Sterre Leufkens*

Measuring redundancy: the relation between concord and complexity

https://doi.org/10.1515/lingvan-2020-0143
Received December 22, 2020; accepted August 19, 2021; published online September 8, 2022

Abstract: Syntagmatic redundancy involves constructions in which a single meaning relates to more than one form in a clause or phrase. Redundancy has been claimed to facilitate language processing and acquisition, and to be a linguistic universal, because of these advantages. At the same time, the repeated expression of meaning has been argued to make languages more complex because of the violation of principles of economy and transparency. This paper first investigates the relation between redundancy and complexity by examining how redundancy scores on well-known metrics of complexity. This assessment shows that redundancy makes languages more complex according to metrics of objective complexity, while it has contradictory effects according to metrics of subjective complexity. To empirically test whether redundancy makes acquisition and processing easier or harder, a cross-linguistically valid way of measuring redundancy first needs to be established. The second part of the paper involves an attempt in this direction. It presents a typological study on cross-linguistic variation in four types of concord, showing that languages differ with respect to the number and types of redundancy constructions they allow. The ranking of the 50 sample languages as to their redundancy-based complexity serves as a testable hypothesis for acquisition and processing research.

Keywords: complexity; learnability; redundancy; typology

1 Introduction

Linguists have long been interested in the question of what makes languages complex. A phenomenon that is highly relevant in this matter is syntagmatic redundancy: multiple expressions of a single meaning within the same phrase or clause (the terminology follows Trudgill 2011). For example, in (1), person and number of the subject are expressed both by the subject pronoun and the verbal suffix.

(1) Dutch

Zij	*koop-t*	*een*	*appeltaart*
3SG.F	buy-3SG	an	apple_pie

'She is buying an apple pie.'

Of course, the pronoun offers more information than the verbal marker and is, therefore, not redundant in the literal sense of the word. The two elements do not express the exact same meaning: their meaning is merely overlapping and good reasons may exist why a speaker wants to use the pronoun in addition to the verbal marker.[1] This does not take away from the fact that sentences like (1) involve multiple manifestations of a single piece of information, which justifies their analysis as cases of redundancy.

Syntagmatic redundancy (henceforth: redundancy) is often seen as a fundamental characteristic and universal feature of languages. The omnipresence of redundancy is then accounted for by its communicative and cognitive advantages. For example, Trudgill (2009: 100) states: "All languages contain redundancy, which

1 Many linguists acknowledge that redundancy has in fact various communicative functions (e.g. Dahl 2004; Nichols 2009; Trudgill 2011; Leufkens 2015, 2020; Petré 2019).

***Corresponding author: Sterre Leufkens**, Utrecht Institute of Linguistics UiL OTS, Universiteit Utrecht, Utrecht, The Netherlands, E-mail: s.c.leufkens@uu.nl. https://orcid.org/0000-0002-2251-1309

seems to be necessary for successful communication, especially in less than perfect (i.e. natural) circumstances." Apart from strengthening the salience of utterances, redundancy has been claimed to facilitate processing and acquisition, as it increases the predictability and robustness of the sentence (Dye et al. 2017, 2018; Gibson et al. 2019; Pate and Goldwater 2015). Using such arguments, one could argue that redundancy makes languages simpler. At the same time, redundancy involves a repetition of information and, therefore, the use of forms that are strictly speaking unnecessary to convey a message. Moreover, redundancy violates the principle of transparency, defined in this paper as a one-to-one relation between meaning and form in a clause or phrase (Hengeveld and Leufkens 2018; Leufkens 2015).[2] Economy and transparency have been claimed to make languages simpler and easier to learn (e.g. Slobin 1980; Kusters 2003; Leufkens 2015), which entails that the violation of these principles by redundancy would make languages more complex. The current paper aims to address this apparent paradox by examining how redundancy increases or decreases complexity according to existing metrics.

The outcome of the investigation is a literature-based prediction with respect to the effect of redundancy on the learnability and processability of languages. Testing this prediction requires a comparison of languages with different degrees of redundancy. However, to my knowledge, no metrics have ever been proposed to quantify redundancy in a cross-linguistically valid way. In a first attempt at doing so, the second issue addressed in this paper is the typology of redundancy. Even if redundancy is a linguistic universal, this does not mean that there is no variation regarding the types of redundancy languages exhibit, or the extent to which they do so. In fact, redundancy comes in a multitude of shapes and sizes. One type, known as "user-level redundancy" (Dahl 2004: 11), involves accidental repetitions that occur in everyday speech, for example in a sentence like "I think … I think I like you". The current study will focus on another category of redundancy phenomena, called "system-level redundancy" (Dahl 2004: 11), which occurs when the grammar requires multiple expressions of the same information. Such grammatical redundancy can, for instance, take the form of agreement between arguments and verbs, NP agreement, negative concord, determiner doubling, and much more (see Barbiers 2008 for a non-exhaustive overview).

