) Ali*
 706 ENCM Hasan work-DAT go-PST ENCM and Ali
 707 *çarşı-ya çık-tı*
 708 market-DAT go.out-PST
 709 ‘Neither did Hasan go to work nor did Ali go shopping’

710

711 In (17a) as well as in (17b) functionally equivalent finite clauses are coordinated, yet the
 712 conjunctive coordinator *de* is required in (17a), whereas it is only optional in (17b). This
 713 suggests, on the one hand, that *ne ... ne ...* is enough on its own as an ENC construction.
 714 On the other hand, the fact that *de* can co-occur with *ne ... ne ...* suggests that ENC
 715 constructions tend to be reinforced with additional elements. More details on the

716 tendency for ENC constructions to be reinforced are discussed further below as well as in
 717 Section 5.

718 Another language with type- α_3 ENC is Sinhala. In this language standard negation is
 719 expressed by a negative marker *nææ*, which occurs on its own in clause-final position
 720 (18a). In addition, when used correlatively and in a position following each of the
 721 coordinands, *nææ* indicates ENC, as in (18b).

722

723 SINHALA (Indo-European > Indo-Iranian, Sri Lanka; Chandralal 2010: 186, 201)

724 (18) a. *Kolambə rassaawa-ka-tə giyot aye gamə-tə*
 725 Colombo job-INDF-DAT go.COND again village-DAT
 726 *e-nn-e nææ*
 727 come-NPT-FOC NEG

728 ‘If you go to a job in Colombo, you’ll not return to your native village’

729 b. *Balla piduru ka-nne-t nææ gonaatə*
 730 dog hay eat-NPT.FOC-CNJ ENCM bull-DAT
 731 *ka-nnə de-nne-t nææ*
 732 eat-INF give-NPT.FOC-CNJ ENCM

733 ‘The dog neither eats hay nor allows the bull to eat it’

734

735 In (18b) each of the correlative ENC markers *nææ ... nææ ...* follows the functionally
 736 equivalent members *balla piduru kannet* ‘the dog eats hay’ and *gonaatə kannə dennet*
 737 ‘(the dog) allows the bull to eat (the hay)’ respectively. However, the correlative
 738 markers are not enough to express ENC: a conjunction *-t* must also be added to each
 739 coordinated verb (Chandralal 2010: 186). Therefore, in Sinhala ENC markers require the
 740 presence of coordinating conjunctions. The difference in comparison to Turkish is that
 741 in Sinhala conjunctions must adhere to each coordinated verb instead of occurring just
 742 once, and that their presence is obligatory.

743 So far ENC constructions have been discussed which consist of correlative markers,
 744 whether on their own (type α_1 , as in Brahui and Huangascar-Topara-Yauyos Quechua),
 745 or accompanied by negative markers (type α_2 , as in Southern Yukaghir and Huastec), or
 746 in combination with conjunctions (type α_3 , as in Turkish and Sinhala). A fourth
 747 possibility is for ENC to be conveyed by a single (non-correlative) ENC marker in

748 combination with the standard negator. This type is here referred to as type β , and it is
 749 extant in 6/37 (16.2%) of the sample languages with dedicated, language-specific ENC
 750 constructions.

751 An example of a language with type- β ENC is Iquito. In this language standard
 752 negation ensues via a negative particle *caa*, which precedes all elements of the utterance
 753 except topics (19a). In addition, *caa* can combine with the emphatic negative
 754 conjunction *nacaaja* in order to convey ENC (19b).

755

756 IQUITO (Zaparoan > Iquito-Arabela, Peru; Lai 2009: 56, 208)

- 757 (19) a. *Iina msaji caa nu=niqui- \emptyset -cura iina icuani*
 758 DEF womanNEG 3SG=see-GNR.PFV-RPST DEF man
 759 ‘That woman did not see that man (yesterday)’
- 760 b. *Ca=p=paji-i- \emptyset amicaáca asaani nacaaja*
 761 NEG=1PL=can-IPFV-EC one.day.away eat.INF ENCM
 762 *mayasiini*
 763 dance.INF
 764 ‘We can neither eat nor dance tomorrow’

765

766 In (19b) the functionally equivalent members *asaani* ‘to eat’ and *mayasiini* ‘to dance’
 767 are coordinated by a single ENC marker *nacaaja* in combination with standard negative
 768 marking (*ca=*) on the verb. Addition of a second ENC marker is not possible. Therefore,
 769 in Iquito ENC markers are not correlative, unlike in languages belonging to types α 1-3.

770 A similar case in point is Kalaallisut. In this language coordination of negative
 771 predicates is expressed through the conjunctive coordinator *aammalu* in combination
 772 with the standard negator *-nngil-* and the presence of a negative verb, in this case an
 773 allomorph of *juminaat* ‘to not be good’ (20a). In turn, ENC ensues via a different
 774 coordinator, namely the clitic *=lu* ‘and’, presence of the standard negator *-nngil-* and, in
 775 this case, *-rani*, a fourth-person singular form of the negative contemporary mood (20b).

776

777 KALAALLISUT (Eskimo-Aleut > Eskimo, Greenland; Fortescue 1984: 124)

- 778 (20) a. *Mattak mama-nngil-aq aammalu*
 779 mattak taste.good-NEG-3SG.IND also

780 *immiaq* *imi-ruminaap-puq*
781 home.made.beer drink-NEG.good-3SG.IND
782 ‘The mattak doesn’t taste good, nor is the home-made beer drinkable’
783 b. *Sila* *ajur-luinna-rani=lu*
784 weather completely.bad-be-4SG.NEG.CONT=and
785 *pitsaa-lluinna-nngil-aq*
786 completely.good-be-NEG-3SG.IND
787 ‘The weather was neither completely bad nor good’
788

789 Stated differently, in Kalaallisut negative coordination (20a) and ENC (20b) have
790 different means of expression. Moreover, in ENC constructions (20b) there is a single
791 negative marker (*-rani*) in combination with standard negative marking (*-nngil-*) on the
792 verb. Addition of a further negative element to the same construction is not possible.
793 Therefore, like in Iquito and unlike in languages belonging to types α 1-3, in Kalaallisut
794 ENC markers are not correlative. The difference in comparison to Iquito is that in
795 Kalaallisut conjunction (*=lu*), negation (*-rani*) and emphasis (co-occurrence of *-nngil-*
796 and *-rani* in the same clause) are expressed separately instead of by a single marker, as
797 is the case of Iquito *nacaaja*.

798 So far languages have been analyzed which display a single strategy to encode ENC,
799 either by means of correlative (types α 1-3) or non-correlative (type β) ENC markers. A
800 final possibility is for the same language to have two or more different strategies. This is
801 the case in 3/37 (8.1%) of the sample languages with dedicated, language-specific ENC
802 constructions. In Jamiltepec Mixtec, for example, ENC of noun phrases is realized via
803 the correlative marker *nī ... nī ...*, which can be optionally accompanied by the
804 conjunctive coordinator *tā* (21a). By contrast, when coordinating clauses *nī ... nī ...* must
805 co-occur with the standard negator *mā* in each of the coordinands, and *tā* may not be
806 added (21b).

807

808 JAMILTEPEC MIXTEC (Otomanguean > Eastern Otomanguean, Mexico; Johnson 1988:
809 81, 127)

810 (21) a. *Nī* *shita* (*tā*) *nī* *ññ*
811 ENCM tortilla and ENCM salt

812 ‘Neither tortillas nor salt’
 813 b. *Nī mā kāchī ra nī mā kusu*
 814 ENCM NEG POT.eat 3SG.M ENCM NEG POT.sleep
 815 *ra*
 816 3SG.M
 817 ‘He will neither eat nor sleep’

818
 819 Therefore, in Jamiltepec Mixtec ENC of noun phrases is realized via type- α 1 markers —
 820 i.e., plain correlative ENC markers, which change into a type- α 3 construction if
 821 coordinator *tā* is added—, whereas ENC of clauses ensues by type- α 2 markers, namely,
 822 correlative ENC markers accompanied by markers of standard negation. This suggests
 823 that a correlation may exist between kind of coordinand and type of ENC marker.

824 In fact, a similar contrast between phrasal and clausal uses of ENC markers can be
 825 observed in languages outside the sample. In Croatian, for example, the ENC
 826 construction *ni ... ni ...* can only be used to coordinate noun phrases (22a), whereas it is
 827 ungrammatical for clauses. In turn, clauses can only be connected by the ENC markers
 828 *niti ... niti ...* (22b), a possibility that is excluded for noun phrases.

829
 830 CROATIAN (Indo-European > Balto-Slavic, Croatia; Van der Auwera et al. 2021)

831 (22) a. *Ni Iris ni Lena nisu išle*
 832 ENCM Iris ENCM Lena NEG.be.PRS.IPFV.3.PL go.PTCP.PST.PL.F
 833 *u kino*
 834 to cinema.ACC.SG

835 ‘Neither Iris nor Lena went to the cinema’
 836 b. *Niti je Iris ispekla kolač,*
 837 ENCM be.PRS.IPFV.3SG Iris bake.PTCP.PST.SG.F cake
 838 *niti je Lena kupila mlijeko*
 839 ENCM be.PRS.IPFV.3SG Lena buy.PTCP.PST.SG.F milk
 840 ‘Iris neither baked a cake nor did Lena buy milk’

841
 842 In (22a) the functionally equivalent noun phrases *Iris* and *Lena* are coordinated by the
 843 type- α 2 ENC markers *ni ... ni ... nisu*, whereas the type- α 1 ENC elements *niti ... niti ...*

844 coordinate the clauses *Iris ispekla kolač* ‘Iris baked a cake’ and *Lena kupila mlijeko*
 845 ‘Lena bought milk’. Therefore, in Croatian the kind of coordinand determines the type
 846 of ENC marker, just like in Jamiltepec Mixtec.

847 However, languages may also have multiple kinds of ENC markers without them
 848 depending on the kind of coordinand at all. In Palula, for example, there is the type- α 1
 849 correlative ENC construction *na ... na ...*, which can coordinate noun phrases such as
 850 *zinaawurá* ‘wild animal’ and *ghrastá* ‘wolf’ (23a) as well as clauses like *se kasíi xaadí*
 851 *daçhéeni* ‘she doesn’t have a care for anyone’s happiness’ and *(se) kasíi marg daçhéeni*
 852 ‘(she) doesn’t have a care for anyone’s sorrow’ (23b).

853

854 PALULA (Indo-European > Indo-Iranian, Pakistan; Liljegren 2016: 349)

- 855 (23) a. *Na zinaawur-á tas the ga asár thíil-i*
 856 ENCM beast-OBL 3SG.ACC to what effect do.PFV-F
 857 *de na ghrast-á thíil-i de*
 858 PST ENCM wolf-OBL do.PFV-F PST
 859 ‘No wolf or any other wild animal had touched him’
- 860 b. *Se na kasíi xaadí daçh-éen-i*
 861 3SG.F.NOM ENCM anyone.GEN happiness look-PRS-F
 862 *na kasíi marg daçh-éen-i*
 863 ENCM anyone.GEN death look-PRS-F
 864 ‘She doesn’t have a care for anyone’s happiness or sorrow’

865

866 At the same time, in Palula there is a different correlative ENC marker, namely type- α 3
 867 *na ta ... (ee) na ba ...*. This element can also coordinate both noun phrases like *şooíi*
 868 *tarapíi ga faaidá* ‘any benefit from the king’s side’ and *barawulxáanii tarapíi ga faaidá*
 869 ‘any benefit from Barawul Khan’ (24a) as well as clauses such as *tanaám the díti* ‘they
 870 gave them to someone else’ and *asaám the díti* ‘they gave them to us’ (24b).¹²

871

872 PALULA (Indo-European > Indo-Iranian, Pakistan; Liljegren 2016: 350)

- 873 (24) a. *Méejī na ta şoo-íi tarap-íi ga*

12 A parallel to Palula outside the sample can be found in geographically near but genealogically unrelated Purik-Sham-Nubra, which contrasts ENC markers *na ... na ...* and *hanna ... hanna ...* without them correlating with the phrasal-clausal distinction (Zemp 2018: 360–361).

874 between ENCM CNTR king-GEN direction-GEN any
875 *faaidá=ee na ba barawulxáan-ii tarap-ii ga*
876 benefit=CNJ ENCM TOP Barawul.Khan-GEN direction-GEN any
877 *faaidá*
878 benefit
879 ‘There were no benefits attached, neither from the king’s side, nor from
880 Barawul Khan’
881 b. *Na ta tanaám the díit-i na ba*
882 ENCM CNTR 3PL.ACC to give.PFV-F ENCM TOP
883 *asaám the díit-i*
884 1PL.ACC to give.PFV-F
885 ‘They didn’t give them to us or to anyone else’

886

887 The Palula, Croatian and Jamiltepec Mixtec data thus allow for the generalization that,
888 if a language has more than one ENC marker, there may be a division of labor relating to
889 the phrasal vs. clausal distinction of coordinands. This division of labor does not,
890 however, necessarily occur in all languages with more than one ENC marker.

