

# A quantitative analysis of counterfactual conditionals in cross-linguistic perspective

Jesús Olguín Martínez,<sup>a</sup> Nicholas Lester<sup>b</sup>

<sup>a</sup> University of California, Santa Barbara, United States <olguinmartinez@ucsb.edu>

<sup>b</sup> University of Zürich, Switzerland <nicholas.a.lester@gmail.com>

People often reason about states of the world that could have been, but which are not, or those which could be, given that certain conditions are satisfied. When we make statements about such relationships, we usually divide them into two parts: the condition (protasis) and the result (the apodosis). While most languages signal the relationship between protasis and apodosis in counterfactual conditional constructions explicitly, they vary widely in the structures they use to do so. The present study addresses several questions related to cross-linguistic variation in this domain. How are the clauses marked for tense, aspect, and modality? What kind of clause-linking strategies are used to combine them? Are the clauses marked using the same or different morphosyntax? Through qualitative and quantitative analysis of a large sample of carefully selected languages, we demonstrate widespread differences between languages. We also uncover general patterns of features that correlate both with the symmetry and the morphosyntax of protases and apodoses in counterfactual conditionals across languages.

**KEYWORDS:** counterfactual conditional, adverbial clause, irrealis, TAM markers, clause-linking devices, clause combining, linguistic typology, subordination, coordination.

## 1. Introduction

Although adverbial clauses have long been a topic of strong interest to linguists, there have been relatively few studies of particular semantic relations from a cross-linguistic perspective. One semantic relation that has received little attention is that of counterfactual conditionals. Counterfactual conditionals convey the idea that the state of affairs denoted by them did not happen or could not happen (cf. Givón 2001: 332-333).<sup>1</sup> Various studies have addressed this semantic relation in individual languages (e.g. Arkadiev 2020 on Kuban Kabardian), in particular macro-areas (e.g. Nicolle 2017 on African languages), and in particular language families (e.g. Bhatt 1998 on Indo-Aryan languages). However, to the best of our knowledge, only a few broad cross-linguistic studies have explored counterfactual conditionals (e.g. Comrie 1986; Wierzbicka 1997; Haiman & Kuteva 2001; Xrakovskij 2005; Karawani 2014; Qian 2016).

One important aspect to bear in mind is that those cross-linguistic studies that have addressed counterfactual conditionals have examined them in a monofactorial context (e.g. Haiman & Kuteva 2001 on the symmetric and asymmetric patterns of counterfactual conditionals). In monofactorial studies, one independent variable (at a time) is investigated without reference to any other independent variables. However, one independent variable hardly ever accounts for all variation in a dependent variable (Wulff *et al.* 2014: 276-277). There is only one cross-linguistic study that has explored counterfactual conditionals in a monofactorial context by taking into account several parameters independently. Qian (2016: 104), in his typological work, considers the formal type of subordinating device, the order of the protasis and apodosis, and the deranking status of the protasis. Although the study of these parameters in a monofactorial context provides an important point of departure, we should like to know how each behaves when the others have been controlled for (Wulff *et al.* 2014: 276-277). Monofactorial methodologies cannot shed light on this theoretical aspect.

This paper has two goals. First, we explore counterfactual conditionals by taking into account a genetically and areally balanced sample of 107 languages. As will be seen, counterfactual conditionals vary widely across languages. To keep the scope of the discussion manageable, we focus on three parameters: (i) the symmetric and asymmetric morphological patterns of counterfactual conditionals, (ii) the range of Tense-Aspect-Mood values (henceforth **TAM**) that tend to appear in the protasis and apodosis of counterfactual conditional constructions, and (iii) the range of clause-linking strategies used in the encoding of counterfactual conditionals. Next, we apply two statistical analyses of these distributions to a subset of the overall database (selected to maximize sample sizes across the classes that we compare statistically). Here we test which factors lead to symmetric *vs* asymmetric systems, as well as which morphosyntactic features distinguish protasis from apodosis within this subsample. Based on these two studies, we hope (a) to uncover the cross-linguistic distribution of the features associated with counterfactual conditionals, (b) to determine what conglomerations of factors produce symmetric *vs* asymmetric systems, and (c) to see whether protases and apodoses show reliable morphosyntactic properties across languages.

The paper is structured as follows. In Study 1, we first introduce the sample of languages. Then, we provide some theoretical remarks on counterfactual conditionals and propose a comparative concept required to compare this construction across languages. Based on this concept, we explore the symmetric and asymmetric morphological patterns of counterfactual conditionals in the world's languages. We further document the range of

TAM values that tend to appear in both the protasis and apodosis and provide a concise overview of the clause-linking strategies used in the encoding of counterfactual conditionals. In Study 2 (see section 3 for a detailed discussion), we report the results of two statistical analyses that test (i) which variables best predict the morphological symmetry/asymmetry of this construction across languages and (ii) whether there are any typological associations between TAM and clause type within the construction. Finally, we synthesize the results of the two studies to arrive at a statistically informed picture of the typological variability of counterfactual conditionals.

## *2. Study 1: Typological variables associated with counterfactual conditionals*

Since this study is one of the first attempts to explore counterfactual conditionals from a typological perspective, the ideal strategy is to build a sample that is genetically stratified at the level of genus, which aims at avoiding a genetic bias.<sup>2</sup> Having a genetically balanced sample is critical given that our main aim is to find statistical tendencies and correlations. Genetic biases could produce misleading or unreliable statistical conclusions, that is, conclusions that can only be generalized to specific (sub)groups of languages. Not only genetic biases need to be removed, but also areal biases. In this research, we take into account a genetically and areally balanced sample of 107 languages based on the Genus-Macroarea method proposed by Miestamo *et al.* (2016). In this method, the primary genealogical stratification is made at the genus level, and the primary areal stratification at the level of macroareas. While the Genus-Macroarea method adopts the genealogical and areal stratification proposed by Dryer (2013), it does not follow the procedure(s) Dryer adopts for the selection of languages in Dryer (1989).<sup>3</sup> In Dryer's (1989) method of sampling, languages are first included in the sample without a systematic method of selection, and this bottom-up approach is then complemented by a more systematic stratification at the stage of testing generalizations (Miestamo *et al.* 2016: 242). The structure and motivations behind the selection of languages deserve some explanation.