While some of the shapes of redundancy are indeed very common cross-linguistically, others are not (Hengeveld and Leufkens 2018). The fact that there is a great deal of cross-linguistic variation with regards to redundancy means that languages may also differ in the extent to which redundancy increases and/or decreases their complexity. This paper compares the complexity of 50 genetically diverse languages by measuring to what extent they allow for different types of redundancy. Investigating the complexity of redundancy in all its forms in 50 languages is not feasible within the scope of this paper; for this reason, I will focus on concord, that is, the expression of a single meaning by means of both a grammatical and a lexical item.

In the next section, concord phenomena will be introduced by means of examples from different languages. Section 3 investigates the relation between concord and existing metrics of complexity, thus establishing in what way redundancy contributes to the complexity of language. Section 4 presents the typological study of concord in a 50-language sample, and Section 5 concludes the paper.

2 Measuring concord

One of the many forms in which redundancy may appear is concord: the expression of a single meaning by means of both a lexical and a grammatical element. Concord, too, comes in many shapes and sizes, some of

2 This paper looks at transparency from a syntagmatic point of view, which means it is defined as the relation between meaning and form in a phrase or clause. As an anonymous reviewer rightly points out, it could also be defined from a paradigmatic point of view. From that perspective, each individual marker is transparent as long as it has a single function in a paradigm (barring allomorphy and syncretism). Even though the transparency (and redundancy) of paradigms is surely relevant in measuring complexity, this paper will take only syntagmatic transparency and redundancy as its object of study for reasons of scope.

which are commonly known in linguistics as 'agreement'. In this paper, the term 'concord' will be used consistently, because not all of the phenomena discussed below would be considered agreement. For example, plural concord and temporal concord (see below for definitions and examples) do not appear in Corbett's (2006) standard work on agreement.

I will investigate four types of concord. The first is argument concord, that is, the expression of argument properties by means of both a (semi-)lexical element (such as a free pronoun) and a grammatical element (typically verbal inflection). Argument concord is illustrated by example (2).

(2) Abkhaz
(sarà) *(barà)* *(yarà)* *ø-bə-s-te-yt'*
1SG 2SG.F 3SG 3SG-2SG.IO-1SG-give-PST.PFV
'I gave it to you.'
(Hewitt 1979: 155)

If a language lacks grammatical expression of argument information, it could never exhibit argument concord. I will return to this issue in the description of the typological study in Section 4.

The second type of concord is negative concord, which combines the use of a lexical negator, such as a negative quantifier, with a grammatical negator. As illustrated by example (3), the two negative forms relate to a single semantic negation.

(3) Hungarian
itt *senki* *sem* *beszél* *magyarul*
here nobody not speak Hungarian
'No one speaks Hungarian here.'
(Rounds 2009: 130)

Negative concord should be distinguished from so-called double negation, which is shown in example (4). In the case of double negation, the separate negative elements each relate to a semantic negation. The semantic negations cancel each other out, resulting in an affirmative interpretation. Double negation does not involve concord, because there are no multiple formal manifestations relating to a single semantic negation.

(4) Dutch
niemand *wordt* *niet* *geraakt* *door* *deze* *film*
nobody becomes not touched by this movie
'Nobody is not touched by this movie.' (i.e. everybody is touched by this movie)
(based on Zeijlstra 2004: 59)

A third type of concord to be investigated is plural concord: the combination of a nominal plural marker in combination with a >1 numeral or quantifier, as illustrated in example (5).

(5) Pipil
ne *ye:y* *pipu-tsi-tsín*
the 3 boy-PL-DIM
'the three boys'
(Campbell 1995: 104)

Some languages avoid plural concord. For example, in Turkish, the nominal plural marker is not overtly expressed in the presence of a >1 numeral (Rijkhoff 2002: 43). This should be distinguished from languages that do not exhibit plural concord because they lack nominal plural marking altogether, for example classifier languages.

The fourth type of concord studied in this paper is temporal concord, which involves the combination of a tense marker with a lexical time expression such as a temporal adverb, as in example (6).

(6) Korean
Mia ka ecey Mikwuk ulo ttena-ss-e.yo
Mia NOM yesterday America to leave-PST-POL
'Mia left for America yesterday.'
(Sohn 1999: 362)

Concord phenomena offer a well-suited test case for the cross-linguistic study of redundancy, because they are cross-linguistically frequent and relatively well-described, enabling the collection of sufficient typological data. Moreover, the four types of concord presented here are similar enough to be comparable, but different enough to showcase the diverse relations that exist between redundancy and complexity, which will be the topic of the following section.