891 The data enable a few more cross-linguistic generalizations with regard to the
892 properties of ENC constructions. First of all, it is quite rare for ENC markers A and B of
893 correlative (i.e., type- α 1-3) ENC constructions to differ in form: this is true of 2/31
894 (6.5%) sample languages with type- α 1-3 ENC constructions. Thus Icelandic has the ENC
895 markers *hvorki ... né ...* (Einarsson 1949: 175), whereas Maltese has *la ... u lanqas ...*
896 (Čéplö & Lucas 2020). In those languages in which ENC markers A and B do differ in
897 form, it is always the second one that is repeated when three or more functionally
898 equivalent elements are coordinated. Thus in English ENC is repeated as *neither ...*
899 *nor ... nor ...*, and not as **neither ... neither ... nor ...* (Quirk et al. 1985: 766). The same
900 is true of German *weder ... noch ... noch ...* (Fehringer 2014: 80) and Maltese *la ... u*
901 *lanqas ... u lanqas ...* (Čéplö & Lucas 2020: 192). Moreover, in correlative constructions
902 ENC markers A and B cannot switch positions. This is illustrated by German, where the
903 order *weder ... noch ...* is possible (25a), whereas *noch ... weder ...* is ungrammatical
904 (25b).

905

906 GERMAN (Indo-European > Germanic, Germany; Fehringer (2014: 80) and native
 907 judgments)

- 908 (25) a. *Es soll weder mir noch dir gehör-en*
 909 3SG.N should ENCM 1SG.DAT ENCM 2SG.DAT belong-INF
 910 ‘It should belong neither to me nor to you (sg.)’
 911 b. **Es soll noch mir weder dir gehören-en*
 912 3SG.N should ENCM 1SG.DAT ENCM 2SG.DAT belong-INF
 913 ‘It should belong neither to me nor to you (sg.)’

914

915 In addition, there are no Iquito examples in Lai (2009) in which ENC marker *nacaaja*
 916 precedes standard negative marker *ca=*, no Kalaallisut examples in Fortescue (1984) in
 917 which negative markers follow standard negator *-nngil-*, and no Palula examples in
 918 Liljegren (2016) where ENC marker B *na ba* precedes ENC marker A *na ta*. This suggests
 919 that neither ENC types $\alpha 1-3$ nor type β tolerate word order variation.

920 The data also suggest that in some languages with dedicated, language-specific ENC
 921 constructions ENC marker A can be omitted or replaced with a different negative marker,
 922 whereas this is never the case of successive ENC markers (i.e., B, C and so on). In
 923 French, for example, the first of ENC markers *ni ... ni ...* can be left out if preceded by
 924 the standard negative marker *ne* (26a). Alternatively, the position of ENC marker A can
 925 be taken up by negator *pas* (26b). None of those changes is possible for ENC marker B
 926 (26c).

927

928 FRENCH (Indo-European > Italic, France; Van der Auwera 2021 and native judgments)

- 929 (26) a. *Marie n'=aime (ni) le théâtre ni*
 930 Marie NEG=love.3SG.PRS ENCM DEF.M theater ENCM
 931 *l'=opéra*
 932 DEF=opera
 933 ‘Marie likes neither theater nor opera’
 934 b. *Marie n'=aime pas le théâtre ni*
 935 Marie NEG=love.3SG.PRS NEG DEF.M theater ENCM
 936 *l'=opéra*
 937 DEF=opera

938 'Marie likes neither theater nor opera'
 939 c. *Marie n'=aime ni le théâtre *(pas/Ø)*
 940 Marie NEG=love.3SG.PRS ENCM DEF.M theater NEG/Ø
 941 *l'=opéra*
 942 DEF=opera
 943 'Marie likes neither theater nor opera'

944

945 A similar case involves Dutch. At least in phrasal uses, the first of ENC markers *noch ...*
 946 *noch ...* can be omitted even when not preceded by the standard negative marker (27):

947

948 DUTCH (Indo-European > Germanic, Belgium, Netherlands; Van der Auwera 2021)

949 (27) (*Noch*) *oester-s noch mossel-en konden hem*
 950 ENCM oyster-PL ENCM mussel-PL could 3SG.OBL
 951 *bekor-en*
 952 tempt-INF
 953 'Neither oysters nor mussels could tempt him'

954

955 In short, then, ENC markers A and B of correlative constructions differ not only in form,
 956 but also with respect to word order, the possibility to be repeated in multiple
 957 coordination as well as the potential to be omitted or replaced with different negative
 958 elements.

959 Further generalizations are possible concerning the interaction between ENC markers
 960 and other negative markers, i.e., whether or not ENC markers trigger so-called negative
 961 concord. The details of this interaction are, however, beyond the scope of this study. For
 962 more on the topic see, among others, Jeretič (2018), Čeplö & Lucas (2020), Van der
 963 Auwera (2021), Van der Auwera et al. (2021) as well as literature cited therein.

964 One final abstraction concerns the emphatic nature of ENC. In this sense, first and
 965 foremost it should be mentioned that the first author to use the term 'emphatic' in
 966 reference to this clause linkage function, namely Haspelmath (2007: 3, 17–19), does not
 967 provide arguments in favor of viewing emphasis as an inherent property of ENC. In fact,
 968 research has been done on this topic without drawing on emphasis (Payne 1985b, Van
 969 der Auwera 2021). Nevertheless, a few arguments may be adduced to support the

970 insight that emphasis is part of the clause linkage function under discussion.

971 Firstly, as mentioned in Section 2 some languages make a formal distinction between
972 negative coordination and ENC. The functional motivation for this formal distinction is
973 unclear unless it is assumed that ENC provides an additional meaning. In this sense, it
974 has been shown above that most (31/37, 83.8%) languages with dedicated, language-
975 specific ENC constructions encode this clause-linkage function by means of correlative
976 markers, which in most (29/31, 93.5%) languages consist of formally identical members
977 A and B. Since ENC markers have different meanings —negative, emphatic negative,
978 conjunctive etc., see Section 5— when occurring individually, one may argue that
979 correlative ENC markers are in fact reduplicated elements. Bearing in mind that
980 reduplication often indicates emphasis (Hurch 2005) and under the aforementioned
981 assumption that ENC provides an additional meaning in comparison to negative
982 coordination, one might think that this additional meaning is, in fact, emphasis. Stated
983 differently, if ENC had no emphatic value, then a single (non-correlative) ENC marker
984 would suffice to coordinate and negate. Moreover, languages would not need to
985 formally differentiate negative coordination and ENC.

986 Secondly, studies on ENC constructions in individual languages have argued that they
987 encode emphasis, and that each coordinand is considered separately. Mairal & Ruiz de
988 Mendoza (2008: 177–178), for example, argue that in such English sentences as *I won't*
989 *eat that garbage, nor pay for it* the construction *not ... nor ...* indicates that the two
990 clauses *I won't eat that garbage* and *(I won't) pay for it* are complementary alternates. In
991 turn, the information that both clauses are alternates is not available in a plain negative
992 coordinated construction like *I won't eat that garbage; and I won't pay for it* (ibid.).

993 Jeschull (2004: 259–260) contrasts Chechen negative coordination, which ensues via
994 negated verbs —in this case *jaac* ‘be.NEG.PRS’— and the conjunctive coordinator ‘*a*
995 (28a), with ENC, which is conveyed by correlative markers *ja ... 'a, ja ... 'a* together
996 with negated verbs —in this case *ca weema* ‘NEG learn.PRS’ and *ca xae'a* ‘NEG
997 know.PRS’— (28b).

998

999 CHECHEN (Nakh-Daghestanian > Nakh, Russia; Jeschull 2004: 259–260)

- 1000 (28) a. *T'e-j-ucha* *bedar* 'a *j-aac*,
1001 on-CL-dress.INF garment and CL-be.NEG.PRS

1002		<i>t'e-xa'a</i>		<i>govr</i>		<i>'a</i>		<i>j-aac,</i>
1003		on-sit.down.INF		garment		and		CL-be.NEG.PRS
1004		<i>juq'-axdwa-d-iexka</i>		<i>giarz</i>		<i>'a</i>		<i>d-aac</i>
1005		waist-LAT		over-CL-tie.INF		weapon		and CL-be.NEG.PRS
1006		<i>san</i>						
1007		1 SG.GEN						
1008		'I don't have a garment to put on, a horse to sit down or a weapon to tie						
1009		around the waist'						
1010	b.	<i>Ja</i>		<i>caarna</i>		<i>ghullaq</i>		<i>d-an</i>
1011		or		3PL.DAT		service		CL-do.INF
1012		<i>weema,</i>		<i>ja</i>		<i>hwuuna</i>		<i>caerga</i>
1013		learn.PRS		or		2SG.DAT		3PL.ALL
1014		<i>muuxa</i>		<i>d-aita</i>		<i>d-ieza</i>		<i>'a</i>
1015		how		CL-let.make.INF		CL-must		and NEG
1016		<i>xae'a</i>						
1017		know.PRS						
1018		'Neither do they learn to do the service, nor do you know how you must						
1019		let them do this service'						

1020

1021 In the same vein as Mairal & Ruiz de Mendoza (2008: 177–178) claim for English,
1022 Jeschull (2004: 263) states concerning Chechen ENC (28b) that “bisyndesis in these
1023 cases conveys some kind of contrast”. Similarly, Alruwaili & Sadler (2019) argue that in
1024 Turaif Arabic the difference between negative coordination (*mā/mu* ‘NEG’ ... *w* ‘and’
1025 *mā/mu* ‘NEG’ ...) vs. ENC (*mā/lā* ‘NEG/ENCM’ ... *wala* ‘ENCM’ ...) is that the latter
1026 “provides an emphatic or focused alternative”. These statements are in line with the
1027 general view that emphasis is inherent to ENC and that each coordinand is considered
1028 separately (Haspelmath 2007: 15).

1029 The third and final argument in favor of emphasis as an inherent property of this
1030 clause linkage function involves the diachronic sources of ENC markers: one of the
1031 grammaticalization paths for ENC markers involves emphatic negators, cf. Hup *né* ...
1032 *né* ... in Section 2. Perhaps more importantly, marker B of ENC constructions seems to
1033 show a tendency for renewal by means of strengthening elements, usually adverbs (Van

1034 der Auwera 2021). One may point out, however, that ENC markers are no different from
1035 verbal negative markers in having emphatic particles as a source (Croft 1991: 5), and
1036 that emphasis may bleach over time (Jespersen 1917: 4). These points are further
1037 discussed in Section 5.

1038

1039 4.2. Areal tendencies of emphatic negative coordination

1040

1041 Another matter that relates to the phenomenon under discussion concerns the
1042 geographic distribution of each ENC type, i.e., α 1-3 and β . This is illustrated by Figure 4
1043 (where black dot (●) = type α 1; black square (■) = type α 2; black diamond (◆) = type
1044 α 3; white dot (○) = type β ; gray dot (◐) = any combination of the former):

1045

1046 **Figure 4.** Geographic distribution of ENC types as illustrated by the languages in the sample



1048 A number of generalizations are possible in light of Figure 4. First of all, dedicated,
1049 language-specific means for the expression of ENC are conspicuously absent from a
1050 number of macroareas including Australia-Papua New Guinea, North America, most of
1051 sub-Saharan Africa and large parts of South America. In turn, this kind of clause linkage
1052 clusters in very specific areas including Europe (Basque, Eastern Armenian, Hungarian,
1053 Icelandic, Kabardian, Karaim, Maltese, Turkish, Udmurt), South Asia (Brahui, Kharia,
1054 Khasi, Palula, Sinhala) and Mesoamerica (Comaltepec Chinantec, Jamiltepec Mixtec,
1055 Highland Popoluca, Huastec, Pipil, San Dionisio del Mar Huave, Yucatec Maya).

1056 By contrast, ENC types do not seem to be areally distributed. Plain correlative
1057 negative junction (type $\alpha 1$) is not only the most frequent ENC strategy as mentioned in
1058 Section 4.1, but also the most widespread one: it is found in all macroareas in which
1059 dedicated, language-specific ENC markers occur. Instead, correlative negative junction
1060 accompanied by the standard negator (type $\alpha 2$) is mostly found in Eurasia (Hungarian,
1061 Kabardian, Southern Yukaghir), whereas it is present in a single language in Africa
1062 (Midob) and by two in Mesoamerica (Huastec, Jamiltepec Mixtec). The distribution of
1063 correlative negative junction together with a coordinating conjunction (type $\alpha 3$) is
1064 likewise quite limited, as it is only found in five languages with one or more ENC
1065 strategy (Elamite, Palula, Rapanui, Sinhala, Turkish). The combination of negative
1066 junction and standard negator (type β) is found in a few languages located far apart from
1067 each other (Comaltepec Chinantec, Iquito, Japanese, Kalaallisut, San Dionisio del Mar
1068 Huave and Xamtanga), in the same way as a combination of two or more strategies
1069 (Basque, Jamiltepec Mixtec and Palula).

1070 The data thus indicate that dedicated ENC markers are not a peculiarity of European
1071 languages, as previously suggested (Haspelmath 2007: 17), but are, in any case, most
1072 widespread in Eurasia. This is shown by the fact that, apart from European languages,
1073 others like Brahui, Kharia, Khasi, Nivkh (Amur) and Southern Yukaghir, which are
1074 spoken in different areas of Asia, also display ENC markers. Indeed, the geographic
1075 clustering of languages with ENC markers suggests contact in general to be a relatively
1076 strong trigger for their development. This is in line with previous claims: as pointed out
1077 by Mithun (1988: 351–352), Matras (1998: 285) and Haspelmath (2007: 7–8),
1078 coordinating conjunctions are particularly prone to spreading under contact.