In the Genus-Macroarea method, constructing a sample without predetermined sample size will, at its simplest, mean picking one language from every genus. This means that we attempted to find one language from each of Dryer's genera for which the available literature gives sufficient information on the grammar of counterfactual conditionals. Dryer's (2013) classification in WALS contains 543 genera. It is important to mention that in Dryer's classification each language belongs to a genus and each genus belongs to a family. Note that a language can be

the only member of its genus, and a genus may form a family on its own (Miestamo *et al.* 2016: 239). For some genera, we were not able to find any language that meets that criterion. Therefore, these genera do not (indeed, cannot) figure in our discussion. We were able to find sufficient information on one language in each of exactly 165 genera (i.e. 165 genera out of 543, or 30%), which accounts for our Core Sample.<sup>4</sup>

One aspect to consider is that some macro-areas are better represented than others in the languages of the Core Sample because of the availability and quality of the sources. Bibliographic bias tends to introduce areal bias (Miestamo *et al.* 2016: 251). This areal bias is problematic for the reason that we are exploring relationships between linguistic parameters. In order to avoid this areal bias, we followed the method for achieving a better areal balance that was introduced in Miestamo (2005). In this regard, the Restricted Sample is a subsample drawn from the Core Sample with the aim to balance the representation of each macro-area, and therefore to avoid areal biases. In the Restricted Sample the number of genera of the least well-represented area defines the maximal size of the Restricted Sample that can be drawn from the Core Sample (Miestamo *et al.* 2016: 252).

In our study, as can be seen in Table 1, the least well represented area is South America with 19.09% of its genera covered in the Core Sample. The Restricted Sample will thus include 19.09% of the total number of genera in each macro-area. For instance, 19.09% of the total number of genera in Africa (77) gives the number of African languages in the Restricted Sample as 15, 19.09% of the total number of genera in Papunesia (136) gives the number of Papunesian languages in the Restricted Sample as 26, etc.

MACRO-AREA	NUMBER OF GENERA	NUMBER OF GENERA IN THE CORE SAMPLE	COVERAGE	NUMBER OF GENERA IN THE RESTRICTED SAMPLE	COVERAGE
Africa	77	23	29.87%	15	19.09%
Australia	43	17	39.53%	10	19.09%
Eurasia	82	27	35.36%	17	19.09%
North America	95	28	31.57%	18	19.09%
Papunesia	136	43	32.35%	26	19.09%
South America	110	21	19.09%	21	19.09%
Total	543	165	—	107	—

Table 1. Genera covered in the Core Sample and Restricted Sample.

With the areal bias removed, the Restricted Sample is better suited to serve as a basis for quantitative analysis. Table 2 provides the list of 107 languages of the Restricted Sample arranged by macro-area.

MACRO AREA	SAMPLE LANGUAGES	SUM
Africa	Bangime, Boko, Emai, Eton, Gumuz, Koyra Chiini, Lango, Lumun, Maba, Ngiti, Sandawe, Supyire, Tamashek, Tommo So, Ts'ixa	15
Australia	Bininj Gun-Wok, Gaagudju, Gooniyandi, Gurr-Goni, Kuku Yalanji, Malakmalak, Mangarrayi, Ngankikurungkurr, Ungarinjin, Wardaman	10
Eurasia	Armenian, English, Finnish, Georgian, Hungarian, Ket, Kharia, Kodava, Korean, Lao, Lezgian, Mongsen Ao, Palula, Spanish, Tangsa, Udihe, Yukaghir	17
North America	Ayutla Mixe, Barbareño Chumash, Buglere, Central Alaskan Yupik, Chol, Cree, Creek, Crow, Garifuna, Haida, Huasteca Nahuatl, Jamul Tiipay, Slave, Sochiapan Chinantec, Teribe, Wappo, Warihio, Yuchi	18
Papunesia	Abau, Awtuw, Balantak, Barai, Barupu, Dadibi, Daga, Duna, Golin, Ilocano, Imonda, Inanwatan, Kaluli, Kombio, Komnzo, Lavukaleve, Makasae, Manambu, Motuna, Paiwan, Rapanui, Rotokas, Sulka, Taba, Urama, Urim	26
South America	Aguaruna, Ashéninka Perené, Awa Pit, Baure, Bora, Epena Pedee, Guna, Hup, Iquito, Kokama-Kokamilla, Kwaza, Mapudungun, Mosetén, Movima, Murui Huitoto, Muylaq' Aymara, Puinave, Tariana, Tiriyó, Urarina, Yurakaré	21
Total		107

**Table 2.** List of languages of the Restricted Sample arranged by macro-area.

Before leaving the present section, some remarks on the languages that were left outside the Restricted Sample are in order. We followed Miestamo *et al.* (2016: 253), who explain that the languages that must be left outside the Restricted Sample should come from the family (or families) with the greatest number of languages in the sample in each macro-area, and leave out a language or languages from that family (or families). This reduces genetic bias by reducing the influence of large families. For instance, for the African macro-area, we decided to leave out languages from different genera of the Niger-Congo and Afro-Asiatic families, the two largest families in Africa. This approach therefore maximizes the independence of the languages sampled in that it avoids including languages from different genera of the same family which might share a feature inherited from the proto-language of the family. Note that unlike this method, in Dryer's 1989 method two or more languages from different genera of the same family may be

included, which may result in a sample not suitable for statistical analysis. It is important to stress that whenever we noticed that two or more languages belonging to different families have been in contact, we decided to choose other languages from the same families. However, sometimes it was not possible to establish whether the languages belonging to different families shared a feature because of intense contact or chance. It is important to mention that these cases are rather few and do not detract from the validity of our overall statistical conclusions.

### *2.1. Theoretical remarks on counterfactual conditionals and comparative concept*

Conditional clauses have been traditionally divided into different types. For instance, Bennett (1908: 198) divides Latin conditional clauses into conditional clauses in which nothing is implied as to the reality of the supposed case, hypothetical conditional clauses, which refer to imagined state of affairs that might hypothetically happen, and counterfactual conditionals, which refer to imagined state of affairs that did not happen. Smyth (1920: 517-520) considers different types of conditional clauses in Greek, such as simple past and present conditionals (i.e. conditions which simply state a supposition with no implication as to its reality or probability), present and past unreal conditionals (i.e. the protasis implies that the state of affairs cannot be realized because contrary to a known fact), more vivid future conditionals (i.e. the speaker sets forth a thought as prominent in his mind), less vivid future conditionals (i.e. it expresses suppositions less distinctly conceived and of less immediate concern to the speaker), and general conditionals (i.e. they refer to a state of affairs that is very likely to occur).

In a recent typological study, Qian (2016: 108) shows that most languages tend to make a morpho-syntactic tripartite differentiation in hypotheticality in conditionals clauses. He mentions that only a few languages lack a formal distinction among *when*-clauses, *if*-clauses and *if...would*-clauses. In languages where the distinction is not encoded, the differentiation between temporal and hypothetical constructions is therefore contextually dependent. The focus of this paper is on counterfactual conditionals. Counterfactual conditionals are considered to be semantic primitives in that every language should have a construction which allows the speaker to express this meaning (Wierzbicka 1997: 28). Before launching into this discussion, the reader should bear in mind one remaining general point.