3 How complex is redundancy?

At the risk of lapsing into cliché: measuring the complexity of a language is a complex endeavor, for one thing because there are so many different sides to complexity. A language can be complex for a learner (L1 or L2, child or adult), it can be complex to use for a speaker, to interpret and process for a hearer, or to describe for the grammarian. Another complicating factor is that for all these different aspects of complexity, a range of metrics has been proposed. As yet, it is unclear how different metrics of complexity relate to each other. Are they measuring the same thing? Should different complexity metrics be combined to determine the complexity of a domain of a language, or a language as a whole? If so, how should they be weighted relative to each other? This paper aims to take a small step in answering such questions, by relating the notion of redundancy to some well-known and often used metrics of complexity, thus analyzing how redundancy contributes to linguistic complexity. First, I will address metrics of objective complexity (also referred to as absolute, system-based, or theory-oriented complexity), which aim to measure the complexity of the language system in terms of the number of forms and rules that it employs. Second, I will relate redundancy to metrics of subjective complexity (also known as relative or user-oriented complexity), i.e. the difficulty of language processing and acquisition (Miestamo 2008).

Metrics of objective complexity have been proposed by, among others, Dahl (2004), Hawkins (2004), McWhorter (2001, 2007, 2011), Nichols (2009), Gil (2009), Pallotti (2015), Anderson (2015), Audring (2017) and Seinhorst (2021). There are fundamental differences between these metrics, but what they have in common is that they count the number of overt forms in sentences or paradigms, and/or the number and length of rules required to describe those sentences or paradigms. The underlying axiom is that a language or sentence is more complex the more linguistic material (features, phonemes, morphemes, overly expressed semantic or pragmatic categories, rules, principles, constraints, etc.) it exhibits. Syntagmatic redundancy as defined in this paper involves multiple forms relating to a single meaning, which means that, by definition, it increases the number of forms in a sentence. Therefore, all four types of concord investigated increase complexity according to this type of complexity metrics.

A different way of quantifying complexity involves the measurement of aspects of language use rather than the language system. In this category of metrics, the complexity of language is sometimes assessed in terms of the difficulty of language processing, for example by measuring response times, response accuracy to probe questions, or performance on lexical decision tasks (e.g. Gibson 1998; Jaeger and Tily 2011). Another type of subjective complexity metrics aims to assess the difficulty of language acquisition, i.e. the learnability of languages. Metrics of this type commonly take L2 acquisition as their basis for measuring complexity (Bulté and Housen 2012; Housen and Simoens 2016; Kusters 2003; Szmrecsanyi and Kortmann 2009; Trudgill 2011). To my knowledge, no one defines complexity as L1 difficulty, presumably because children all seem to acquire their native languages with seemingly equal effort (or effortlessness). There are attempts, however, to establish which linguistic features are relatively hard for children to acquire, for example Slobin (1973, 1980).

Many subjective complexity metrics take into account the principles of economy and transparency. For example, Kusters (2003) states that a language that morphologically expresses fewer semantic categories is relatively simple in the sense that it is easier to acquire for an L2 learner than a language with more elaborate morphology. He also argues that a language is simpler the more its morphology maintains a one-to-one relation between meaning and form. Trudgill (2011: 41) adopts the same view when he states that "[p]ost-threshold learners have less difficulty in coping with regularity and transparency than irregularity and opacity; and loss of redundancy reduces the burden for learner speakers." Hence, in this category of metrics, redundancy makes languages more complex as it violates both principles.

At the same time, it has been demonstrated that redundancy facilitates processing, because it increases predictability and robustness of the linguistic signal (Dye et al. 2017, 2018; Gibson et al. 2019; Pate and Goldwater 2015). Moreover, psycholinguistic studies show that redundancy facilitates language learning. For example, in an artificial language learning experiment, Taraban (2004) shows that case and nominal gender are easier for adults to acquire if morphological marking is accompanied by consistent phonological cues providing the same information. Audring (2014), too, points out that gender agreement facilitates acquisition of grammatical gender, basing her argument on diachronic data. She shows that complex gender systems only survive if there is sufficient marking:[3] Only in that case, learners receive enough clues to acquire the gender of each noun. In sum, there is evidence from multiple data sources that redundancy simplifies languages according to measures of subjective complexity.