1079 The relative impact of contact in the emergence of dedicated, language specific ENC
1080 constructions is further supported by documented cases of borrowing that involve ENC
1081 markers: Sawknah-Fogaha, a Berber language of Libya, is reported to have acquired the
1082 ENC marker *la-bâ* ‘neither, nor’ on the basis of the North African Arabic negator *la* and
1083 the verb form (*ma*) *ba* ‘will (not)’ (Lakfioui & Brugnatelli 2020: 974, fn. 6). Karaim, a
1084 Turkic language spoken in eastern Europe, has apparently adopted the ENC marker *ni*
1085 ‘neither, nor’ from Slavic *ni* ‘neither, nor’ (Musaev 2003: 26–27). Furthermore, Kharia,
1086 an Austroasiatic language native to eastern India and Nepal, is claimed to have
1087 borrowed the item *na*, which acts both as a negator and an ENC marker (cf. Section 4.1),

1088 from Indo-Aryan, where it has the same uses (Peterson 2011: 341).

1089 Borrowing of ENC markers often involves contact between colonial and indigenous
1090 languages. This is particularly evident in Mesoamerica and parts of South America.
1091 Stolz & Stolz (1996: 100) find that borrowing of conjunctions is prevalent in the area,
1092 where nearly 30 languages from different genealogical groupings have each adopted at
1093 least one of up to 18 Spanish conjunctions. Of those 30 languages, nine (Huastec,
1094 Mayo, Mezontla Popoloca, Papantla Totonac, Pipil, Sierra de Juárez Zapotec as well as
1095 different varieties of Mixtec, Nahuatl and Otomi) are reported to have borrowed the
1096 Spanish correlative ENC marker *ni ... ni ...*. To these the following cases should be
1097 added: according to Sakel (2004: 332), Mosetén-Chimane, a Bolivian language isolate,
1098 adopted the ENC marker *ni ... ni ...* from Spanish, just the same as San Dionisio del Mar
1099 Huave *ñing ... ñing ...* (Salminen 2017: 99) and Garifuna *ní ... ní ...* (Munro & Gallagher
1100 2014: 44). The Hup correlative ENC marker *né ... né ...* may likewise be the result of
1101 borrowing of the Portuguese conjunction *nem* ‘nor’ via Tariana or Eastern Tucanoan
1102 (Epps 2008: 736–737).

1103 The fact that conjunctions in general and ENC markers in particular are prone to
1104 borrowing does not imply that adopted elements cannot be subject to changes in the
1105 borrowing language. In Jamiltepec Mixtec, for example, the borrowed ENC markers *nī...*
1106 *nī...* must be obligatorily accompanied by a standard negator (*mā*) in each of the clausal
1107 coordinands (29), as mentioned in Section 4.1.

1108

1109 JAMILTEPEC MIXTEC (Otomanguean > Eastern Otomanguean, Mexico; Johnson 1988:
1110 127)

1111 (29) *Nī mā kāchī ra nī mā kusu ra*
1112 ENCM NEG POT.eat 3SG.M ENCM NEG POT.sleep 3SG.M
1113 ‘He will neither eat nor sleep’

1114

1115 In (29) the coordinands *kāchī ra* ‘he will eat’ and *kusu ra* ‘he will sleep’ are
1116 accompanied by one instance of *nī* and one instance of *mā* each, and leaving out any of
1117 the two would result in ungrammaticality (Stolz & Stolz 1996: 94). This is unlike the
1118 source language, Spanish, where *ni* usually stands alone in every coordinand and may
1119 be optionally accompanied by the adverb *tampoco* ‘either’. Thus in Jamiltepec Mixtec

1120 the correlative ENC marker *nī* does not replace negation, but is rather reinforced by it, in
 1121 line with the emphatic nature of ENC.

1122 Furthermore, adoption of foreign ENC markers does not entail the loss of native
 1123 clause-linkage functions. Instead, original and borrowed structures can coexist and
 1124 complement each other. This is exemplified by Udmurt, where the native negative
 1125 coordinating structure that ensues via the disjunct *no* ‘either, or’ (30a-b) exists in
 1126 parallel with the ENC construction based on *ńe* (30c), a borrowing from Russian.

1127

1128 UDMURT (Uralic > Permian, Kazakhstan, Russia; Edygarova 2015: 287)

- 1129 (30) a. *So kīrʒa-nī no, ektī-nī no u-g*
 1130 3SG sing-INF either dance-INF either NEG.PRS-3
 1131 *bigatī*
 1132 can.SG
 1133 ‘(S)he can neither sing nor dance (lit. she cannot either sing or dance)’
- 1134 b. *Až-iz no, ber-iz no evel*
 1135 front-3SG either back-3SG either NEG
 1136 ‘It has neither front nor back (lit. it does not have either front or back)’
- 1137 c. *Ńe vań, ńe evel, ńe sil, ńe čorig –*
 1138 ENCM EX ENCM NEG ENCM meat ENCM fish
 1139 *šu-e žuč kalik*
 1140 say-PRS.3SG Russian people
 1141 ‘The Russians say: it neither exists nor does not exist, it is neither fish
 1142 nor meat’

1143

1144 According to Edygarova (2015: 287), the ENC construction based on *ńe* in (30c) can
 1145 replace the native negative coordinating function that ensues via *no* (30a-b), but this is
 1146 not obligatory.

1147 The insight that the presence of dedicated, language-specific ENC constructions
 1148 correlates with macroarea can be tested on grounds of the statistical analysis laid out in
 1149 Section 3.3. A look at the results provides weak support for this view: there is a medium
 1150 correlation between the dependent variables ENC_AVAILABLE and MACRO_AREA
 1151 according to Cramer’s V (.46), which is, however, quite low in keeping with Theil’s U

1152 (.24). The weakness of this correlation can probably be explained by the fact that,
1153 despite being most widespread in Eurasian languages, ENC constructions are present to a
1154 lesser degree in languages of Africa, North America-Mesoamerica, South America and
1155 Southeast Asia-Oceania.

1156 In the same vein, the correlation between the dependent variables ENC_AVAILABLE
1157 and GENEALOGICAL_AFFILIATION is quite weak according to Cramer's V (.28). In turn,
1158 the results of Theil's U show a very strong correlation (.82) between
1159 GENEALOGICAL_AFFILIATION and ENC_AVAILABLE. These apparently contradictory data
1160 can be interpreted as follows: the presence of dedicated, language-specific ENC
1161 constructions does not presuppose a specific genealogical affiliation for a given sample
1162 language. By contrast, being a member of a specific family strongly conditions the
1163 likelihood for a sample language to have dedicated ENC markers. Indeed, most (24/37,
1164 64.9%) sample languages with dedicated, language-specific ENC constructions belong to
1165 one of ten genealogical groupings: these are Indo-European (Eastern Armenian,
1166 Icelandic, Palula, Sinhala), Afro-Asiatic (Maltese, Sawknah-Fogaha, Xamtanga),
1167 Austronesian (Rapanui, Sirayaic), Austroasiatic (Kharia, Khasi), Uralic (Hungarian,
1168 Udmurt), Otomanguean (Comaltepec Chinantec, Jamiltepec Mixtec), Mayan (Huastec,
1169 Yucatec Maya) and Turkic (Karaim, Turkish). There are also three language isolates
1170 (Basque, Elamite, Nivkh (Amur)) and two pidgins and creoles (Papiamentu, Ternateño).

1171 Further correlations worth mentioning include those between ENC_AVAILABLE and
1172 RELATED_TO_SN (Cramer's V = 1, Theil's U = .81), ENC_AVAILABLE and
1173 DISTINCT_FROM_SN (Cramer's V = 1, Theil's U = .87) as well as DISTINCT_FROM_SN
1174 and RELATED_TO_SN (Cramer's V = .8, Theil's U = .86). These data imply that, if a
1175 language has dedicated ENC markers, they will very likely be diachronically related to
1176 and, at the same time, formally identical with markers of standard negation. For further
1177 details on this topic, see Section 5.

1178 In summary, the sample data suggest that dedicated, language-specific ENC
1179 constructions are widespread among Eurasian languages. By contrast, many instances of
1180 this clause-linkage function outside the Eurasian macroarea —as well as some within it
1181 — can be traced back to contact between indigenous and colonial languages. This
1182 finding is in line with previous claims that conjunctions are prone to spreading in
1183 contact situations. Borrowing of ENC markers does not, however, necessarily imply that

1184 native clause-linkage functions are replaced. Rather, both kinds of construction can
1185 coexist and even complement each other. Moreover, the properties of adopted ENC
1186 constructions do not have to be identical to those of the donor language. Instead, the
1187 borrowed structures undergo adaptations specific to the target language.

1188

1189 **5. Grammaticalization paths leading to ENC markers**

1190

1191 Languages have been observed to exhibit a considerable variety of coordinating
1192 constructions emerging as the result of diverse grammaticalization paths (Mithun 1988:
1193 331–349). By extension, one would expect to find different sources of
1194 grammaticalization leading to many kinds of ENC markers. Nevertheless, the fact that
1195 ENC and negation are semantically close (cf. Section 2) suggests that ENC markers and
1196 negators should tend to be related at the formal level as well, which narrows down the
1197 potential number of diachronic sources.¹³

1198 The diachronic data laid out in this section have been gathered systematically, i.e.,
1199 they originate in the 37 sample languages that were argued in Section 4.1 to make use of
1200 dedicated, language-specific ENC markers. Furthermore, when discussing specific cases
1201 additional information is drawn from languages outside the sample so as to underpin the
1202 existence of particular grammaticalization paths.

1203 In line with the aforementioned prediction, the simplest attested diachronic scenario
1204 involves coexpression of ENC marker(s) and standard negator. This state of affairs is
1205 present in 4/37 (10.8%) of the sample languages with dedicated, language-specific ENC
1206 markers (Basque, Eastern Armenian, Elamite, Sinhala). In these languages there is no
1207 formal distinction between standard negator and ENC marker, cf. Basque *ez ~ ez ... ez ...*
1208 (Hualde & Ortiz de Urbina 2003: 562), Eastern Armenian *oč ~ oč ... oč ...* (Dum-Tragut
1209 2009: 289), Elamite *in- ~ in- ... in- ...* (Khačikjan 1998: 49, 55) and Sinhala *nææ ~ nææ*
1210 *... nææ ...* (Chandralal 2010: 186). Parallels are also found outside the sample, including
1211 Gurani *na ~ na ... na ...* (Mahmoudveysi et al. 2012: 236) and Northern Tosk Albanian
1212 *as ~ as ... as ...* (Camaj 1984: 243). Stilo (2004) mentions ENC marker-standard negator
1213 coexpression in three Iranian languages: Vafsi, Persian and Gilaki. Bearing in mind

13 This prediction has been articulated in previous literature: “[the] coding [of ENC markers] is obviously strictly connected with the strategies that languages use to express negation, which is a different, albeit related, notion” (Mauri 2008: 51).

1214 aforementioned Sinhala and Gurani, the formal identity of ENC markers and standard
1215 negators thus seems to be particularly prevalent in Indo-Iranian languages.

1216 More frequently, however, standard negators combine with other elements in the
1217 process of grammaticalization of ENC markers. These other elements essentially
1218 encompass coordinators and adverbs. Among the languages of the sample, 10/37
1219 (27.0%) (Hungarian, Icelandic, Japanese, Kabardian, Kalaallisut, Karamojong, Maltese,
1220 Palula, Sirayaic, Yucatec Maya) have ENC constructions that represent this development.
1221 Specifically, negators may join, for example, with adverbs or scalar focus particles
1222 which under the scope of negation translate as ‘(not) even’. An example is Yucatec
1223 Maya, whose ENC marker *mix* is believed to result from blending of standard negator
1224 *ma* ‘NEG’ and adverb *ix* ‘even’ (31b, Bolles & Bolles 2001: 56). *Mix* also has a ‘not
1225 even’ meaning outside of coordination (31a):

1226

1227 YUCATEC MAYA (Mayan > Core Mayan, Guatemala, Mexico; Bolles & Bolles 2001: 56,
1228 Yoshida 2011: 125)

1229 (31) a. *Mix in u-ohel*
1230 not.even 1SG 1SG-know
1231 ‘I don’t even know’

1232 b. *Ma’ k’-ahóolt-a’an-i’, mix ohelt-a’an tu’ux u*
1233 NEG HAB-know-3PL-PART ENCM know-3PL place from
1234 *taal-i’*
1235 come-PART

1236 ‘(S)he was not well-known, nor did they know where (s)he came from’

1237

1238 Beyond the sample of languages a less grammaticalized counterpart of Yucatec Maya
1239 *ma + ix > mix* is found in Sikkimese. In the latter language, unlike in the former, the
1240 standard negative prefix *mi-*, *ma-* and the adverb $=j\ddot{a}$: ‘(not) even’ have not
1241 univerbated.¹⁴ Apparently this is because these elements have different hosts: $=j\ddot{a}$:
1242 cliticizes to coordinands, whereas *mi-*, *ma-* attaches to the verb, i.e., $m\grave{e}^? < mi- + j\grave{o}^?$
1243 ‘personal existential’ (Yliniemi 2019: 459). This can be seen in (32):

1244

14 Lehmann (2020: 205) defines univerbation as “the syntagmatic condensation of a sequence of words recurrent in discourse into one word”.

1245 SIKKIMESE (Sino-Tibetan > Bodic, India; Yliniemi 2019: 459)

1246 (32) *Rap=jã:* *mè?* *t^hama=jã:* *mè?*
1247 supreme=even NEGEX.PER last=even NEGEX.PER
1248 ‘(It) is neither first-class nor last (in quality)’

1249

1250 Examples (31a-b) and (32) thus illustrate that adverbs and scalar focus particles
1251 meaning ‘(not) even’ also constitute the basis for ENC markers outside European
1252 languages. Therefore, this grammaticalization path is more widespread than was
1253 previously believed (Haspelmath 2007: 17–18). This state of affairs is moreover in
1254 agreement with claims on the diachronic behavior of coordinators, since the
1255 grammaticalization of adverbial particles as coordinators seems to be quite frequent
1256 (Mithun 1988: 340).