One of the biggest challenges of typology is coping with different terminological traditions across languages while exploring one particu-

lar phenomenon. For instance, different terminology has been used to talk about non-finite adverbial forms, such as converbs in Altaic languages, gerunds and adverbial participles in languages from Europe, medial verbs in languages from New Guinea, and conjunctive participles in languages from South Asia (Haspelmath 1995: 23). Another example comes from apprehensive markers. The terminology used to refer to apprehensive markers varies a lot, especially from one geographical area to another and across language families. In this regard, Vuillemer (2018: 259) explains that she has identified about 20 terms, such as admonitive, avertive, warning clitic, timitive, volitive of fear or fear case marker, among others. Unlike these phenomena, the terminology used in different grammars to refer to counterfactual conditionals seems not to vary a lot. Protasis and apodosis are the most common ways to refer to the counterfactual conditional clause and the main clause respectively. Other less common ways are antecedence and consequent, subordinate clause and matrix clause, and dependent clause and superordinate clause. In the present study, we have chosen to use the terms 'protasis' and 'apodosis'. This is due to the fact that, as explained by Traugott (1985: 304), the concepts 'protasis' and 'apodosis' are the traditional terms, whereas antecedent and consequent are associated more directly with the philosophical tradition. We now turn to the definition of counterfactual conditionals adopted in this study.

The definition in (1) is the comparative concept put forward in this paper.<sup>5</sup> This definition facilitates cross-linguistic comparability and does not impose any a priori restrictions on the form of counterfactual conditionals.

(1) Counterfactual conditional: A counterfactual conditional clause is a type of complex sentence construction in which the relation between the protasis and apodosis is that of an imagined state of affairs that did not happen.

There are two key components that can be highlighted from the definition in (1): complex sentence construction and imagined state of affairs that did not happen. The first component (i.e. complex sentence construction) refers to a specific relationship between (at least) two states of affairs in (at least) two clauses (Longacre 1985: 255; Croft 2001: 320-321).<sup>6</sup> Complex sentence constructions are thus sentences that contain more than one clause. A clause, in turn, can be defined as a unit minimally consisting of a predication that may be accompanied by its arguments and modifiers (Lehmann 1988: 182; Haspelmath 1995: 11; Gast & Diessel 2012: 4 – among many others). The syntactic relation between these two states of affairs may be one of coordination or subordination, among others.<sup>7</sup>

Conceived of in this way, the notion of complex sentence construction is useful because it has enabled us to incorporate counterfactual conditional constructions which show different types of syntactic relations and are encoded by various types of clause-linking strategies. Therefore, if the component complex sentence construction is substituted by a more particular syntactic relation, such as subordination, many languages will have to be excluded from the present study, such as the example in (2), from Imonda (West Sepik). Note that in this example both the protasis and apodosis are simply juxtaposed rather than marked by any specific morpho-syntactic device(s). Complicating the picture further, the distinction between subordinate and main clauses is regarded by many linguists to be gradual (Gast & Diessel 2012: 5; Lehmann 1988: 190 – among many others), making it difficult to define or to compare across languages.

(2) Imonda (Seiler 1985: 206)

*ka heulō-ta-ba, ne-m ka eg-t.*  
1SG.SBJ hear-IRR-TOP 2SG.OBJ-GL 1SG.SBJ follow-CF  
'If I had heard (you), I would have followed you.'

With this in mind, the following range of complex sentence constructions and clause-linking strategies are taken into account in the present study. First, languages may encode counterfactual conditionals by means of paratactic structures, as in (3). By parataxis is meant two clauses without any structural element linking them. The relation arises by implicature, usually due to contextual or common knowledge and/or iconicity of sequencing (Greenberg 1966; Haiman 1980). As pointed out by Mauri & Sansò (2009), it is not infrequent to find languages lacking grammaticalized strategies (e.g. free adverbial subordinators) and expressing counterfactual conditional relations by means of paratactic constructions. Mauri & van der Auwera (2012: 396) explain that in this scenario not all is left to inferential processes. Rather, if a language expresses counterfactual conditionals by means of paratactic constructions, at least one of the linked state of affairs has to be marked as irrealis (by means of irrealis, dubitative, or hypothetical elements) in order for the counterfactual conditional relation to be inferable. Verstraete (2014: 223) mentions that TAM markers, in paratactic counterfactual conditionals, may serve as a pragmatic trigger of the counterfactual conditional interpretation. This pattern shows a clear areal pattern, as has been shown previously by Haiman (1983). Its mainstay is Papua New Guinea and Australia (see section 2.4. for a similar cross-linguistic distribution attested in the present study). In the Yimas (Lower Sepik) example in (3), two clauses appear one after the other without any grammaticalized strategy. In order for the counterfactual conditional relation to be inferable, it is necessary that the

two clauses are overtly marked as potential, otherwise the hearer could interpret the construction as a purely temporal or causal relation.

(3) Yimas (Foley 1991: 442)

*tunkurŋ ant-ka-tay-c-mp-n, ant-ka-tu-r-ak.*  
eye.VI.SG POT-1SG.SBJ-see-PFV-VII.SG-OBL POT-1SG.SBJ-kill-PFV-VII.SG.OBJ  
'If I had seen the eye (of the crocodile), I would have killed it.'

Counterfactual conditionals may also be encoded by a general coordinating device, as is shown in (4). General coordinating devices are coordinating linkers, such as 'and' (Haspelmath 2004), that occur in a biclausal construction, from which a counterfactual conditional relation is inferred due to iconicity of sequencing and/or contextual factors. Given the underspecification of general coordinating devices, Mauri & van der Auwera (2012: 396) also explain that not all is left to inferential processes and at least one of the linked state of affairs has to be marked as irrealis (by means of irrealis, dubitative, or hypothetical elements) in order for the counterfactual conditional relation to be inferable, as is shown in the Sulka (Isolate) example in (4). In this example, if one of the clauses does not appear with *-ngoe*, the hearer could interpret the construction as a purely temporal or causal relation (Tharp 1996: 153).

(4) Sulka (Tharp 1996: 153)

*ip-ngoe va nap-ngoe.*  
2SG.SBJ-go.PST.COND and 3SG.SBJ-go.PST.COND  
'If you had gone, then he would have gone.'