Thus, the literature on subjective complexity presents us with a paradox: redundancy is argued to simplify languages by making processing and acquisition easier, but at the same time it is said to complexify languages by violating the principles of economy and transparency. However, the paradox disappears once we distinguish between the learnability of the redundantly expressed feature and the learnability of redundancy itself. Let us turn to gender agreement again for an explanation. Gender agreement as such may be difficult to learn: the learner has to know on which elements gender has to be marked, which forms to use, whether it is optional or obligatory, et cetera. At the same time, as we have seen above, gender agreement helps learners to acquire the gender of each noun. Hence, redundant marking of gender may make the language less learnable because it is difficult to acquire agreement rules, but at the same time more learnable because it becomes easier to acquire the gender feature of nouns. The same could hold for processing: while redundant marking of, for example, gender makes an upcoming noun more predictable and therefore facilitates processing (Gibson et al. 2019), redundancy may at the same time impede processing simply because there are more forms to be processed. In more general terms, redundancy has a facilitating effect on acquisition and processing of specific features, but at the same time an inhibitory effect on the learnability of clausal structure and the processability of clauses.

To my knowledge, no research exists that acknowledges the conflict between the simplifying and complexifying forces of redundancy on language processing and acquisition. If we want to weight these forces relative to one another and evaluate whether redundancy makes language easier or harder, we need processing and acquisition research that compares the processability and learnability of languages with different degrees of redundancy. Such research requires some way of measuring redundancy cross-linguistically. In the next section, I will make a first attempt to quantify the extent to which languages allow for a type of redundancy, namely concord. This will result in a ranking of languages as to their degree of redundancy. This makes it possible to test whether languages that rank high, that is, languages with a high degree of redundancy, are relatively easy or difficult to learn and process.

3 Note that sufficient marking does not necessarily involve redundant marking. The idea is that the gender of a noun becomes more learnable the more often it is expressed, regardless of the location of that marking (i.e. on controller, target(s), or both). According to the definition in this paper, gender is redundantly marked when it is expressed overtly multiple times. This means that 'sufficient marking' and 'redundant marking' are not the same. Still, the generalization holds that redundant marking (of gender or any other information) by definition provides the learner with more clues than non-redundant marking.

4 Typology of concord

Instances of redundancy are found in all natural languages, but there is a great deal of variation as to the types of redundancy that languages exhibit and the extent to which they do so. For example, while subject-verb agreement is cross-linguistically a very common phenomenon, determiner doubling is not. Typological variation in redundancy is relevant in measuring complexity, because different forms of redundancy may contribute to complexity in a different way and to a different extent.

In order to map the extent to which the four types of concord are attested in the languages of the world, I carried out an explorative typological study on concord in 50 languages. The language sample used in this study is copied from Rijkhoff (2002), who conducted a typological study of the noun phrase that includes a cross-linguistic investigation of how numerals combine with plural marking. Thus, copying Rijkhoff's sample had as an advantage that the data on plural concord had already been collected. Rijkhoff's sample is drawn by means of the Diversity Value technique (Rijkhoff and Bakker 1998; Rijkhoff et al. 1993), a sampling strategy that quantifies the degree of internal variation of each language family. By sampling languages from families proportionate to their Diversity Value, the strategy ensures that a sample incorporates the highest degree of linguistic variation. Since the current study aims to map out linguistic variation in redundancy, the sample was appropriate for this study. The complete sample, including genetic affiliation and ISO-codes, is provided in Appendix A below.

Data on the exponence of argument concord, negative concord, and temporal concord in the sample languages has been collected primarily from reference grammars and other descriptive sources. If evidence for a type of concord could not be found in descriptive literature, an expert or native speaker of the language was consulted if possible. In some cases, neither grammatical descriptions nor experts were available within the time frame of the study, so that no data could be recorded. The complete set of codings is provided in Appendix B below.

In the sample languages, a type of concord was considered attested if it was possible to combine a lexical expression and a grammatical expression of a single meaning in a grammatical clause or phrase. Occurrence was measured as a binary feature, meaning that syntactic restrictions to the use of concord have not been considered. A type of concord was regarded as non-existent in a language if there was evidence that combining the lexical and the grammatical element was ungrammatical, or, in the case of double negation, would not result in an overlap of meaning (see Section 2). Some languages did not exhibit a type of concord because one of the constitutive elements was lacking. For example, in languages without pronominal marking other than pronouns, argument concord is simply impossible. In the same way, temporal concord cannot exist if there is no tense marking, there is no plural concord in a language that does not possess nominal plural marking (e.g. in languages with classifiers), and negative concord cannot possibly arise if a language lacks negative lexical items altogether. Such languages have been counted separately from languages in which a type of concord was possible, but not allowed.

It should be noted that redundancy has its effect on objective or subjective complexity only to the extent that it actually occurs: if a language permits concord optionally, but speakers do not make use of the option, the effect is lost. For example, argument concord can only increase predictability and facilitate processing if both the lexical and grammatical item are overtly expressed: In the case of pro-drop, there will presumably not be a facilitatory effect. To quantify the extent to which redundancy complexifies or simplifies language use, we should not only test the grammatical possibility of redundant constructions, but measure their occurrence in actual sentences by means of a cross-linguistic corpus study. This would be a worthwhile endeavor, but one that lies outside the scope of this paper.