1257 Other elements that may combine with standard negators to produce ENC markers
1258 include conjunctive coordinators. For example, Hungarian is believed to have
1259 grammaticalized ENC marker *sem* from the fusion of conjunctive coordinator *is* ‘also’
1260 and standard negator *nem* (33a, Károly 1984: 35). When used monosyndetically, *sem*
1261 has a ‘and not, also not’ meaning (33b, *ibid.*):

1262

1263 HUNGARIAN (Uralic > Hungarian, Hungary, Slovakia; De Groot 1994: 155, Kenesei et
1264 al. 1998: 117)

1265 (33) a. *Nem szabad sem inni, sem enni*
1266 NEG allow ENCM drink.INF ENCM eat.INF
1267 ‘It is not allowed to eat nor to drink’
1268 b. *Nem olvasta a könyvet Anna sem*
1269 NEG read.3SG.PST DEF book.ACC Anna not.also
1270 ‘(In addition to others) Anna too didn’t read the book’

1271

1272 ENC markers also stem from the univerbation of negative marker with both scalar focus
1273 particle and conjunctive coordinator. In South Bolivian Quechua, for instance, ENC
1274 marker *nillataj* is believed to result from a merger of negative marker *ni* —presumably a
1275 Spanish borrowing—, the adverb *lla* ‘just’ and a conjunction, *taj* ‘and’ (Camp &
1276 Liccardi 1967: 93). The ENC marker is sometimes reduced to *nitaj*, as illustrated by (34):

1277

1278 SOUTH BOLIVIAN QUECHUA (Quechuan > Quechua II, Argentina, Bolivia; Herrero &
1279 Sánchez de Lozada 1978: 195)

1280 (34) *Mana Jirmin ni(-lla-)taj doña Amelya*

1281 NEG Fermín NEG-just-and doña Amelya

1282 *mana-pu-ni(-lla-)taj Damyana risqan-ku-ču*

1283 NEG-MOD-NEG-just-and Damyana look-PL-NONFACT

1284 ‘Neither Fermín nor doña Amelya nor Damyana are looking’

1285

1286 The South Bolivian Quechua case is similar to Maltese, where the ENC construction
1287 *la ... u lanqas ...* consists of negator *la*, conjunction *u* ‘and’ and *lanqas* (< *la* + *anqas*
1288 ‘less’) (Čeplö & Lucas 2020: 182).

1289 Another possibility illustrated by the sample languages is for non-standard negators
1290 to coexpress ENC, either on their own or in combination with other elements. Among the
1291 languages of the sample 2/37 (5.4%) (Hup, Rapanui) represent this situation. As
1292 mentioned in Section 2, in Hup correlative emphatic negators *né ... né ...* are used to
1293 indicate ENC. In turn, in Rapanui constituent negator *ta’e* combines with negative
1294 coordinator *ni* in order to encode this kind of clause linkage (35):

1295

1296 RAPANUI (Austronesian > Malayo-Polynesian, Chile; Kieviet 2017: 506)

1297 (35) *A Hiero poki ta’e porio ni ta’e pāpaku*

1298 PROP Hiero child NEG fat ENCM NEG thin

1299 ‘Hiero was neither a fat nor a skinny child’

1300

1301 To summarize so far, negators are behind the emergence of dedicated, language-specific
1302 ENC markers in 16/37 (43.2%) of all sample languages with this kind of clause linkage.
1303 Depending on the language, standard negators or non-standard negators may participate
1304 in grammaticalization, they may coexpress ENC on their own or in combination with
1305 other elements, and they may or may not fuse with other elements.

1306 Negators are not, however, an exclusive source of ENC markers. In fact, there is at
1307 least another grammaticalization path that involves verbs. Two languages in the sample
1308 (2/37, 5.4%, Comaltepec Chinantec, Midob) represent this development. In Comaltepec

1309 Chinantec for example, there is the stative negative verb ʔóʔ^L (Anderson 1989: 30).
 1310 When used on its own, ʔóʔ^L negates predications (36a), but “it seems to give more
 1311 attention to the negation than does the negative prefix” (ibid.). This suggests that ʔóʔ^L
 1312 may in fact be used as an emphatic negator. In turn, when combined with the standard
 1313 negative prefix ha^L - and an augmentative suffix $-gi^{H}$, ʔóʔ^L expresses ENC (36b-c).

1314

1315 COMALTEPEC CHINANTEC (Otomanguean > Western Otomanguean, Mexico; Anderson
 1316 1989: 29–30)

1317 (36) a. $Zánʔ^L$ ʔóʔ^L $so.^L$ zi^{LM} ʔi^L $zā^L$ na^L
 1318 truly NEG rise heart REL person that
 1319 ‘That man is definitely not mean’

1320 b. $Ha^L-sinʔ^{LM}-r$ ki^Lni^{HM} ʔi^L $hiénʔ^L$ ʔóʔ^L-gi^{H} ki^Lni^{HM}
 1321 NEG-stand-3 before REL other NEG-AUG before
 1322 $dié^{LH}$
 1323 god

1324 ‘They do not stand before another nor before a god’

1325 c. $Ha^L-ni^L-tiú^L$ hmi^{LM} ʔóʔ^L-gi^{H} ʔmoʔ^{LH} $niú^{LM}$
 1326 NEG-IPFV-pour rain NEG-AUG dew star
 1327 ‘It will neither rain nor will it dew’

1328

1329 In short, then, the stative negative verb ʔóʔ^L , which has emphatic negative uses, also
 1330 encodes ENC in Comaltepec Chinantec. A similar development can be observed in
 1331 Midob, where the correlative ENC marker $ínyén \dots inyéén \dots$ (37) apparently comes from
 1332 $-ín$, a variant of the copula stem $-an$ (plural $-jûm$) ‘to be’, in its 3rd-person singular
 1333 continuous subjunctive form:

1334

1335 MIDOB (Nubian > West-Central Nubian, Sudan; Werner 1993: 63)

1336 (37) $Án$ $ittir$ $ínyén$ $àn$ $éd$ $ínyén$ $iir-áa-hàm$
 1337 this woman ENCM this man ENCM come-NEG-PST
 1338 ‘Neither this woman nor this man came’

1339

1340 Stated differently, sentences such as (37) originally translate as ‘be it a woman, be it a

1341 man, (s)he did not come'. Following Werner (1993: 67), this implies that this variant of
1342 the copula must have taken on an inherently negative meaning as the result of its being
1343 reanalyzed as an ENC marker.

1344 In addition to negators (16/37, 43.2%) and verbs (2/37, 5.4%), the languages of the
1345 sample illustrate two more diachronic scenarios. The first of these involves borrowing,
1346 which as mentioned in Section 4.2 is a recurrent possibility that involves 13/37 (35.2%)
1347 of the sample languages with dedicated, language-specific ENC markers (Highland
1348 Popoluca, Huangascar-Topara-Yauyos Quechua, Huastec, Jamiltepec Mixtec, Karaim,
1349 Kharia, Papiamento, Pipil, San Dionisio del Mar Huave, Sawknah-Fogaha, Southern
1350 Yukaghir, Ternateño, Udmurt). The second possibility refers to ENC markers whose
1351 etymology is not straightforward, i.e., 6/37 (16.2%) languages (Brahui, Iquito, Khasi,
1352 Nivkh, Turkish, Xamtanga). Of the latter group, however, Brahui *na ... na ...*, *ney ...*
1353 *ney ...* and Turkish *ne ... ne ... (de)* are conspicuously reminiscent of the Indo-Iranian *nV*
1354 *... nV ...* ENC marker pattern. Therefore, the possibility that these markers are due to
1355 diffusion from the geographically adjacent Indo-Iranian languages cannot be excluded.

1356 A final point worth analysis involves the subsequent diachronic changes undergone
1357 by grammaticalized ENC markers and constructions. An argument has been made in
1358 Section 4.1 in favor of considering emphasis as an inherent property of ENC. However,
1359 emphasis is known to bleach over time, for example in the case of negators (Jespersen
1360 1917: 4), which have also been shown to play an important part in ENC. This
1361 observation suggests that ENC markers lose strength over time, and it predicts two
1362 possible outcomes: ENC markers can lose emphasis and thus lack a functional
1363 motivation to be formally distinguished from negative coordination. Even though no
1364 such diachronic change is documented in the sample, this possibility does seem to
1365 account for the fact that most (213/250, 85.2%) sample languages do not discriminate
1366 ENC and negative coordination.

1367 Alternatively, languages can implement strengthening mechanisms in order to
1368 counterbalance emphasis bleaching. This development is indeed attested: in Spanish
1369 ENC marker B of construction *ni ... ni ...* can be optionally reinforced by means of the
1370 adverb *tampoco* 'either', both with phrasal (38a) and clausal (38b) coordinands.

1371

1372 SPANISH (Indo-European > Italic, Spain; Sánchez 2017: 673, Van der Auwera 2021)

- 1373 (38) a. *Ni Fulano ni (tampoco) Mengano salieron*
 1374 ENCM Fulano ENCM either Mengano leave.3PL.PST
 1375 ‘Neither Fulano nor Mengano left’
 1376 b. *Él ni ha escrito una novela ni (tampoco)*
 1377 3SG.M ENCM AUX write INDEF.F novel ENCM either
 1378 *quiere escrib-ir=la*
 1379 want.3SG.PRS write-INF=3SG.F
 1380 ‘He has neither written a novel nor wants to write one’
 1381

1382 In French, as well, ENC marker B of construction *ni ... ni ...* can be strengthened by *non*
 1383 *plus* ‘not either’ (39a) and *davantage* ‘either’ (39b), as long as the coordinands are
 1384 phrasal.

1385
 1386 FRENCH (Indo-European > Italic, France; Van der Auwera 2021 and native judgments)

- 1387 (39) a. *Marie n’=aime pas le théâtre, ni non*
 1388 Marie NEG=love.3SG.PRS NEG DEF.M theater ENCM NEG
 1389 *plus l’=opéra*
 1390 either DEF=opera
 1391 ‘Marie doesn’t love theater and not opera either’
 1392 b. *L’=exclusion sociale ne produit ni plus*
 1393 DEF=exclusion social NEG produce.3SG.PRS ENCM more
 1394 *de croissance ni davantage d’=emplois*
 1395 of growth ENCM either of=employment
 1396 ‘Social exclusion produces neither more growth nor more employment’
 1397

1398 Thus in those cases in which an ENC construction is reinforced, it is apparently ENC
 1399 marker B (*ni tampoco, ni non plus*) that undergoes strengthening and form renewal,
 1400 whereas ENC marker A (*ni, ni*) remains intact.¹⁵ This is a parallel to Jespersen’s Cycle
 1401 (Jespersen 1917).

1402 The effects of Jespersen’s Cycle seem to account for two more features of ENC
 1403 constructions: on the one hand, strengthening of ENC marker B and preservation of ENC

¹⁵ An exception to this generalization may be Jamiltepec Mixtec, which as shown in example (29) strengthens both ENC markers A and B by means of the standard negator *mā*.

1404 marker A may be the reason why in some languages the form of ENC markers differs (cf.
1405 Section 4.1), i.e., Maltese *la ... u lanqas ...*, Turaif Arabic *mā/lā ... wala ...*, Icelandic
1406 *hvorki ... né ...*, German *weder ... noch ...* etc. On the other hand, the tendency for ENC
1407 marker B to renovate may explain why the different ENC types do not allow for markers
1408 A and B to switch positions (cf. Section 4.1).

1409 In summary, the data collected for this study support the prediction that ENC markers
1410 tend to be etymologically related and formally similar to negative markers (Mauri 2008:
1411 51). Nevertheless, negators are far from constituting an exclusive diachronic source of
1412 ENC markers. Rather, ENC markers have a sizable number of diachronic sources, and
1413 their degree of grammaticalization varies cross-linguistically. This is in accordance with
1414 previous claims on the diachronic sources of coordinators (Mithun 1988: 336–349,
1415 Haspelmath 2007: 48–49). Furthermore, ENC markers are not diachronically stable, but
1416 rather undergo processes of renovation through the addition of strengthening elements.

1417

1418 **6. Conclusions**

1419

1420 An overview of emphatic negative coordination in a representative sample of 250
1421 languages has revealed that only a reduced number of languages in the world have
1422 dedicated, language-specific means to express this clause linkage function. Moreover,
1423 ENC strategies can be classified into a limited number of types and subtypes. In spite of
1424 these restrictions, languages have been shown to exhibit considerable cross-linguistic
1425 variety concerning the number and type of strategies used by each, the kinds of markers
1426 that may participate in ENC, and the manner in which different ENC markers are
1427 exploited. The existence of ENC marker ‘splits’ in some languages of the sample thus
1428 shows that it is fruitful and necessary to compare the behavior of ENC constructions both
1429 within and across languages.

1430 In addition, the areal distribution of ENC types indicates, on the one hand, that
1431 dedicated, language-specific ENC constructions are mostly particular to the Eurasian
1432 macroarea, and on the other hand, that ENC types tend to cluster geographically. These
1433 findings have been considered to signal that contact and borrowing are relevant
1434 contributors to the development of dedicated ENC markers, which can be observed in
1435 many cases of contact between colonial and indigenous languages. In any case,

1436 borrowed ENC markers do not tend to replace indigenous linkage devices, but rather
1437 complement them. In the same vein, genetic affiliation seems to have a considerable
1438 impact on the likelihood for a language to present dedicated, language-specific ENC
1439 markers.