We also take into account, in the present study, counterfactual conditionals encoded by grammaticalized strategies, i.e. dedicated devices, which explicitly encode the semantic relation of the adverbial clause to the state of affairs expressed in the main clause. The most common dedicated devices by which counterfactual conditionals tend to be encoded in the languages of the sample are dedicated adverbial subordinators and specialized converbs. Some comments on the properties of these devices and the challenges in defining them are in order.

A dedicated adverbial subordinator is a morpheme that marks a subordinate adverbial clause for its semantic relationship to the main clause. For the most part dedicated adverbial subordinators are associated with free subordinating items, illustrated in the San Andrés Otomi (Oto-Manguenan) example in (5), where the counterfactual conditional relation is encoded by the free adverbial subordinator *bì* 'if'. However, there are languages in which dedicated adverbial subordinators may be bound morphemes, as can be seen in the Rama (Chibchan) example

in (6), where the counterfactual conditional relation is encoded by the bound adverbial subordinator *-kata* ‘if’.

(5) San Andrés Otomi (Lastra de Suárez 2001: 136)

*bì* *k<sup>w</sup>a-ní*, *k<sup>w</sup>a-ó-hpí* *r<sup>^</sup>* *másçité*.  
 if 1SG.SBJ.SUBJ-see 1SG.SBJ.SUBJ-ask-3SG.OBJ ART machete  
 ‘If I had seen (it), I would have asked him the machete.’

(6) Rama (Craig 1990: 165)

*nah* *maa* *alkuk-kata*, *nah* *uwaik*  
 1SG.SBJ 2SG.SBJ hear-if 1SG.SBJ long.time  
*siik-ut*.  
 come-IRR  
 ‘If I had heard (that) you (had come), I would have come a long time ago.’

The greatest obstacle in defining dedicated adverbial subordinators in the present study has been to define what a subordinate clause is (Kortmann 1997: 57). However, given that subordination is a multidimensional phenomenon (Lehmann 1988) described by a set of independent formal parameters (e.g. dependent clause reduces its range of TAM values, dependent clause increasingly acquires nominal properties), there will be instances in which the dedicated adverbial subordinator will clearly operate in a subordinate clause and others in which it will not. In the Movima (Isolate) example in (7), the free adverbial subordinator *disoy* ‘if’ introduces a clause that is clearly subordinate in that it is deprived of any TAM markers, it appears with nominalizing morphology (i.e. the suffix *-wa*), and it occurs with the oblique article *nokos*, commonly found in nominal elements. The opposite situation is shown in the Wardaman (Yangmanic) example in (8) in that the free adverbial subordinator *bujun* ‘if’ appears in a dependent clause that is marked for its own TAM markers and shows overt participant coding. This clause appears with the same properties of main clauses (Merlan 1994: 188).<sup>8</sup>

(7) Movima (Haude 2006: 532)

*disoy* *no-kos* *dinkaye-wa-nkweł*,  
 if OBL-ART hurry-NMLZ-2PL.SBJ  
*di’* *man <a>ye = nkweł* *ney* *dira*.  
 HYP meet <DR> = 2PL.SBJ here still  
 ‘If you had hurried, you might still have met them here (but you didn’t).’

(8) Wardaman (Merlan 1994: 188)

*bujun* *yi-ngan-wo-ndi* *ma-jad*,  
 if IRR-3SG.SBJ.1.SG.OBJ-give-PST big-ABS  
*yi-ngong-wo-ndi*.  
 IRR-2SG.SBJ.1SG.OBJ-give-PST  
 ‘If he had given me a lot, I would have given you (some).’

Another important aspect regarding dedicated adverbial subordinators should be mentioned here. Since language is a not static, but rather a dynamic system that is in a constant state of flux (Croft 2003: 283), it is expected that languages may have dedicated adverbial subordinators that may not (yet) fully grammaticalized. When building the sample of the present study, we came across languages in which counterfactual conditionals are encoded by verbs meaning ‘to say’. Whether this form has become grammaticalized as a dedicated adverbial subordinator or not is unclear to us. For instance, in Anejom (Oceanic) the expression of counterfactual conditionals by means of the verb *ika* ‘say’ is very frequent, as in (9). In Araki (Oceanic) the form *co de* is a free adverbial subordinator related to the verb ‘say’, as can be seen in (10). Note that *de* ‘say’ is accompanied by the first person inclusive plural irrealis pronoun *co* which refers to the speaker and his addressee. However, it may also be accompanied by other types of person markers, which seems to suggest that it may not (yet) fully grammaticalized. François (2002: 177) explains that in this case *co de* has to be understood as ‘let us say that’, in a very similar way to English ‘let us suppose’. Other Oceanic languages in which this pattern is found are Bariai (Gallagher & Baehr 2005: 160), Big Nambas (Fox 1979: 108-109), Daakaka (Von Prince 2015: 378), Kwamera (Lindstrom & Lynch 1994: 35) and Mangap-Mbula (Bugenhagen 1995: 404). For the sake of transparency, the policy adopted in this study has been to exclude these instances from the present research on the grounds that it has not been possible to determine whether these strategies are dedicated adverbial subordinators or verbs. It is important to stress that these problematic cases are rather few and do not detract from the validity of our overall conclusions.

(9) Anejom (Lynch 2000: 161)

*et wut ika et idim itiyi ehe,*  
 3SG.AOR TEMP.CONJ.FUT say 3SG.AOR really NEG rain  
*ek pu idim apan m-asjan-ya.*  
 1SG.AOR FUT really go ES-throw-line  
 ‘If it really hadn’t rained, I would have gone fishing.’

(10) Araki (François 2002: 178)

*co de na maci, na pa avu.*  
 1.INCL.IRR say 1SG.SBJ bird 1SG.SBJ.IRR SEQ fly  
 ‘If I were a bird, I would have flown.’

One important aspect to bear in mind is that, in some languages, counterfactual conditionals may be encoded by two dedicated adverbial subordinators. For instance, in the Urarina (Isolate) example in (11), the dependent clause appears with *baana* ‘if’ and *hananiane* ‘if’.

(11)Urarina (Olawsky 2006: 255)

*baana itçuu-a=ne hananiane, raj kalau-tçuru muk-akatçe.*  
 if be.near-3SG.SBJ=SUB if POSS son-PL catch-1PL.SBJ  
 'If its creatures had been near, we would have caught it (about a peccary).'