The data show that there is indeed variation as to the types of concord that languages exhibit. Argument concord is attested in 35 of the 48 languages (73%) on which data could be collected, and all languages that possess the constitutive elements of argument concord allow them to co-occur. Similarly, temporal concord is attested in 25 of 32 languages (78%), and languages that possess tense marking always allow it to co-occur with

lexical temporal elements. Negative concord is attested in seven of the 28 languages (25%), so that it appears much less common than argument and temporal concord. However, if we only take languages into consideration that possess the constitutive elements for negative concord, the proportion of concord languages is larger: seven out of eight languages (88%) exhibit this type of concord. Finally, plural concord is allowed in 18 out of 50 languages (36%), and considering the languages with the constitutive elements only, it is allowed in 18 out of 42 languages (43%) – a much smaller proportion than any of the other concord types.

The data in Table 1 show that languages always allow argument concord and temporal concord: if the constitutive elements are there, they can be combined in one clause. Languages also allow for negative concord most of the time, but since the constitutive parts of negative concord are relatively infrequent, negative concord is less frequent in the sample. The numbers paint a truly different picture for plural concord, as it is the only kind of concord that is often not allowed, even if both constitutive elements exist. The reason for the low frequency of plural concord in the sample may be that nominal plural marking is often optional in comparison to the marking of sentential negation, tense, and pronominal marking on the verb. This means that in the case of plurality, speakers have the choice to drop one of the markers, while in the other cases of concord their grammar obliges them to be redundant (see Leufkens 2020 for a more thorough discussion of this point). In any case, the data in Table 1 prove that there is cross-linguistic variation regarding types of redundancy, even if we look at a relatively small subset of redundancy phenomena.

In order to evaluate how redundancy contributes to the complexity of individual languages, it is interesting to look at the dispersion of concord types per language. Table 2 provides an overview, displaying the 16 possible combinations of the four concord types. The 50 sample languages have been categorized in terms of these 16 categories. Languages for which one or more data points were missing have been excluded, as it was impossible to categorize them. In Table 2, languages in which a type of concord was not possible because one of the constitutive elements was lacking, have been taken together with languages disallowing that type of concord.

Table 2 shows once more that there is cross-linguistic variation in languages, this time with respect to the number and combination of concord types they exhibit. Regarding the number of concord types, Abkhaz ranks highest, as it is the only language in the sample possessing all types of concord. A group of nine languages exhibits three out of the four types of concord, while eight languages possess two types. Three languages have only one type of concord, and Mandarin and Gude rank lowest by allowing no type of concord at all. This ranking of languages in terms of the number of concord constructions reflects a ranking of objective complexity, Abkhaz being the most complex and Mandarin and Gude being the simplest.[4] Regarding subjective complexity, the ranking from Table 2 provides testable hypotheses for acquisition and processing research. If it is the case that redundancy makes languages harder to acquire and process, Abkhaz should be

Table 1: Four types of concord in 50-language sample.

	Argument concord	Temporal concord	Negative concord	Plural concord
Total number of languages for which data could be found	48	32	28	50
Concord not possible for lack of constitutive element	13	7	20	8
Concord not allowed	0	0	1	24
Concord allowed	35	25	7	18

4 Note that this only concerns the complexity of concord phenomena – taking into account other redundancy phenomena or other complexity-inducing phenomena could change the language ranking presented in Table 2.

Table 2: Distribution of types of concord over languages.

	Argument concord	Temporal concord	Negative concord	Plural concord	Languages of this type
Type 1	+	+	+	+	Abkhaz (1)
Type 2	+	+	+	–	Hungarian, Turkish (2)
Type 3	+	+	–	+	Dutch, Burushaski, Pipil, Quechua, Sarcee, Tamil, West-Greenlandic (7)
Type 4	+	–	+	+	
Type 5	–	+	+	+	
Type 6	+	–	–	+	
Type 7	+	–	+	–	
Type 8	+	+	–	–	Basque, Cayuga, Chukchi, Samoan (4)
Type 9	–	+	–	+	Babungo, Kisi (2)
Type 10	–	+	+	–	Bambara, Kayardild (2)
Type 11	–	–	+	+	
Type 12	+	–	–	–	Lango (1)
Type 13	–	+	–	–	Korean, Vietnamese (2)
Type 14	–	–	+	–	
Type 15	–	–	–	+	
Type 16	–	–	–	–	Mandarin, Gude (2)

the most difficult language. If, however, redundancy simplifies acquisition and processing, Mandarin and Gude would prove to be the hardest languages of the sample.