1440 Finally, a diachronic analysis of ENC markers has revealed that they are related to
1441 negation not only on a semantic level, but on a formal level as well, since the majority
1442 of ENC markers develop out of and are frequently identical to negators. Furthermore,
1443 markers of ENC are often similar to other kinds of coordinators in the sense that they
1444 have a considerable number of diachronic sources, and in that their degree of
1445 grammaticalization varies cross-linguistically. Typological generalizations such as
1446 Jespersen's Cycle can also help account for some features of ENC constructions, i.e.,
1447 their word order rigidity and their tendency to renovate ENC marker B by means of
1448 strengthening elements as a reaction to emphasis bleaching.

1449

1450 **Funding**

1451

1452 This project received financial support from the research group *Diacronía de la*
1453 *animacidad: aproximación tipológica al origen de las marcas animadas* headed by
1454 Professor Dr. Iván Igartua (University of the Basque Country) and funded by the
1455 Spanish Ministry of Science, Innovation and Universities through grant PGC2018-
1456 098995-B-I00. This support is gratefully acknowledged.

1457

1458 **Acknowledgements**

1459

1460 Earlier versions of this paper were presented virtually on 20 November 2020 at the *14th*
1461 *Biennial Conference of the High Desert Linguistics Society* organized by the University
1462 of New Mexico and on December 15th at the *53rd Annual Conference of the Australian*
1463 *Linguistic Society*. I would like to thank the audiences of these conferences, including
1464 William Croft, Jens Van Gysel and Yi-Yang Cheng for their valuable comments. Thanks
1465 are also due to Camilo Enrique Díaz Romero, Johan van der Auwera, the editors and
1466 two anonymous reviewers for constructive criticism, corrections and contributions to
1467 earlier versions of this paper. I would also like to express my deepest gratitude to Haritz

1468 Salaberri for indispensable help with regression analysis of the data and for making the
1469 corpus accessible publicly. Anne Wolfsgruber and Maitena Duhalde helped me with
1470 native judgments on German and French examples, respectively. Any remaining errors
1471 are the author's sole responsibility.

1472

1473 **References**

1474

1475 Adelaar, Alexander. 2011. *Siraya: Retrieving the phonology, grammar and lexicon of a*
1476 *dormant Formosan language*. Berlin: De Gruyter.

1477 Adelaar, Willem F. H. 2007. Ensayo de clasificación del Katawixí dentro del conjunto
1478 Harakmbut-Katukina. In Andrés Romero-Figueroa, Ana Fernández-Garay & Ángel
1479 Corbera Mori (eds.), *Lenguas indígenas de América del Sur: Estudios descriptivo-*
1480 *tipológicos y sus contribuciones para la lingüística teórica*, 159–169. Caracas:
1481 Universidad Católica Andrés Bello.

1482 Aikhenvald, Alexandra Y. 2003. *A grammar of Tariana, from northwest Amazonia*.
1483 Cambridge: Cambridge University Press.

1484 Alruwaili, Shatha & Louisa Sadler. 2019. Negative coordination in (Turaif) Arabic. In
1485 Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG'18 conference*, 25–45.
1486 Stanford: Stanford University.

1487 Anderson, Judi L. 1989. *Comaltepec Chinantec syntax*. Dallas: Summer Institute of
1488 Linguistics, University of Texas at Arlington.

1489 Andronov, Mikhail S. 1980. *The Brahui language*. Moscow: Akademija Nauk.

1490 Andronov, Mikhail S. 2001. *A grammar of the Brahui language in comparative*
1491 *treatment*. Munich: Lincom.

1492 Bakker, Dik. 2011. Language sampling. In Jae J. Song (ed.), *The Oxford handbook of*
1493 *linguistic typology*, 100–127. Oxford: Oxford University Press.

1494 Barjasteh, Behrooz. 2018. Two Brahui texts with glossary and grammatical analysis.
1495 *Iranian Journal of Applied Language Studies* 12(1). 89–122.

1496 Bernini, Giuliano & Paolo Ramat. 1996. *Negative sentences in the languages of*
1497 *Europe: A typological approach*. Berlin: De Gruyter.

1498 Bickel, Balthasar, Bernard Comrie & Martin Haspelmath. 2015. *Leipzig glossing rules*.
1499 (Available online at <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>)

- 1500 Black, Paul E. 2006. *Dictionary of algorithms and data structures*. Gaithersburg:
 1501 National Institute of Standards and Technology.
- 1502 Bolles, David & Alejandra Bolles. 2001. *A grammar of the Yucatecan Mayan language*.
 1503 Lancaster: Layrinthos Press.
- 1504 Bond, Oliver. 2011. Negation in clause linkages. In Julia Sallabank (ed.), *Language*
 1505 *documentation and description*, Vol. 9, 77–120. London: School of Oriental and
 1506 African Studies.
- 1507 Bowler, Margit. 2014. Conjunction and disjunction in a language without ‘and’. In Todd
 1508 Snider, Sarah D’Antonio & Mia Weigand (eds.), *Proceedings of the 24th semantics*
 1509 *and linguistic theory conference*, 137–155. Ithaca: Linguistic Society of America.
- 1510 Briley, David. 1997. Four grammatical marking systems in Bauzi. In Karl J. Franklin
 1511 (ed.), *Papers in Papual Linguistics*, Vol. 2, 1–131. Canberra: Australian National
 1512 University.
- 1513 Brown, Keith & Jim Miller. 2013. *The Cambridge dictionary of linguistics*. Cambridge:
 1514 Cambridge University Press.
- 1515 Bussmann, Hadumod. 1998. *Routledge dictionary of language and linguistics*.
 1516 London/New York: Routledge.
- 1517 Camaj, Martin. 1984. *Albanian grammar with exercises, chrestomathy and glossaries*.
 1518 Wiesbaden: Otto Harrassowitz.
- 1519 Camp, Elizabeth L. & Millicent R. Liccardi. 1967. Quechua. In Esther Matteson (ed.),
 1520 *Bolivian indian grammars*, Vol. 2, 27–98. Norman: University of Oklahoma,
 1521 Summer Institute of Linguistics.
- 1522 Čéplö, Slavomír & Christopher Lucas. 2020. *Langas*, negative concord and predicate
 1523 negation in Maltese. In Slavomír Čéplö & Jaroslav Drobný (eds.), *Maltese*
 1524 *linguistics on the Danube*, 177–198. Berlin: De Gruyter.
- 1525 Chandralal, Dileep. 2010. *Sinhala*. Amsterdam/Philadelphia: Benjamins.
- 1526 Churchward, Maxwell C. 1953. *Tongan grammar*. Oxford: Oxford University Press.
- 1527 Cramer, Harold. 1946. *Mathematical methods of statistics*. Princeton: Princeton
 1528 University Press.
- 1529 Croft, William. 1991. The evolution of negation. *Journal of Linguistics* 27(1). 1–27.
- 1530 Croft, William. 2003. *Typology and universals*, 2nd edition. Cambridge: Cambridge
 1531 University Press.

- 1532 Crystal, David. 2008. *A dictionary of linguistics and phonetics*, 6th edition. London:
1533 Blackwell.
- 1534 De Groot, Casper. 1994. Hungarian. In Peter Kahrel & René van den Berg (eds.),
1535 *Typological studies in negation*, 143–162. Amsterdam/Philadelphia: Benjamins.
- 1536 De Swart, Henriëtte. 2020. Double negation readings. In Viviane Déprez & M. Teresa
1537 Espinal (eds.), *The Oxford handbook of negation*, 479–496. Oxford: Oxford
1538 University Press.
- 1539 Dik, Simon C. 1997. Coordination. In Kees Hengeveld (ed.), *The theory of functional*
1540 *grammar*, Vol. 2, 189–214. Berlin/New York: De Gruyter.
- 1541 Dos Anjos, Zoraide. 2011. *Fonologia e gramática katukina-kanamari*. Amsterdam: Vrije
1542 Universiteit PhD dissertation.
- 1543 Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in*
1544 *Language* 13(2). 257–292.
- 1545 Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1).
1546 81–138.
- 1547 Dum-Tragut, Jasmine. 2009. *Armenian: Modern Eastern Armenian*.
1548 Amsterdam/Philadelphia: Benjamins.
- 1549 Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). *Ethnologue:*
1550 *Languages of the world*, 24th edition. Dallas: Summer Institute of Linguistics.
- 1551 Edmonson, Barbara W. 1988. *A descriptive grammar of Huastec (Potosino dialect)*.
1552 New Orleans: Tulane University PhD dissertation.
- 1553 Edygarova, Svetlana. 2015. Negation in Udmurt. In Matti Miestamo, Anne Tamm &
1554 Beáta Wagner-Nagy (eds.), *Negation in Uralic languages*, 265–292.
1555 Amsterdam/Philadelphia: Benjamins.
- 1556 Einarsson, Stefán. 1949. *Icelandic grammar, text, glossary*. Baltimore: The Johns
1557 Hopkins Press.
- 1558 Epps, Patience. 2008. *A grammar of Hup*. Berlin/New York: De Gruyter.
- 1559 Fehring, Carol. 2014. *German grammar in context: Analysis and practice*, 2nd
1560 edition. London/New York: Routledge.
- 1561 Fortescue, Michael. 1984. *West Greenlandic*. London/Sydney/Dover: Croom Helm.
- 1562 Göksel, Asli & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London/New
1563 York: Routledge.

- 1564 Grimes, Barbara F. (ed.). 1997. *Ethnologue: Languages of the world*, 13th edition.
 1565 Dallas: Summer Institute of Linguistics.
- 1566 Groth, Christa. 1985. Syntax of the phrase types in Canamari. In David L. Fortune (ed.),
 1567 *Porto Velho workpapers*, 93–129. Brasil: Summer Institute of Linguistics.
- 1568 Hammarström, Harald. 2009. Sampling and genealogical coverage in *WALS*. *Linguistic*
 1569 *Typology* 13(1). 105–119.
- 1570 Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2020.
 1571 *Glottolog 4.3*. Jena: Max Planck Institute for the Science of Human History.
 1572 (Available online at <http://glottolog.org>, accessed on 24-03-2021.)
- 1573 Haspelmath, Martin. 2007. Coordination. In Timothy Shopen (ed.), *Language typology*
 1574 *and syntactic description*, Vol. 2, 2nd edition, 1–51. Cambridge: Cambridge
 1575 University Press.
- 1576 Herrero, Joaquín & Federico Sánchez de Lozada. 1978. *Diccionario español-quechua:*
 1577 *Estructura semántica del quechua cochabambino contemporáneo*. Cochabamba:
 1578 Editorial Universo.
- 1579 Hualde, José I. & Jon Ortiz de Urbina. 2003. *A grammar of Basque*. Berlin/New York:
 1580 Mouton de Gruyter.
- 1581 Hurch, Bernhard (ed.). 2005. *Studies on reduplication*. Berlin/New York: De Gruyter.
- 1582 Ishy de Magalhães, Priscila H. 2018. *Kanamari do Juruá (família katukina): Aspectos*
 1583 *fonológicos e morfosintáticos*. Campinas: Universidade Estadual de Campinas PhD
 1584 dissertation.
- 1585 Jeretič, Paloma. 2018. Hybrid and optional negative concord in Turkish: A unified
 1586 analysis. *Papers in Turkish and Turkic Linguistics* 18(2).
- 1587 Jeschull, Liane. 2004. Coordination in Chechen. In Martin Haspelmath (ed.),
 1588 *Coordinating constructions*, 241–266. Amsterdam/Philadelphia: Benjamins.
- 1589 Jespersen, Otto. 1917. *Negation in English and other languages*. Copenhagen: Bianco
 1590 Lunos Bogtrykkeri.
- 1591 Johnson, Audrey F. 1988. A syntactic sketch of Jamiltepec Mixtec. In Henry Bradley &
 1592 Barbara E. Hollenbach (eds.), *Studies in the syntax of Mixtecan languages*, Vol. 1,
 1593 11–150. Dallas: Summer Institute of Linguistics, University of Texas at Arlington.
- 1594 Károly, Sándor. 1984. Meaning and the changes of the morphological articulation of
 1595 words and constructions (synthetic and analytic types of changes in Hungarian). *Acta*

- 1596 *Linguistica Academiae Scientiarum Hungaricae* 34(1/2). 27–64.
- 1597 Kenesei, István, Robert Michael & Anna Fenyvesi. 1998. *Hungarian*. London:
1598 Routledge.
- 1599 Khačikjan, Margaret. 1998. *The Elamite language*. Rome: Consiglio nazionale delle
1600 ricerche.
- 1601 Kieviet, Paulus. 2017. *A grammar of Rapa Nui*. Berlin: Language Science Press.
- 1602 Kornfilt, Jaklin. 2006. *Turkish*. London/New York: Routledge.
- 1603 Lai, I-Wen. 2009. *Time in the Iquito language*. Austin: University of Texas PhD
1604 dissertation.
- 1605 Lakfioui, Mena B. & Vermondo Brugnatelli. 2020. Negation in Berber: Variation,
1606 evolution, typology. *Linguistics* 58(4). 967–1008.
- 1607 Lehmann, Christian. 2020. Univerbation. *Folia Linguistica* 54(s41). 205–252.
- 1608 Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference
1609 grammar*. Berkeley/Los Angeles: University of California Press.
- 1610 Liljegren, Henrik. 2016. *A grammar of Palula*. Berlin: Language Science Press.
- 1611 Mahdmoudveysi, Parvin, Denise Bailey, Ludwig Paul & Geoffrey Haig. 2012. *The
1612 Gorani language of Gawraǰū, a village of West Iran: Texts, grammar, and lexicon*.
1613 Wiesbaden: Ludwig Reichert.
- 1614 Mairal, Ricardo & Francisco Ruiz de Mendoza. 2008. Levels of description and
1615 explanation in meaning construction. In Christopher S. Butler & Javier Martín (ed.),
1616 *Deconstructing constructions*, 153–200. Amsterdam/Philadelphia: Benjamins.
- 1617 Maslova, Elena. 2003. *A grammar of Kolyma Yukaghir*. Berlin/New York: De Gruyter.
- 1618 Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing.
1619 *Linguistics* 36(2). 281–331.
- 1620 Mauri, Caterina. 2008. *Coordination relations in the languages of Europe and beyond*.
1621 Berlin/New York: De Gruyter.
- 1622 Miestamo, Matti. 2005. *Standard negation: The negation of declarative verbal main
1623 clauses in a typological perspective*. Berlin/New York: De Gruyter.
- 1624 Miestamo, Matti, Anne Tamm & Beáta Wagner-Nagy (eds.). 2015. *Negation in Uralic
1625 languages*. Amsterdam/Philadelphia: Benjamins.
- 1626 Miestamo, Matti, Dik Bakker & Antti Arppe. 2016. Sampling for variety. *Linguistic
1627 Typology* 20(2). 233–296.