Counterfactual conditionals may also be encoded by specialized converbs, that is, special verb forms that do not appear in independent declarative clauses (Cristofaro 2003: Chapter 3) and mark the adverbial clause for its semantic relationship to the main clause, as in the Ingush (Nakh-Daghestanian) example in (12). Although specialized converbs and bound adverbial subordinators may look similar at first glance, there are some clear-cut differences between them. While specialized converbs are part of the inflectional paradigm of verbs and thus in paradigmatic contrast to other inflectional morphemes, bound adverbial subordinators are not. What this means is that specialized converbs cannot be analyzed as a verb plus a subordinating affix (Haspelmath 1995: 4). Another important difference between these devices has to do with their lexical autonomy. Specialized converbs never have the degree of autonomy associated with the status of lexemes (Haspelmath 1995: 4), but bound adverbial subordinators do. These criteria have played an important role when exploring the sources of the sample.

(12)Ingush (Nichols 2011: 305)

*ehw dalaarie, mocagha hwa-dea xuddar.*  
 conscience GEND.IRR.CVB long\_ago DEIC-GEND.ANT.CVB go.GEND.COND  
 'If they had had any conscience, they would have done it long ago.'

Another thought-provoking example comes from Chamacoco in (13). In this language, counterfactual conditionals are encoded by parhypotaxis.<sup>9</sup> In this example the protasis appears with a dedicated adverbial subordinator and the apodosis appears with a general coordinating device that is obligatory. Interestingly, there are instances in which both the protasis and apodosis appear with a dedicated adverbial subordinator, as in the Paiwan example in (14), and the Lango example in (15).

(13)Chamacoco (Bertinetto & Ciucci 2012: 98)

*kēhe, uu like ishir lishi sēhe,*  
 if DET.SG.M this indigenous.SG.M poor.SG.M want  
*teehe, s-ohnimichi=ke, hn uhu oy-ihye ire.*  
 INTERJ 3.IRR-get.off=PST and 2SG.CAUS 1PL-arrest 3SG  
 'If the indigenous had wanted to get off (the bus), you would have made us arrest him.'

(14)Paiwan (Chang 2006: 318)

*kana na=meLay sa ?udal, kana=ken a vaik=anga.*  
 CF1 PERF=rain.stop.AV this.NOM rain CF2=1SG.NOM LK go.AV=COMPL  
 'If this rain had stopped, I would have already left.'

(15)Lango (Noonan 1992: 233)

*kónò* ànwòhò àtíé cem, *kónò* àmýí.  
 if 3SG.find.PFV 1SG.SBJ.be.PRES.HAB food if 1SG.SBJ-give-PFV-2SG.OBJ  
 ‘If I had had food, I would have given it to you.’

Having explained the constructions that are included in the present study due to the notion of complex sentence construction, we turn briefly to the constructions that are excluded due to this criterion. The examples in (16) and (17) are discarded from the study because they do not establish a relationship between two states of affairs, that is, both examples lack an apodosis.

(16)Ma’di (Blackings & Fabb 2003: 143)

*jí dri dri dzè kú.*  
 2SG.SBJ then hand wash NEG  
 ‘Had you not washed your hand.’ (you’d have been in real trouble) (the event of washing took place a few moments ago)

(17)Hunzib (Van den Berg 1995: 106)

*zuq’u-r q’ədə dízi y-at’əru řadam.*  
 be-PRÉT IRR me.DAT 2-love-PST.PTCP person  
 ‘If I only had a lover.’

The second component of the comparative concept used in the present study is that of an imagined state of affairs that did not happen. This component refers to past counterfactual conditionals, which express a counterfactual state of affairs in the past (e.g. *If John had come yesterday, we would have had fun*) and present counterfactual conditionals, which express a counterfactual state of affairs in the present (e.g. *If only John were here now, we would be happy*). The sources of the languages of the sample explain for the most part the encoding of past counterfactual conditionals rather than present counterfactual conditionals.

Before leaving the present section, it is important to bear in mind that we also take into account languages in which counterfactual conditionals and hypothetical conditionals are expressed in the same way and therefore they leave the interpretation to be inferred from the context. This theoretical fact has not gone unnoticed and echoes Qian (2016: 101), who explains that in some languages (e.g. Mising, Hmong, Tagalog, Dolakha Newar, Zuni, Vietnamese), there is a clear differentiation between real and hypothetical conditional clauses. However, in these languages a hypothetical or a counterfactual conditional reading is contextually dependent. This is shown in the Gumawana (Oceanic) example in (18) and the Longgu (Oceanic) example in (19), in which

there is a construction that allows both a hypothetical and counterfactual conditional reading.

(18)Gumawana (Olson 1992: 360)

*neta i-tagona, dedei-na, ta-tupa.*  
 if 3SG-offer good-3SG 1PL.INCL-sail  
 'If he offered, then good, we would sail.'  
 'If he had offered, then good, we would have sailed.'

(19)Longgu (Hill 1992: 286)

*zuhu no beata roporopo-i, gaoa ho la bweubweu.*  
 if IRR fine morning-SG 1DU.INCL IRR go walking  
 'If it were fine this morning, we would go for a walk.'  
 'If it had been fine this morning, we would have gone for a walk.'

The general spirit of this section has been to bring greater conceptual clarity to the understanding of counterfactual conditionals. In doing so, this section provided a brief survey of the main components of counterfactual conditional in the light of cross-linguistic data. In the following sections, we explore the three parameters mentioned in §1.

## 2.2. Symmetric and asymmetric patterns of the protasis and apodosis

Cross-linguistically, the verbs in the protasis and the apodosis of a counterfactual conditional may be encoded by different TAM values. This property may be called the asymmetry of conditionals (Haiman & Kuteva 2001: 101), as is illustrated by Mparntwe Arrernte (Pama-Nyungan) in (20), where the protasis appears with *-ke* and the apodosis with *-mere*.<sup>10</sup> However, sometimes the protasis and apodosis, irrespective of their particular morphological form, have parallel structures, which we refer to as a symmetric pattern. In the Quiégolani Zapotec (Oto-Manguean) example in (21), both the protasis and apodosis occur with the counterfactual mood marker *ny-*. Interestingly, there are languages in which counterfactual conditionals may be symmetric or asymmetric, as is shown in the examples in (22) and (23), from Huasteca Nahuatl (Uto-Aztecan). Another possibility is that neither the protasis nor the apodosis shows any TAM values, as is illustrated in (24), from Tetun (Austronesian). Note that these instances are treated as symmetric counterfactual conditionals because they show parallel structures.

(20)Mparntwe Arrernte (Wilkins 1989: 234)

*unte apmwerrke petye-ke,*  
 2SG.SBJ yesterday come-PST.COMPL  
*arrayte unte te-nhe are-mere.*  
 true 2SG.SBJ 3SG.OBJ-ACC see-HYP  
 'If you had come yesterday, then you certainly would have seen her.'