5 Conclusions

This paper has shown how redundancy is related to common metrics of complexity. I have focused on a particular category of redundant phenomena, namely concord, which involves the combination of a lexical and a grammatical element relating to a single meaning within one phrase or clause. I have demonstrated that all types of concord increase complexity according to metrics that assess objective complexity, by counting linguistic forms and rules. According to metrics of subjective complexity, measuring ease of processing and acquisition, redundancy could be argued to make language harder, as it violates transparency and economy, but at the same time simpler, as it increases predictability and robustness.

Comparative studies of language acquisition and processing are required to test whether redundancy increases or decreases subjective complexity. To make such studies possible, a way of quantifying redundancy cross-linguistically is needed. A typological study of concord, presented in the second part of the paper, is a first attempt to rank languages on their degree of redundancy. The study has shown that even though all languages exhibit some type of redundancy, there is large cross-linguistic variation as to the types and number of concord constructions that languages possess. Languages have been categorized over 16 language types according to the types of concord they exhibited, reflecting different degrees of complexity in the objective and subjective sense. Future research will need to determine how these degrees of redundancy relate to difficulty of processing and acquisition of the languages involved.

Acknowledgments: I would like to thank Klaas Seinhorst, as well as the editors and two anonymous reviewers, for their constructive and insightful feedback on earlier versions of this paper.

Appendix A

Sample language	ISO 639-3	Language family	
		Ruhlen (1991)	Eberhard et al. (2020)
Abkhaz	abk	Caucasian	Abkhaz-Adyghe
Alamblak	amp	Indo-Pacific (Sepik-Ramu)	Sepik
Babungo (Vengo)	bav	Niger-Kordofanian (Niger-Congo, Niger-Congo Proper, Central Niger-Congo)	Niger-Congo
Bambara (Bamanankan)	bam	Niger-Kordofanian (Niger-Congo, Mande)	Niger-Congo
Basque	eus	Basque	Basque
Berbice Dutch Creole (Berbice Creole Dutch)	brc	Pidgins and Creoles	Creole
Bukiyip	ape	Indo-Pacific (Torricelli)	Torricelli
Burmese	mya	Sino-Tibetan (Tibeto-Karen)	Sino-Tibetan
Burushaski	bsk	Burushaski	Burushaski
Cayuga	cay	Amerind (Northern Amerind, Almosan-Keresiouan)	Iroquoian
Chukchi	ckt	Chukchi-Kamchatkan	Chukotko-Kamchatkan
Dutch	nld	Indo-Hittite (Indo-European)	Indo-European
Galela	gbi	Indo-Pacific (West-Papuan)	West Papuan
Georgian	kat	Kartvelian	Kartvelian
Gilyak (Nivkh)	niv	Gilyak	Gilyak
Guarani (Guaraní, Paraguayan)	gug	Amerind (Equatorial-Tucanoan)	Tupian
Gude	gde	Afro-Asiatic (Chadic)	Afro-Asiatic
Hittite	hit	Indo-Hittite (Anatolian)	Not listed
Hixkaryána	hix	Amerind (Ge-Pano-Carib)	Cariban
Hmong Njua	hnj	Austric (Miao-Yao)	Hmong-Mien
Hungarian	hun	Uralic-Yukaghir	Uralic
Hurrian	xhu	Hurrian	Not listed
Ika (Arhuaco)	arh	Amerind (Chibcan-Paezan)	Chibchan
Kayardild	gyd	Australian (Pama-Nyungan)	Australian
Ket	ket	Ket	Yeniseian
Kisi (Kisi, Southern)	kss	Niger-Kordofanian (Niger-Congo, Niger-Congo Proper, West Atlantic)	Niger-Congo
Koasati	cku	Amerind (Northern Amerind, Penutian)	Muskogean
Korean	kor	Korean-Japanese-Ainu	Koreanic
Krongo	kgo	Niger-Kordofanian (Kordofanian)	Nilo-Saharan
Lango	laj	Nilo-Saharan (East Sudanic)	Nilo-Saharan
Mandarin	cmn	Sino-Tibetan (Sinitic)	Sino-Tibetan
Nahali (Nihali)	nll	Nahali	Nihali
Nama Hottentot (Khoekhoe)	naq	Khoisan	Khoe-Kwadi
Nasioi (Naasioi)	nas	Indo Pacific (East Papuan)	South Bougainville
Ngalakan (Ngalakgan)	nig	Australian (Gunwinygun)	Australian
Ngiti	niy	Nilo-Saharan (Central Sudanic)	Nilo-Saharan
Nung	nut	Austric (Austro-Tai, Daic)	Kra-Dai
Nunggubuyu (Wubuy)		Australian (Nunggubuyu)	Australian
Oromo (Oromo, Borana-Arsi-Guji)	gax	Afro-Asiatic (Cushitic)	Afro-Asiatic
Pipil (Nahuat)	ppl	Amerind (Central Amerind)	Uto-Aztecan
Quechua, Imbabura (Quichua, Imbabura Highland)	qvi	Amerind (Andean)	Quechuan
Samoan	smo	Austric (Austro-Tai, Austronesian, Malayo-Polynesian)	Austronesian
Sarcee (Sarsi)	srs	Sarcee (Na-Dene)	Eyak-Athabaskan
Sumerian	sux	Sumerian	Not listed
Tamil	tam	Elamo-Dravidian	Dravidian