- 1628 Mithun, Marianne. 1988. The grammaticization of coordination. In John Haiman &
 1629 Sandra A. Thompson (eds.), *Clause combining in grammar and discourse*, 331–360.
 1630 Amsterdam/Philadelphia: Benjamins.
- 1631 Munro, Pamela & Caitlin E. Gallagher. 2014. Garifuna negatives. In Lev Michael &
 1632 Tania Granadillo (eds.), *Negation in Arawak languages*, 13–53. Leiden/Boston: Brill.
- 1633 Musaev, Kenesbaj M. 2003. *Sintaksis karaimskogo jazyka*. Moscow: Akademija Nauk.
- 1634 Payne, John R. 1985a. Negation. In Timothy Shopen (ed.), *Language typology and*
 1635 *syntactic description*, Vol. 1, 197–242. Cambridge: Cambridge University Press.
- 1636 Payne, John R. 1985b. Complex phrases and complex sentences. In Timothy Shopen
 1637 (ed.), *Language typology and syntactic description*, Vol. 2, 3–41. Cambridge:
 1638 Cambridge University Press.
- 1639 Pearson, Karl. 1895. Notes on regression and inheritance in the case of two parents.
 1640 *Proceedings of the Royal Society of London* 58. 240–242.
- 1641 Perkins, Revere D. 1989. Statistical techniques for determining language sample size.
 1642 *Studies in Language* 13(2). 293–315.
- 1643 Peterson, John. 2011. *A grammar of Kharia: A South Munda language*. Leiden/Boston:
 1644 Brill.
- 1645 Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A*
 1646 *comprehensive grammar of the English language*. London/new York: Longman.
- 1647 Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2(3). 263–
 1648 314.
- 1649 Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter Kahrel. 1993. A method of
 1650 language sampling. *Studies in Language* 17(1). 169–203.
- 1651 Rongier, Jacques. 2004. *Parlons éwé: Langue du Togo*. Paris: L’Harmattan.
- 1652 Sakel, Jeanette. 2004. *A grammar of Mosestén*. Berlin/New York: De Gruyter.
- 1653 Salminen, Mikko B. 2017. *A grammar of Umbeyajts as spoken by the Ikojts people of*
 1654 *San Dionisio del Mar, Oaxaca, Mexico*. Cairns: James Cook University PhD
 1655 dissertation.
- 1656 Sánchez, Cristina. 2017. Coordination and correlatives. In Andreas Dufter & Elisabeth
 1657 Stark (eds.), *Manual of Romance morphosyntax and syntax*, 647–688. Berlin: De
 1658 Gruyter.
- 1659 Shimelman, Aviva. 2017. *A grammar of Yauyos Quechua*. Berlin: Language Science

- 1660 Press.
- 1661 Sneddon, James N. 1996. *Indonesian: A comprehensive grammar*. London/New York:
1662 Routledge.
- 1663 Stilo, Donald. 2004. Coordination in three West Iranian languages: Vafsi, Persian and
1664 Gilaki. In Martin Haspelmath (ed.), *Coordinating constructions*, 269–330.
1665 Amsterdam/Philadelphia: Benjamins.
- 1666 Stolz, Christel & Thomas Stolz. 1996. Funktionswortentlehnung in Mesoamerika:
1667 Spanisch-amerindischer Sprachkontakt (Hispanoindiana II). *Sprachtypologie und*
1668 *Universalienforschung* 49(1). 86–123.
- 1669 Tate, Robert F. 1954. Correlation between a discrete and a continuous variable: Point-
1670 biserial correlation. *The Annals of Mathematical Statistics* 25(3). 603–607.
- 1671 Theil, Henri. 1966. *Applied economic forecasting*. Amsterdam/Chicago: North Holland,
1672 Rand McNally.
- 1673 Van der Auwera, Johan. 2021. Quirky negative concord: Croatian, Spanish and French
1674 *ni*'s. *Jezikoslovlje* 22(1).
- 1675 Van der Auwera, Johan, Motoki Nomachi & Olga Krasnoukhova. 2021. Connective
1676 negation and negative concord in Balto-Slavic. In Peter Arkadiev, Jurgis Pakerys,
1677 Inesa Šeškauskienė & Vaiva Žeimantienė (eds.), *Studies in Baltic linguistics: A*
1678 *Festschrift for Axel Holvoet on the occasion of his 65th birthday*. Vilnius: Vilnius
1679 University Press.
- 1680 Voorhoeve, Clemens L. 1975. *Languages of Irian Jaya, checklist: Preliminary*
1681 *classification, language maps, wordlists*. Canberra: Australian National University.
- 1682 Werner, Roland. 1993. *Tidn-Áal: A study of Midob (Darfur-Nubian)*. Berlin: Dietrich
1683 Reimer.
- 1684 Wright, Raymond E. 1995. Logistic regression. In Laurence G. Grimm & Paul R.
1685 Yarnold (eds.), *Reading and understanding multivariate statistics*, 217–244.
1686 Washington: American Psychological Association.
- 1687 Yliniemi, Juha. 2019. *A descriptive grammar of Denjongke (Sikkimese Bhutia)*.
1688 Helsinki: University of Helsinki PhD dissertation.
- 1689 Zemp, Marius. 2018. *A grammar of Purik Tibetan*. Leiden: Brill.
- 1690
- 1691 **Non-standard abbreviations**

1692				
1693	ADD	additive	LK	linker
1694	ANR	action nominalizer	MOD	modal suffix
1695	AO1-4	actor-oriented affixes 1-4	NEGEX	negative existence
1696	AOR	aorist	NF	non-feminine
1697	AUG	augmentative	NONFACT	non-factual
1698	CL	gender-class marker	NPT	non-past tense
1699	CNJ	conjunction	PART	partitive
1700	CNTR	contrast	PER	personal
1701	CONT	contemporative mood	POT	potential mood
1702	DEP	dependent marker	PROP	proper article
1703	DST	distant past	RDP	reduplication
1704	DF	default case marker	REMP	remote past
1705	EC	extended current tense	RPST	recent past
1706	EMPH	emphatic	SS	same-subject marker
1707	ENC	emphatic negative coordination	SUB	subordinator
1708	ENCM	ENC marker	TAG2	interactive tag 2
1709	EX	existential	UO	undergoer voice
1710	GNR	general tense	V1	class 1 verb
1711	HAB	habitual	VIS	visual
1712	ITER	iterative		
1713				
1714	Appendix. Languages of the sample			