*A quantitative analysis of counterfactual conditionals in cross-linguistic perspective*

(21) Quiégolani Zapotec (Black 1994: 44)

*che-bel ny-oon=t Min, ny-oon-t Lawer.*  
 when-if CF-cry=NEG Jazmin CF-cry-NEG Laura  
 ‘If Jazmin had not cried, Laura would have cried.’

(22) Huasteca Nahuatl (Olguín Martínez 2016: 75)

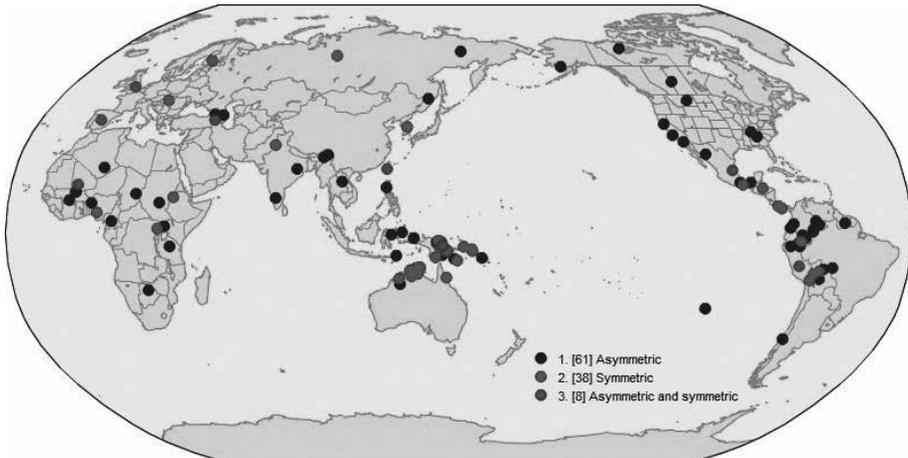
*tlān kin-kuah-toskia tama-li, amo mayana-toskia.*  
 if 3PL.OBJ-eat-COND.PST tamal-Abs NEG be.hungry-COND.PST  
 ‘If he had eaten tamales, he would not have been hungry.’

(23) Huasteca Nahuatl (Olguín Martínez 2016: 76)

*ach-ia-toya okichpil ihui-tl, ach-miki-toskia.*  
 NEG-go-PST.PERF boy party-Abs NEG-die-COND.PST  
 ‘Had the boy not gone to the party, he wouldn’t have died.’

(24) Tetun (Van Klinken 1999: 312)

*kalo ha’u fetu,*  
 if 1SG.SBJ woman  
*ha’u la bele k-akur tasi wé-n.*  
 1SG.SBJ NEG can 1SG.SBJ-cross sea water-GEN  
 ‘If I were not a woman, I wouldn’t have been able to cross the sea.’



**Map 1.** Distribution of symmetric and asymmetric counterfactual conditionals.

MACRO-AREA	SYMMETRIC	ASYMMETRIC	BOTH
Africa	3	11	1
Australia	9	1	0
Eurasia	5	9	3
North America	4	11	1
Papunesia	13	12	2
South America	4	17	1
Total	38	61	8

**Table 3.** Distribution of symmetric and asymmetric counterfactuals per macro-area.

As can be observed in Map 1, asymmetric counterfactual conditionals are the most robust type ( $61/107 = 57\%$ ). They are found in all macro-areas, but the preference for this type is especially strong in South America in the languages of the sample (i.e.  $17/61 = 27.86\%$ ), as is shown in Table 3. Symmetric counterfactual conditionals are the second most common type ( $38/107 = 35.51\%$ ). They are also found in all the macro-areas, but they are mostly attested in languages from Papunesia ( $13/38 = 34.21\%$ ) and Australia ( $9/13 = 23.68\%$ ), as is illustrated in Table 3. Haiman & Kuteva (2001: 109) explain that the symmetric morphological pattern of counterfactual conditionals is predominantly an areal typological feature in languages from Papua New Guinea. The authors mention that it occurs in almost every Papuan language they are aware of. Brooks (2018: 187) mentions that this symmetric pattern may be due to contact-induced language change by showing evidence from Chini and other languages from Papua New Guinea. In this regard, he mentions that the forms are not always cognate across Chini and other languages from Papua New Guinea, but the symmetric pattern is the same.

Having addressed the symmetric and asymmetric morphological patterns of counterfactual conditionals, we now turn our attention to the range of TAM values that tend to appear in these complex sentence constructions.

### 2.3. TAM values of counterfactual conditionals

Since counterfactual conditionals express non-actualized state of affairs, one would expect that they should appear with TAM markers whose semantics is appropriate to the counterfactual conditional context, such as irrealis markers, conditional mood markers, and counterfactual mood markers, among others (Mithun 1995: 384). However, it has long

been observed that, across a large number of unrelated languages, past tense markers, and other TAM markers whose semantics does not harmonize with the counterfactual conditional meaning (e.g. perfective, completive), tend to appear in counterfactual conditional constructions (Comrie 1986). This is a clear mismatch for the reason that past tense marking, perfective, and perfect marking tend to occur in state of affairs that are actualized and, as was mentioned above, counterfactual conditionals express non-actualized state of affairs. Different linguists have tried to offer a possible explanation to this mismatch. These can be divided into two lines of reasoning, that is, those who have adopted a remoteness-based approach and those who have adopted a back-shifting approach (see von Prince 2019 for a detailed explanation). First, those who have adopted the remoteness-based approach explain that past and counterfactuality share a semantic core of distance from the actual present (von Prince 2019). For instance, Steele (1975) explains that the connection between past tense and counterfactual conditionals is that the past tense marker has as its basic meaning not past tense but something like distant from present reality. Karawani (2014: 15) mentions that the connection between past tense and counterfactual conditionals stems from the fact that there is an inherent nature of the past as being closed and therefore the condition is impossible or false. Second, von Prince (2019) explains that past tense markers, in the back-shifting approach, “are thought to push one’s perspective back in time so that developments that are no longer possible become historically accessible.”

Our study shows that past tense markers and other TAM markers whose semantics do not harmonize with the counterfactual conditional meaning tend to occur in counterfactual conditional constructions. However, there may be more to the story. In this regard, past tense may combine with some other type of TAM marker expected to occur in non-actualized state of affairs (e.g. irrealis, counterfactual mood), showing a mixed pattern. For instance, the protasis of the counterfactual conditional in the Papantla Veracruz Totonac (Totonacan) example in (25) appears with different semantically conflicting TAM values, viz. the past tense marker *ix-* and completive marker *-li* (expected to occur in actualized state of affairs) and the counterfactual mood marker *-ti-* (expected to occur in non-actualized state of affairs).