(continued)

Sample language	ISO 639-3	Language family	
		Ruhlen (1991)	Eberhard et al. (2020)
Tsou	tsu	Austric (Austro-Tai, Austronesian, Tsouic)	Austronesian
Turkish	tur	Altaic	Turkic
Vietnamese	vie	Austric (Austroasiatic)	Austro-Asiatic
Wambon	wms	Indo-Pacific (Trans-New Guinea)	Trans-New Guinea
West Greenlandic (Greenlandic)	kal	Eskimo-Aleut	Eskimo-Aleut

Appendix B

NA = Not Applicable; one of the constitutive elements of this type of concord is not available in the language.
ND = No Data; no sufficient data were available to determine whether this type of concord is allowed in the language.
+ = This type of concord is possible in the language.
– = This type of concord is ungrammatical in the language, or, in the case of negative concord, combined negation elements do not lead to negative concord.

Language	Argument concord	Temporal concord	Negative concord	Plural concord
Abkhaz	+	+	+	+
Alamblak	+	ND	ND	+
Babungo (Vengo)	NA	+	NA	+
Bambara (Bamanankan)	NA	+	+	–
Basque	+	+	NA	–
Berbice Dutch Creole	NA	ND	+	–
Bukiyip	+	ND	ND	+
Burmese	NA	+	ND	NA
Burushaski	+	+	NA	+
Cayuga	+	+	NA	NA
Chukchi	+	+	NA	NA
Dutch	+	+	–	+
Galela	+	NA	ND	NA
Georgian	+	ND	+	–
Guaraní	+	ND	ND	–
Gude	NA	NA	NA	–
Hittite	ND	ND	ND	+
Hixkaryana	+	+	ND	NA
Hmong Njua	ND	ND	ND	–
Hungarian	+	+	+	–
Hurrian	+	ND	ND	–
Ika	+	ND	ND	–
Kayardild	NA	+	+	–
Ket	+	ND	NA	+
Kisi	NA	+	NA	+
Koasati	+	ND	NA	–
Korean	NA	+	NA	–
Krongo	+	ND	ND	+
Lango	+	NA	NA	–
Mandarin	NA	NA	NA	NA
Nahali	NA	ND	ND	–

(continued)

Language	Argument concord	Temporal concord	Negative concord	Plural concord
Nama Hottentot (Khoekhoe)	NA	ND	ND	+
Nasioi (Naasioi)	+	ND	ND	+
Ngalakan	+	ND	NA	–
Ngiti	+	+	ND	+
Nivkh (Gilyak)	+	NA	ND	–
Nung	NA	NA	ND	NA
Nunggubuyu (Wubuy)	+	ND	ND	–
Oromo (Borana-Arsi-Guji Oromo)	+	+	ND	–
Pipil	+	+	NA	+
Quechua, Imbabura	+	+	NA	+
Samoan	+	+	NA	–
Sarcee (Sarsi)	+	+	NA	+
Sumerian	+	ND	ND	–
Tamil	+	+	NA	+
Tsou	+	NA	ND	–
Turkish	+	+	+	–
Vietnamese	NA	+	NA	NA
Wambon	+	+	ND	–
West Greenlandic	+	+	NA	+

References

Anderson, Stephen. 2015. Dimensions of morphological complexity. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds.), *Understanding and measuring morphological complexity*, 11–26. Oxford: Oxford University Press.

Audring, Jenny. 2014. Gender as a complex feature. *Language Sciences* 43. 5–17.

Audring, Jenny. 2017. Calibrating complexity: How complex is a gender system? *Language Sciences* 60. 53–68.

Barbiers, Sjef. 2008. Microvariation in syntactic doubling. An introduction. In Sjef Barbiers, Olaf Koeneman, Marika Lekakou & Margreet van der Ham (eds.), *Microvariation in syntactic doubling*, 1–34. Bingley: Emerald.