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Abun	kgr	Language isolate	Southeast Asia-Oceania	Berry & Berry (1999: 79)	No	—	—
Amarakaeri	amr	Harakmbut, Amarakaeri	South America	Chávez (1984: 327)	No	—	—
Alamblak	amp	Sepik, Sepik Hill	Australia-Papua New Guinea	Bruce (1988: 27)	No	—	—
Amdo Tibetan	adx	Sino-Tibetan, Bodic	Southeast Asia-Oceania	Dpal (2016: 530)	No	—	—
//Ani	hnh	Khoe-Kwadi, Khoe	Africa	Kilian-Hatz (2008: 299)	No	—	—
Antia-Komnzo-Wára-Wère-Kémä	tci	Yam, Morehead-Maró	Australia-Papua New Guinea	Döhler (2018: 104)	No	—	—
A'ou	aou	Tai-Kadai, Kadaic	Southeast Asia-Oceania	Li et al. (2014: 261)	No	—	—
Apalai	apy	Cariban, Apalai	South America	Koehn & Koehn (1986: 75)	No	—	—
Apinayé	apn	Nuclear-Macro-Je, Je	South America	Cunha de Oliveira (2005: 338)	No	—	—
Araweté	awt	Tupian, Maweti-Guarani	South America	Solano (2009: 252)	No	—	—
Arikapú	ark	Nuclear-Macro-Je, Jabuti	South America	Ribeiro (2008: 158)	No	—	—
Awa-Cuaiquer	kwi	Barbacoan, Unclas sified Barbacoan	South America	Cumow (1997: 303)	No	—	—
Ayoreo	ayo	Zamucoan, Zamuco-Ayoreo	South America	Bertinetto (2009: 48)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Bambara	bam	Mande, Western Mande	Africa	Maiga (2001: 72)	No	—	—
Bambassi	myf	Blue Nile Mao, Bambassi	Africa	Ahland (2012: 341)	No	—	—
Bargime	dba	Language isolate	Africa	Hantgan (2013: 315)	No	—	—
Bardi	bcj	iyulanNyulnyulan, Western	Australia-Papua New Guinea	Bowern (2012: 709-710)	No	—	—
Basque	eus	Language isolate	Eurasia	Hualde & Ortiz de Urbina (2003:)	Yes ($\alpha 1 + \alpha 2$)	No	Yes
Bauzi	bvz	Geelvink Bay, Bauzi	Australia-Papua New Guinea	Briley (1997: 85)	No	—	—
Bilua	blb	Language isolate	Southeast Asia-Oceania	Obata (2003: 234)	No	—	—
Bonan	peh	Mongolic-Khitian, Mongolic	Southeast Asia-Oceania	Fried (2010: 95)	No	—	—
Bora	boa	Boran, Boa	South America	Thiesen & Weber (2012: 527)	No	—	—
Bororo	bor	Bororoan, Bororo-Otuke	South America	Nonato (2008: 126, 131)	No	—	—
Brahui	brh	Dravidian, North Dravidian	Eurasia	Andronov (2006: 103)	Yes ($\alpha 1$)	Yes	No
Bullom So	buy	Atlantic-Congo, Mel	Africa	Childs (2011: 144)	No	—	—
Burmese	mya	Sino-Tibetan, Burmo-Qiangic	Southeast Asia-Oceania	Soe (1999: 144, 160-161)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Burushaski (Hunza, Nagar)	bsk	Language isolate	Eurasia	Berger (1998: 183)	No	—	—
Cao Miao	cov	Tai-Kadai, Kam-Tai	Southeast Asia-Oceania	Wu (2015: 299)	No	—	—
Cavineña	cav	Pano-Tacanan, Tacanan	South America	Guillaume (2008: 107)	No	—	—
Central Alaskan Yupik	esu	Eskimo-Aleut, Eskimo	North America-Mesoamerica	Miyaoka (2012: 1460)	No	—	—
Central Kalapuya	kyl	Kalapuyan, Central Kalapuya	North America-Mesoamerica	Banks (2007: 27)	No	—	—
Central Khmer	khm	Austroasiatic, Khmeric	Southeast Asia-Oceania	Haiman (2011: 90)	No	—	—
Central Nicobarese	ncb	Austroasiatic, Nicobaric	Southeast Asia-Oceania	Man (1889: 62)	No	—	—
Central Tunebo	tuf	Chibchan, Core Chibchan	South America	Headland (1994: 62)	No	—	—
Chimariko	cid	Language isolate	North America-Mesoamerica	Jany (2009: 177)	No	—	—
Chokwe	cjk	Atlantic-Congo, Volta-Congo	Africa	Martins (1990: 191)	No	—	—
Chukchi	ckt	Chukotko-Kamchatkan	Eurasia	Dunn (1999: 333)	No	—	—
Comaltepec Chinantec	cco	eanOtomanguean, Western	North America-Mesoamerica	Anderson (1989: 30)	Yes (β)	Yes	No (?)
Coptic (Sahidic)	cop	Afro-Asiatic, Egyptian	Africa	Layton (2000: 178)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Cuiba	cui	Guahiboan, Guahibo-Cuiva	South America	Kerr & Berg (2018: 264)	No	—	—
Dakota	dak	Siouan, Core Siouan	North America- Mesoamerica	Boas & Delloria (1941: 156)	No	—	—
Daga	dgz	Dagan, Daga	Australia-Papua New Guinea	Murane (1974: 229)	No	—	—
Dazaga	dzg	Saharan, Western Saharan	Africa	Walters (2016: 180)	No	—	—
Dieri	dif	Pama-Nyungan, Karnic	Australia-Papua New Guinea	Austin (1981: 208)	No	—	—
Dime	dim	South Omotic, Dime	Africa	Seyoum (2008: 195)	No	—	—
Duna	duc	Language isolate	Australia-Papua New Guinea	San Roque (2008: 255, 315)	No	—	—
Eastern Armenian	hye	Indo-European, Armenic	Eurasia	Dum-Tragut (2009: 289)	Yes (a1)	No	Yes
Eastern Kayah	eky	Sino-Tibetan, Karenic	Southeast Asia- Oceania	Solnit (1997: 227)	No	—	—
Elamite	elx	Language isolate	Eurasia	Khačikjan (1998: 55)	Yes (a3)	Yes	No (?)
Ewe	ewe	Atlantic-Congo, Volta-Congo	Africa	Rongier (2004: 176-177)	No	—	—
Fulniô	fun	Language isolate	South America	Lapenda (2005: 212)	No	—	—
Fwe	fwe	Atlantic-Congo, Volta-Congo	Africa	Gunnink (2018: 187)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Gaagudju	gbu	Language isolate	Australia-Papua New Guinea	Harvey (2002: 331)	No	—	—
Goemai	abk	Afro-Asiatic, Chadic	Africa	Hellwig (2011: 169)	No	—	—
Gooniyandi	gni	Bunaban, Gooniyandi	Australia-Papua New Guinea	McGregor (1990: 494, 530)	No	—	—
Guragone	gge	Maningrida, Burera-Nakara	Australia-Papua New Guinea	Green (1995: 124)	No	—	—
Gwich'in	gwi	Athabaskan-Eyak-Tlingit,	North America-Mesoamerica	McDonald (1972 [1911]: 76)	No	—	—
Halh Mongolian	khk	Mongolic-Khitan, Mongolic	Eurasia	Gaunt & Bayarmandakh (82)	No	—	—
Halkomelem	hur	Salishan, Central Salish	North America-Mesoamerica	Suttles (2004: 124)	No	—	—
Hatam	had	Hatam-Mansim, Hatam	Australia-Papua New Guinea	Reesink (1999: 109)	No	—	—
Hawaiian	haw	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Pukui & Elbert (1986: 488)	No	—	—
Highland Popolucan	poi	Mixe-Zoque, Zoque	North America-Mesoamerica	De Jong (2009: 920-921)	Yes ($\alpha 1$)	Yes	No
Hills Karbi	mjw	Sino-Tibetan, Kuki-Chin-Naga	Southeast Asia-Oceania	Konnerth (2014: 563)	No	—	—
Hokkaido Ainu	ain	Ainu, Hokkaido-Kuril Ainu	Eurasia	Refsing (1986: 163)	No	—	—
Huambisa	hub	Chicham, Shuaric	South America	Peña (2015: 747)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Huangascar-Topara-Yauyos	qux	Quechuan, Quechua I	South America	Shimelman (2017: 108)	Yes ($\alpha 1$)	Yes	No
Huastec	hus	Mayan, Huastecan Mayan	North America- Mesoamerica	Edmonson (1988: 545)	Yes ($\alpha 2$)	Yes	No
Hungarian	hun	Uralic, Hungarian	Eurasia	Bernini & Ramat (1996: 102)	Yes ($\alpha 2$)	Yes	Yes
Hup	jup	Naduhup, Eastern Naduhup	South America	Epps (2008: 736-737)	Yes ($\alpha 1$)	Yes	No
Hurrian	xhu	Hurro-Urartian, Hurrian	Eurasia	Wegner (2000: 162)	No	—	—
Icelandic	isl	Indo-European, Germanic	Eurasia	Einarsson (1949: 175)	Yes ($\alpha 1$)	Yes	No
Ik	ikx	Kuliak, Ik	Africa	Schrock (2014: 581)	No	—	—
Indonesian	ind	AsianAustronesian, Malayo-	Southeast Asia- Oceania	Sneddon (1996: 348)	No	—	—
Ingush	inh	akh Nakh-Daghestanian,	Eurasia	Nichols (2011: 528)	No	—	—
Iquito	iqu	Zaparoan, Iquito-Arabela	South America	Lai (2009: 208)	Yes (β)	Yes	Yes (?)
Jamiltepec Mixtec	mxt	eanOtomanguean, Eastern	North America- Mesoamerica	Johnson (1988: 127)	Yes ($\alpha 1/2 + \alpha 3$)	Yes	No
Japanese	jpn	Japonic, Japanese	Eurasia	Kaiser et al. (2013: 323)	Yes (β)	Yes	No (?)
Jaqaru	jqr	Aymaran, Tupe	South America	Hardman (2000: 82)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Jebero	jeb	Cahuapanan, Jebero	South America	Valenzuela (2014: 46, 839-840)	No	—	—
Jingultu	jig	Mirndi, Jingultu	Australia-Papua New Guinea	Pensalfini (1997: 201)	No	—	—
Kabardian	kbd	Abkhaz-Adyghe, Circassian	Eurasia	Kumakhov (2013: 254)	Yes ($\alpha 2$)	Yes	No (?)
Kako	kkj	Atlantic-Congo, Volta-Congo	Africa	Ernst (1992: 151)	No	—	—
Kakua	cbv	Kakua-Nukak, Kakua	South America	Bolaños (2016: 259)	No	—	—
Kalaallisut	kal	Eskimo-Aleut, Eskimo	North America-Mesoamerica	Fortescue (1984: 124)	Yes (β)	Yes	Yes
Kamayurá	kay	Tupian, Maweti-Guarani	South America	Seki (2000: 238)	No	—	—
Karaim	kdr	Turkic, Common Turkic	Eurasia	Musaev (2003: 164-165)	Yes ($\alpha 1$)	Yes	No
Karamojong	kdj	Nilotic, Eastern Nilotic	Africa	Bader (2008: 120)	Yes ($\alpha 1$)	Yes	No
Katcha-Kadugli-Miri	xtc	.Kadugli-Krongo, Central-Western	Africa	Turner (2016: 212)	No	—	—
Katukina-Kanamari	knm	inamariKatukinan, Katukina-	South America	Ishy de Magalhães (2018: 371)	No	—	—
Kayardild	gyd	Tangkic, Southern Tangkic	Australia-Papua New Guinea	Evans (1995: 395)	No	—	—
Kelabit	kzi	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Hemmings (2016: 610)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Ket	ket	Yeniseian, Northern	Eurasia	Dul'zon (1968: 578)	No	—	—
Kharia	khr	Austroasiatic, Mundaic	Eurasia	Peterson (2011: 341)	Yes (α1)	Yes	No
Khasi	kha	Austroasiatic, Khasi-Palaung	Eurasia	Nagaraja (2014b: 1181)	Yes (α1)	Yes	Yes (?)
Kija	gia	Jarrakan, Kija	Australia-Papua New Guinea	Kofod (1996: 121-126)	No	—	—
Kilmeri	kih	Border, Bewani	Australia-Papua New Guinea	Gerstner-Link (2018: 571)	No	—	—
Kiowa	kio	Kiowa-Tanoan, Kiowa	North America-Mesoamerica	Watkins (1984: 216)	No	—	—
Koasati	cku	Muskogean, Alabaman-Koasati	North America-Mesoamerica	Kimball (1985: 172)	No	—	—
Konai	kxw	East Strickland, Konai	Australia-Papua New Guinea	Årsjö (2016: 54)	No	—	—
Korafe-Yegha	kpr	er Nuclear Trans-Papuan New Guinea, er New Guinea,	Australia-Papua New Guinea	Farr (1999: 157)	No	—	—
Korean	kor	Koreanic, Korean	Eurasia	Yeon & Brown (2013: 136)	No	—	—
Koreguaje	coe	Tucanoan, Western Tucanoan	South America	Cook & Criswell (1993: 102)	No	—	—
Kotiria	gvc	Tucanoan, Eastern Tucanoan	South America	Stenzel (2004: 316)	No	—	—
Koyra Chiini Songhay	khq	onghay Songhay, Northwest	Africa	Heath (1999: 260-261)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Kuku-Yalanji	gvn	Pama-Nyungan, Yimidhir-Yalanji-	Australia-Papua New Guinea	Patz (2012: 113)	No	—	—
Kunama	kun	Language isolate	Africa	Idris (1987: 118)	No	—	—
Kuot	kto	Language isolate	Australia-Papua New Guinea	Lindström (2002: 221)	No	—	—
Kusunda	kgg	Language isolate	Eurasia	Watters (2006: 126)	No	—	—
Kutenai	kut	Language isolate	North America-Mesoamerica	Morgan (1991: 513-514)	No	—	—
Kwomtari	kwo	Kwomtari-Nai, Kwomtari	Australia-Papua New Guinea	Spencer (2008: 118)	No	—	—
Lao	lao	Tai-Kadai, Kam-Tai	Southeast Asia-Oceania	Enfield (2007: 54)	No	—	—
Laven	lbo	Austroasiatic, Bahnaric	Southeast Asia-Oceania	Jacq (2001: 283)	No	—	—
Lavukaleve	lvk	Language isolate	Southeast Asia-Oceania	Terrill (2003: 461)	No	—	—
Laz (Arhavi)	lzz	Kartvelian, Georgian-Zan	Eurasia	Lacroix (2009: 80)	No	—	—
Limilingan	lmc	Limilingan-Wulna, Limilingan	Australia-Papua New Guinea	Harvey (2001: 101)	No	—	—
Lowland Oaxaca Chontal	clo	al Tequistlatecan, Lowland Oaxaca	North America-Mesoamerica	O'Connor (2007: 57)	No	—	—
Lumun	lmd	Narrow Talodi, Lumun-Torona	Africa	Smits (2017: 718)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Maba (Chad)	mde	Maban, Mabang	Africa	Weiss (2009: 382)	No	—	—
Maco	wpc	Jodi-Saliban, Saliban	South America	Rosés (2015: 426)	No	—	—
Madi	jaa	Arawan, Madi-Madiha	South America	Dixon (2004: 167)	No	—	—
Madurese	mad	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Davies (2010: 399)	No	—	—
Maltese	mlt	Afro-Asiatic, Semitic	Eurasia	Sutcliffe (1936: 212)	Yes (α1)	Yes	No (?)
Mamaindé	wmd	Nambiquaran, Nambikwara C.	South America	Eberhard (2009: 581)	No	—	—
Mamvu	mdi	Central Sudanic, Membri-Mangbutu-	Africa	Blackings & Fabb (2003: 708)	No	—	—
Manambu	mle	Ndu, Nuclear Ndu	Australia-Papua New Guinea	Aikhenvald (2008: 190, 301, 305)	No	—	—
Mangarrayi-Maran	mpc	garrayiMangarrayi-Maran,	Australia-Papua New Guinea	Merlan (1989: 38)	No	—	—
Maore Comorian	swb	Atlantic-Congo, Volta-Congo	Africa	Alinet (2009: 262)	No	—	—
Marind	mrz	Anim, Marind-Boazi-Yaqai	Australia-Papua New Guinea	Olsson (2017: 115)	No	—	—
Marithiel	mfr	Western Daly, Bringen	Australia-Papua New Guinea	Green (1989: 159)	No	—	—
Mateq	xem	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Connell (2013: 152)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Mauwake	mhl	langNuclear Trans New Guinea,	Australia-Papua New Guinea	Berghäll (2015: 324)	No	—	—
Mapudungun	arn	Araucanian, Mapudungun	South America	Smeets (2005: 471)	No	—	—
Maquiritari	mch	Cariban, Guianan	South America	Cáceres (2011: 309)	No	—	—
Mbugu	mhd	Pidgins and creoles	Africa	Mous (2003: 133)	No	—	—
Menya	mcr	Angan, Nuclear Angan	Australia-Papua New Guinea	Whitehead (2004: 145, 171)	No	—	—
Meriam	ulk	Eastern Trans-Fly, Meriam	Australia-Papua New Guinea	Piper (1989: 201-202)	No	—	—
Midob	mei	Nubian, West-Central Nubian	Africa	Werner (1993: 63)	Yes (α2)	Yes	No (?)
Migabac	mpp	IuonNuclear Trans New Guinea,	Australia-Papua New Guinea	McEvoy (2008: 85)	No	—	—
Mískito	miq	Misumalpan, Mískito	North America-Mesoamerica	Salamanca (1988: 318)	No	—	—
Mosetén-Chimané	cas	Language isolate	South America	Sakel (2004: 332)	No	—	—
Moskona	mtj	East Bird's head, Meax	Australia-Papua New Guinea	Gravelle (2010: 503)	No	—	—
Mullukmulluk	mpb	Northern Daly, Mullukmulluk	Australia-Papua New Guinea	Lindsay et al. (2017: 51)	No	—	—
Murui Huitoto	huu	Huitotoan, Nuclear Witotoan	South America	Wojtylak (2017: 431)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Nangikurrunggurr	mwf	Southern Daly, Nangikurrunggurr	Australia-Papua New Guinea	Reid (1991: 349)	No	—	—
Nlɪŋg	ngh	Tuu, !Kwi	Africa	5) Collins & Namaseb (2011:	No	—	—
Neverver	lgk	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Barbour (2012: 418, 447)	No	—	—
Ngalkbun	ngk	Gunwinyguan, Marne	Australia-Papua New Guinea	Singer (2001: 71)	No	—	—
Ngizim	ngi	Afro-Asiatic, Chadic	Africa	Schuh (1972: 251, 462)	No	—	—
Nhanda	nha	1 Pama-Nyungan, South-West Pama-	Australia-Papua New Guinea	Blevins (2001: 105)	No	—	—
Nias	nia	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Brown (2001: 575)	No	—	—
Nigerian pidgin	pcm	Pidgins and creoles	Africa	Faracías (2005: 90)	No	—	—
Nihali	nll	Language isolate	Eurasia	Nagaraja (2014a: 78)	No	—	—
Nivaclé	cag	Matacoan, Mataguayo I	South America	Fabre (2016: 226)	No	—	—
Nivkh (Amur)	niv	Language isolate	Eurasia	42) Savel'eva & Taksami (1965:	Yes (α1)	Yes	No
North-Central Ju	knw	Kxa, Ju-Jung	Africa	Heine & König (2015: 276)	No	—	—
Northeast Kiwai (Urama)	kiw	Kiwaian, Northeast Kiwai	Australia-Papua New Guinea	Brown et al. (2016: 90)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
North Slavey	scs	Athabaskan-Eyak-Tlingit,	North America-Mesoamerica	Rice (1989: 1107)	No	—	—
Northern Emberá	emp	Chocoan, Embera	South America	Rojas (1987: 36)	No	—	—
Northern Haida	hdn	Haida, Northern Haida	North America-Mesoamerica	Enrico (2003: 287)	No	—	—
Northern Yokuts	yok	Yokutsan, General Yokuts	North America-Mesoamerica	Collord (1968: 92)	No	—	—
Northwest Gbaya	gya	Atlantic-Congo, Volta-Congo	Africa	Noss (1981: 122)	No	—	—
Northwest Sahaptin	yak	Sahaptian, Sahaptin	North America-Mesoamerica	Jansen (2010: 88)	No	—	—
Nuu-chah-nulth	myh	Wakashan, Wakashan, Southern	North America-Mesoamerica	Davidson (2002: 274)	No	—	—
Nzadi	nzd	Atlantic-Congo, Volta-Congo	Africa	Crane et al. (2011: 177)	No	—	—
Okanagan	oka	Salishan, Interior Salish	North America-Mesoamerica	Davis (2001: 85)	No	—	—
Oksapmin	opm	OkNuclear Trans New Guinea,	Australia-Papua New Guinea	Loughnane (2009: 392)	No	—	—
Ömie	aom	Koiarian, Baraic	Australia-Papua New Guinea	Austing & Upia (1975: 592)	No	—	—
Onondaga	ono	Iroquoian, Northern	North America-Mesoamerica	Woodbury (2018: 212)	No	—	—
Orya	ury	Tor-Orya, Orya	Australia-Papua New Guinea	Neukom & Patnaik (2003: 329)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Ottawa	otw	Algic, Algonquian	North America-Mesoamerica	Valentine (2001: 849)	No	—	—
Palikúr	plu	Arawakan, Palikúr	South America	Launey (2003: 201)	No	—	—
Palula	phl	Indo-European, Indo-Aryan	Eurasia	Lijegren (2016: 349-350)	Yes ($\alpha 1 + \alpha 2$)	Yes	Yes
Papiamentu	pap	Pidgins and creoles	South America	Goilo (2000: 20)	Yes ($\alpha 1$)	Yes	Yes
Papuan Malay	pmy	AsianAustronesian, Malayo-	Australia-Papua New Guinea	Kluge (2017: 520-521)	No	—	—
Pilagá	plg	Guaicuruan, Guaicuru del sur	South America	Vidal (2001: 284)	No	—	—
Pipil	ppl	Uto-Aztecan, Southern Uto-	North America-Mesoamerica	Campbell (1985: 124)	Yes ($\alpha 1$)	Yes	No
Pirahã	myp	Language isolate	South America	Everett (1983: 49-50)	No	—	—
Puinave	pui	Language isolate	South America	Gitrón (2008: 417)	No	—	—
Purepecha	pua	Tarascan, Purepecha	North America-Mesoamerica	Chamoreau (2003: 66)	No	—	—
Qawasqar	alc	Kawasqar, North-Central Alcalufan	South America	Aguilera (2001: 217)	No	—	—
Rapanui	rap	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Kieviet (2017: 503, 504, 506)	Yes ($\alpha 3$)	Yes	No (?)
Rotokas	roo	Oceanian, Bougainville, Papua New Guinea	Australia-Papua New Guinea	Robinson (2011: 86)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
San Blas Kuna	cuk	Chibchan, Core Chibchan	North America-Mesoamerica	Smith (2014: 193)	No	—	—
San Dionisio del Mar Huave	hve	uaveHuavean, San Dionisio-San	North America-Mesoamerica	Salminen (2017: 241)	Yes (β)	Yes	No (?)
Sandawe	sad	Language isolate	Africa	Steehan (2011: 238)	No	—	—
Sanumá	xsu	Yanomamic, Sanumá	South America	Borgman (1990: 86)	No	—	—
Savosavo	svs	Language isolate	Southeast Asia-Oceania	Wegner (2012: 102, 154)	No	—	—
Sawknah-Fogaha	swn	Afro-Asiatic, Berber	Africa	Lakfioui & Brugnattelli (2020: 317)	Yes (α1)	Yes	No
Semelai	sza	Austroasiatic, Aslian	Southeast Asia-Oceania	Kruspe (2004: 317)	No	—	—
Seri	sei	Language isolate	North America-Mesoamerica	Moser & Marlett (2010: 449)	No	—	—
Shasta	sht	Shastan, Shasta	North America-Mesoamerica	Silver (1966: 242)	No	—	—
Shatt	shj	Dajuc, Eastern Dajuc	Africa	Alfira et al. (2016: 211)	No	—	—
Sheko	she	Dizoid, Sheko	Africa	Hellenthal (2010: 425)	No	—	—
Sinhala	sin	Indo-European, Indo-Iranian	Eurasia	Chandralal (2010: 186)	Yes (α3)	No	Yes
Sirayaic	fos	Austronesian, East Formosan	Southeast Asia-Oceania	Adelaar (2011: 219)	Yes (α1)	Yes	Yes