(25) Papantla Veracruz Totonac (Levy 1990: 139)

*para ix-k-ti-akxilh-li, ix-k-ti-maqski'-lh ixmachita.*  
if PST-1SG.SBJ-CF-see-COMPL PST-1SG.SBJ-CF-ask-COMPL machete  
'If I had seen it, I would have asked him the machete.'

For the purposes of the present study, we discuss the range of TAM values of both the protasis and apodosis in a separate way. We use four terms to describe the range of TAM values of both the protasis and apodosis: actualized pattern, non-actualized pattern, mixed pattern, and unmarked pattern.

Actualized patterns refer to those instances in which the protasis or the apodosis is encoded by TAM values whose semantics do not harmonize with the counterfactual conditional context, such as past tense marking, perfect marking, completive marking, and perfective marking. For instance, the protasis of the Bangime (Isolate) example in (26) is encoded by perfective marking and past tense marking. These TAM values are not expected to appear in counterfactual conditionals.

(26) Bangime (Heath & Hantgan 2017: 465)

<i>sé</i>	<i>ɲ</i>	<i>jáá</i>	<i>Séédù</i>	<i>ɲijè</i>	<i>hìngà,</i>
if	1SG	see.PFV	Seydou	yesterday	PST
<i>ɲ</i>	<i>dégé</i>	∅	<i>náw.</i>		
1SG	hit.FUT	1SG	FUT		

'If I had seen Seydou yesterday, I'd have hit him.'

Non-actualized patterns refer to those instances in which the protasis or the apodosis is encoded by TAM values expected to occur in the counterfactual conditional context, such as irrealis, potential mood marking, conditional mood marking, counterfactual mood marking, future tense marking, and hypothetical mood marking. An example appears in (27) from Gooniyandi (Bunuban), where the protasis is encoded by the subjunctive *-ya-* and the irrealis *-ala*.

(27) Gooniyandi (McGregor 1990: 432)

<i>barlanyi</i>	<i>mila-ya-ala,</i>	<i>mangaddi</i>	<i>mood-gila-rni.</i>
snake	see-SUBJ-IRR.1SG.SBJ	NEG	step.on-IRR.1SG.SBJ-POT

'If I had seen the snake, I wouldn't have stepped on it.'

One remark on the irrealis category is in order here. Mithun (1995: 384) explains that the notion irrealis portrays the state of affairs as within the realm of thought, as knowable only through imagination. A source of potential confusion in any discussion on irrealis is that it has been applied to different concepts and constructions in languages from many areas of the world. It is therefore important to clarify what is meant when using this term. In this paper, we consider irrealis as specific markers (rather than notional descriptions of non-encoded meanings of constructions) in the forms of verbal affixes and clausal enclitics (Brooks 2018: 4). There seems to be a strong correlation between counterfactual conditionals and irrealis marking because, as explained

by Mithun (1995: 384), when languages have a grammaticalized realis/irrealis distinction, counterfactual conditionals tend to be encoded by irrealis marking. This study supports this theoretical claim in that most languages of the sample that have a grammaticalized realis/irrealis distinction tend to be marked by irrealis.

Mixed patterns refer to those instances in which the protasis or apodosis is encoded by a combination of two semantically conflicting TAM values, as can be seen in the Hungarian (Uralic) example in (28), where the protasis appears with the past tense marker *-t* and the conditional mood marker *volna*.

(28) Hungarian (Kenesei *et al.* 1998: 52)

*ha Péter-alud-t volna, Anna haragud-ott volna.*  
 if Peter-sleep-PST COND Anna be.angry-PST COND  
 ‘If Peter had been asleep, Anna would have been angry.’

By unmarked is meant those instances in which the protasis or apodosis is deprived of TAM marking, as can be seen in the example in (29) from Inanwatan (Marind). In this example, the protasis does not appear with any TAM values.

(29) Inanwatan (de Vries 2004: 39)

*lwáa-go dókter-e náwe úra-y-aigo,*  
 yesterday-CIRC doctor-M me see-TRANS-NEG  
*máwo-go nú-d-eqo.*  
 now-CIRC die-CF-1SG.SBJ  
 ‘If the doctor had not helped me yesterday, I would have died.’

Before leaving the present section, one remark on actualized and non-actualized patterns is in order here. There are languages in which the protasis will be nominalized, but it may appear with TAM marking that is actualized or non-actualized. This fact has not gone unnoticed and echoes Qian (2016: 156), who explains that in different languages the protasis or apodosis of counterfactual conditionals constructions may be nominalized, but may take TAM verbal inflections, as can be seen in Table 4.<sup>11</sup>

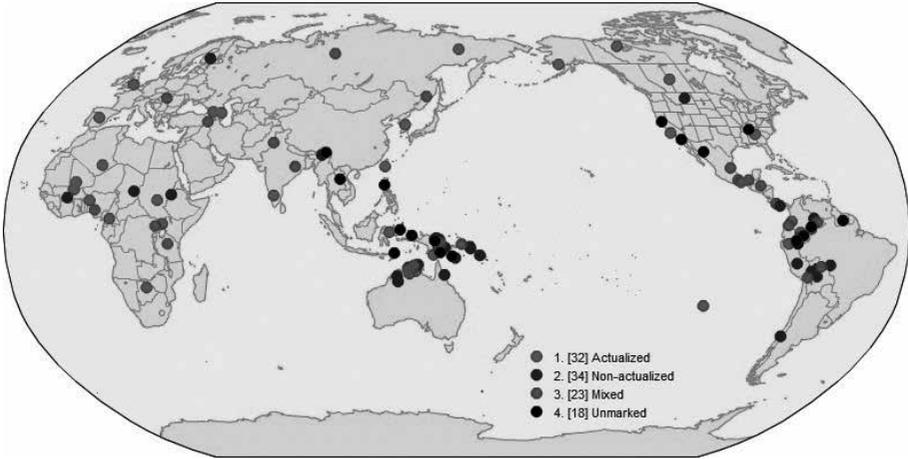
NOMINALIZATION OF PROTASIS	NOMINALIZATION OF APODOSIS
Hup, Kham, Macushi, Warekena	Afar, Kwazá, Movima, Pashto, Savosavo, Yimas

**Table 4.** Languages in which the protasis or apodosis of counterfactual conditional constructions is nominalized (Qian 2016: 158).

Having introduced the terminology that will be used in the following section, we can now proceed to explaining the most common TAM values of both the protasis and apodosis in counterfactual conditional constructions.