Bulté, Bram & Alex Housen. 2012. Defining and operationalising L2 complexity. In Alex Housen, Folkert Kuiken & Ineke Vedder (eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 21–46. Amsterdam: John Benjamins Publishing Company.

Campbell, Lyle. 1995. *The Pipil language of El Salvador*. Amsterdam: Mouton de Gruyter.

Corbett, Greville C. 2006. *Agreement*. Cambridge: Cambridge University Press.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins Publishing Company.

Dye, Melody, Petar Milin, Richard Futrell & Michael Ramscar. 2017. A functional theory of gender paradigms. In Ferenc Kiefer, James Blevins & Huba Bartos (eds.), *Perspectives on morphological structure: Data and analyses*, 212–239. Leiden: Brill.

Dye, Melody, Petar Milin, Richard Futrell & Michael Ramscar. 2018. Alternative solutions to a language design problem: The role of adjectives and gender marking in efficient communication. *Topics in Cognitive Science* 10(1). 209–224.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the world*, 23rd edn. Dallas, TX: SIL International. https://ethnologue.com (accessed 13 October 2020).

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76.

Gibson, Edward, Richard Futrell, Steven P. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–405.

Gil, David. 2009. How much grammar does it take to sail a boat? In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 19–33. Oxford: Oxford University Press.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hengeveld, Kees & Sterre Leufkens. 2018. Transparent and non-transparent languages. *Folia Linguistica* 52(1). 139–175.

Hewitt, B. George. 1979. *Abkhaz*. Amsterdam: North-Holland Publishing Company.

Housen, Alex & Hannelore Simoens (eds.). 2016. Cognitive perspectives on difficulty and complexity in SLA [Special issue]. *Studies in Language Acquisition* 38(2). 163–175.

Jaeger, T. Florian & Harry Tily. 2011. On language 'utility': Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(3). 323–335.
Kusters, Wouter. 2003. *Linguistic complexity*. Amsterdam: University of Amsterdam PhD Dissertation.
Leufkens, Sterre. 2015. *Transparency in language: A typological study*. Amsterdam: University of Amsterdam PhD Dissertation.
Leufkens, Sterre. 2020. A functional typology of redundancy. *Revista da ABRALIN* 19(3). 79–103.
McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2–3). 125–166.
McWhorter, John H. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.
McWhorter, John H. 2011. *Linguistic simplicity and complexity: Why do languages undress?* Boston/Berlin: Walter de Gruyter GmbH & Co.
Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 23–41. Amsterdam: John Benjamins Publishing Company.
Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 110–125. Oxford: Oxford University Press.
Pallotti, Gabriele. 2015. A simple view of linguistic complexity. *Second Language Research* 31(1). 117–134.
Pate, John K. & Sharon Goldwater. 2015. Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language* 78. 1–17.
Petré, Peter. 2019. The interaction between extravagance and competition in syntactic change. *Paper presented at Societas Linguistica Europeaea 52*, Universität Leipzig, 21–24 August.
Rijkhoff, Jan. 2002. *The noun phrase*. Oxford: Oxford University Press.
Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter A. Kahrel. 1993. A method of language sampling. *Studies in Language* 17(1). 169–203.
Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2. 263–314.
Rounds, Carol H. 2009. *Hungarian: An essential grammar*. London: Routledge.
Ruhlen, Merritt. 1991. *A guide to the world's languages: Classification*, vol. i. London: Arnold.
Seinhorst, Klaas. 2021. *The complexity and learnability of phonological patterns: Simulations, experiments, typology*. Amsterdam: University of Amsterdam PhD Dissertation.
Slobin, Dan I. 1973. Cognitive prerequisites for the development of grammar. In Charles Ferguson & Dan I. Slobin (eds.), *Studies of child language development*, 175–208. New York: Holt, Rinehart & Winston.
Slobin, Dan I. 1980. The repeated path between transparency and opacity in language. In Ursula Bellugi & Michael Studdert-Kennedy (eds.), *Signed and spoken language: Biological constraints on linguistic form*, 229–243. Weinheim: Verlag Chemie.
Sohn, Ho-Min. 1999. *The Korean language*. Cambridge: Cambridge University Press.
Szmrecsanyi, Benedikt & Bernd Kortmann. 2009. Between simplification and complexification: Non-standard varieties of English around the world. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 64–79. Oxford: Oxford University Press.
Taraban, Roman. 2004. Drawing leaners' attention to syntactic context aids gender-like category induction. *Journal of Memory and Language* 51(2). 202–216.
Trudgill, Peter. 2009. Sociolinguistic typology and complexification. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 98–109. Oxford: Oxford University Press.
Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
Zeijlstra, Hedde. 2004. *Sentential negation and negative concord*. Amsterdam: University of Amsterdam PhD Dissertation.