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Siwai	siw	South Bougainville, Buin	Australia-Papua New Guinea	Onishi (1994: 161)	No	—	—
Skou	skv	Skou, Skou-Serrapio	Australia-Papua New Guinea	Donohue (2004: 144)	No	—	—
Somali	som	Afro-Asiatic, Cushitic	Africa	Saeed (1999: 257)	No	—	—
Southern-Coastal Tsimshian	tsi	South Tsimshian, Southern-Coastal	North America-Mesoamerica	Flaherty (1979: 179)	No	—	—
Southern Cook Island Maori	mri	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Nicholas (2016: 285)	No	—	—
Southern Yukaghir	yux	Yukaghir, Kolymic	Eurasia	Maslova (2003: 495-496)	Yes (α2)	Yes	No
Sumerian	sux	Language isolate	Eurasia	Zólyomi (2017: 149-150)	No	—	—
Swahili	swh	Atlantic-Congo, Volta-Congo	Africa	Mpiranya (2015: 193)	No	—	—
Tagalog	tgl	AsianAustronesian, Malayo-	Southeast Asia-Oceania	527) Schachter & Otanes (1972:	No	—	—
Tariana	tae	Arawakan, Japuracolumbia	South America	Aikhenvald (2003: 403)	No	—	—
Taulil	tuh	Taulil-Butam	Southeast Asia-Oceania	Meng (2018: 108)	No	—	—
Tehuelche	teh	Chonan Chonan, Continental	South America	Fernández (1998: 326)	No	—	—
Teiwa	twe	Timor-Alor-Pantar, Alor-Pantar	Southeast Asia-Oceania	Klamer (2010: 274)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Ternate	tft	North Halmahera, Northern North	Southeast Asia-Oceania	Hayami-Allen (2001: 180)	No	—	—
Ternateño	tmg	Pidgins and creoles	Southeast Asia-Oceania	Sippola (2011: 270)	Yes (α 1)	Yes	Yes (?)
Thai	tha	Tai-Kadai, Kam-Tai	Southeast Asia-Oceania	Smyth (2002: 143)	No	—	—
Thao	ssf	Asian Austronesian, Western Plains	Southeast Asia-Oceania	Wang (2004: 270)	No	—	—
Tipai	dih	Cochimi-Yuman	North America-Mesoamerica	Miller (2001: 187)	No	—	—
To'abaita	mlu	Asian Austronesian, Malayo-	Southeast Asia-Oceania	Lichtenberk (2008: 1021)	No	—	—
Tommo So Dogon	dto	Dogon, Dogon Escarpment	Africa	McPherson (2013: 443)	No	—	—
Toura (Côte d'Ivoire)	neb	Mande, Eastern Mande	Africa	Beath (1971: 280-281)	No	—	—
Tukang Besi South	bhq	Austronesian, Malayo-Polinesian	Southeast Asia-Oceania	Donohue (1999: 421-422)	No	—	—
Turkish	tur	Turkic, Common Turkic	Eurasia	Kornfilt (2006: 111)	Yes (α 1/3)	Yes	Yes (?)
Tuvalu	tvl	Asian Austronesian, Malayo-	Southeast Asia-Oceania	Besnier (2000: 186)	No	—	—
Udihe	ude	Tungusic, Central Tungusic	Eurasia	i7 Nikolaeva & Tolskaya (2001: 287)	No	—	—
Udmurt	udm	Uralic, Permian	Eurasia	Edygarova (2015: 287)	Yes (α 1)	Yes	No

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Urarina	ura	Language isolate	South America	Olawsky (2006: 555)	No	—	—
Ute-Southern Paiute	ute	Uto-Aztecan, Northern Uto-	North America-Mesoamerica	Givón (2011: 185)	No	—	—
Vaeakau-Taumako	piv	AsianAustronesian, Malayo-	Southeast Asia-Oceania	Næss & Hovdhaugen i-394	No	—	—
Wandala	mfi	Afro-Asiatic, Chadic	Africa	Frajzyngier (2012: 428)	No	—	—
Wappo	wao	Yuki-Wappo, Wappo	North America-Mesoamerica	Thompson et al. (1990: 23)	No	—	—
Warao	wba	Language isolate	South America	Romero-Figeroa (1997: 123)	No	—	—
Wardaman	wmb	Yangmanic, Wardaman	Australia-Papua New Guinea	Merlan (1994: 301)	No	—	—
Wari'	pav	Chapacuran, Moreic-Waric	South America	Everett & Kern (1997: 460)	No	—	—
Warlpiri	wbp	Pama-Nyungan, Desert Nyungic	Australia-Papua New Guinea	Bowler (2014: 139)	No	—	—
Watam	wax	Lower Sepik-Ramu, Ramu	Australia-Papua New Guinea	Foley (2010: 102)	No	—	—
Western Keres	kjq	Keresan, Western Keres	North America-Mesoamerica	Miller (1965: 123, 228-229)	No	—	—
Western Xiangxi Miao	mmr	Hmong-Mien, Hmongic	Southeast Asia-Oceania	Sposato (2015: 436, 572-573)	No	—	—
Wolaytta	wal	Ta-Ne-Omotic, Omoto	Africa	Wakasa (2008: 689)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Woleaian	woe	Asian-Austronesian, Malayo-	Southeast Asia-Oceania	Sohn (1975: 297)	No	—	—
Wolof	wol	African, Atlantic-Congo, North-Central	Africa	Diouf (2009: 107-108)	No	—	—
Xamtanga	xan	Afro-Asiatic, Cushitic	Africa	Teshome (2015: 268-269)	Yes (β)	No	Yes
Xicotepc De Juárez Totonac	tqt	Totonacan, Totonac	North America-Mesoamerica	Reid (1991: 78)	No	—	—
Xinca-Guazacapan	xin	Xinca, Xinca-Guazacapan	North America-Mesoamerica	Rogers (2010: 385)	No	—	—
Yakkha	ybh	Sino-Tibetan, Himalayish	Southeast Asia-Oceania	Schackow (2015: 448)	No	—	—
Yagua	yad	Peba-Yagua, Yagua	South America	Payne (1985: 247)	No	—	—
Yeri	yev	Nuclear Torricelli, West Palai	Australia-Papua New Guinea	Wilson (2017: 376, 513-514)	No	—	—
Yucatec Maya	yua	Mayan, Core Mayan	North America-Mesoamerica	Bolles & Bolles (2001: 66)	Yes (α)	Yes	Yes
Yuchi	yuc	Language isolate	North America-Mesoamerica	Linn (2001: 499)	No	—	—
Yue Chinese	yue	Sino-Tibetan, Sinitic	Southeast Asia-Oceania	Mathews & Yip (1994: 292)	No	—	—
Yuracaré	yuz	Language isolate	South America	Van Gijn (2006: 126)	No	—	—
Yurok	yur	Algic, Yurok	North America-Mesoamerica	Garrett (2010: 48)	No	—	—

Language name	ISO	Genealogical affiliation	Macroarea	Source	ENC available?	Distinct from SN?	Etymologically related to SN?
Yuwaalaraay-Gamilaraay	kld	anPama-Nyungan, Southeastern	Australia-Papua New Guinea	Giacon (2014: 265)	No	—	—
Zuni	zun	Language isolate	North America-Mesoamerica	Bunzel (1938: 400)	No	—	—

1735 *Author's address*
1736
1737 Iker Salaberri
1738 Department of Basque Language and Communication
1739 University of the Basque Country (UPV/EHU)
1740 Lehendakari Agirre, 83
1741 48015 Bilbao, Basque Autonomous Community
1742 Spain
1743 ikersalaberri@gmail.com
1744 <https://orcid.org/0000-0003-3859-4011>