*2.3.1. TAM values of the protasis in counterfactual conditionals*

As can be observed in Map 2, the protases of counterfactual conditionals tend to appear with a non-actualized pattern (34/107 = 31.77%) or actualized pattern (32/107 = 29.90%) in the languages of the sample. While both types are found in all macro-areas, they seem to be more frequent in particular macro-areas. As is shown in Table 5, actualized protases seem to be slightly more common in Africa (9/32 = 28.12%) and non-actualized protases in Papunesia (i.e. 13/34 = 38.23%). Some other observations to be gleaned from Map 2 are the following. First, mixed protases are scattered in all macro-areas, but they seem to be slightly more frequent in Eurasia (i.e. 6/23 = 26.08%). Second, unmarked protases are mostly attested in Papunesia (i.e. 7/18 = 38.88%). Note that this type is not found in Africa and Australia in the languages of the sample.



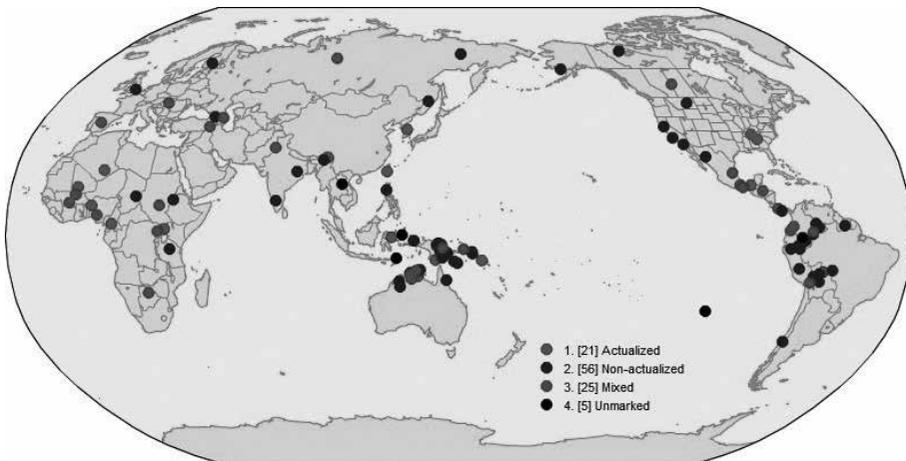
**Map 2.** TAM values of the protasis in counterfactual conditionals.

MACRO-AREA	ACTUALIZED	NON-ACTUALIZED	MIXED	UNMARKED
Africa	9	4	3	0
Australia	1	5	3	0
Eurasia	7	2	6	2
North America	7	2	4	4
South America	4	8	4	5
Papunesia	4	13	3	7
Total	32	34	23	18

**Table 5.** Distribution of TAM values of the protasis in counterfactual conditionals per macro-area.

*2.3.2. TAM values of the apodosis in counterfactual conditionals*

The first and most important finding, as can be observed in Map 3, is that apodoses encoded by non-actualized patterns are the most common pattern worldwide. In the sample, 56 languages (56/107 = 52.33%) show this pattern. In particular, this pattern seems to be more common in Papunesia (16/56 = 28.57%) and South America (15/56 = 26.78%). With respect to mixed patterns (25/107 = 23.36%), they are found in all macro-areas, but they do not seem to cluster in any particular area. Regarding actualized protases (21/107 = 19.62%), they are mostly attested in Africa in the languages of the sample. Note that languages tend not to have apodoses that are unmarked.



**Map 3.** TAM values of the apodosis of counterfactual conditionals.

MACRO-AREA	ACTUALIZED	NON-ACTUALIZED	MIXED	UNMARKED
Africa	7	3	6	0
Australia	1	6	3	0
Eurasia	1	8	6	1
North America	5	8	4	0
South America	1	15	4	1
Papunesia	6	16	2	3
Total	21	56	25	5

**Table 6.** Distribution of TAM values of the apodosis in counterfactual conditionals per macro-area.

In the following section we explain the last parameter addressed in the present study, viz. the range of clause-linking devices used in the encoding of counterfactual conditionals.

#### *2.4. Clause-linking devices used in the encoding of counterfactual conditionals*

Clause-linking devices are among the most important means used to establish subordinative and coordinative relations (Hetterle 2015: 106). These devices may sometimes shed light on the type of semantic relation that holds between clauses (e.g. adverbial subordinators, specialized converbs) in that they serve as devices for labeling complex sentence relations like causal, conditional or temporal relations (Verstraete 2014: 195). Counterfactual conditionals are encoded by different formal types of clause-linking devices. For the purposes of this study, we classify these strategies in the following way.

First, specialized devices refer to devices that are only used to encode counterfactual conditionals. These include dedicated adverbial subordinators and specialized converbs. In the example in (30) from Eton (Niger-Congo), the free clause-linking device *bén* is only used to encode counterfactual conditionals. Therefore, this device is specialized. Second, non-specialized devices refer to devices that encode counterfactual conditionals and other semantic types of conditionals (e.g. real, generic, and hypothetical). In Aguaruna (Jivaroan), all semantic types of conditionals are encoded by the subordinating affix *-ka* as can be observed in (31) and (32). This seems to indicate that *-ka* is a non-specialized device. Third, parataxis refers to those languages in which counterfactual conditionals and other semantic types of conditionals (e.g. real, generic, and hypotheti-

cal) do not appear with any clause-linking device, as can be observed in the examples in (33) and (34) from Gaagudju (Isolate).

(30)Eton (Van de Velde 2008: 365)

*ben nâ nɛ à-dĩdĩá vǎ,*  
 if COMP I.PPR I-FOC~being here  
*mà-lédà-H wò.*  
 1SG.SBJ-show-CONS 2SG.NPPRI.PPR  
 'If it had been here, I would have shown it to you.'

(31)Aguaruna (Overall 2017: 391)

*wi kafini wi-a-ku-nu-ka,*  
 1SG.SBJ tomorrow go-IPFV-SIM-1SG:SS-COND  
*taka-sa-tja-tata-ha-i.*  
 work-ATT-NEG-FUT-1SG.SBJ-DECL  
 'If I go tomorrow, I won't work.'

(32)Aguaruna (Overall 2017: 507)

*ami wi-tfau-aita-ku-mi-ĩ-ka,*  
 2SG.SBJ go.PFV-NEG:REL.COP-SIM-2-COND  
*fihha ani-sa-nu puhu-mai-inu-aita-ha-i.*  
 well be.happy-SUB-1SG:SS live-POT-NMLZ-COP-1SG-DECL  
 'If you had not gone, I would be happy.'

(33)Gaagudju (Harvey 2002: 371)

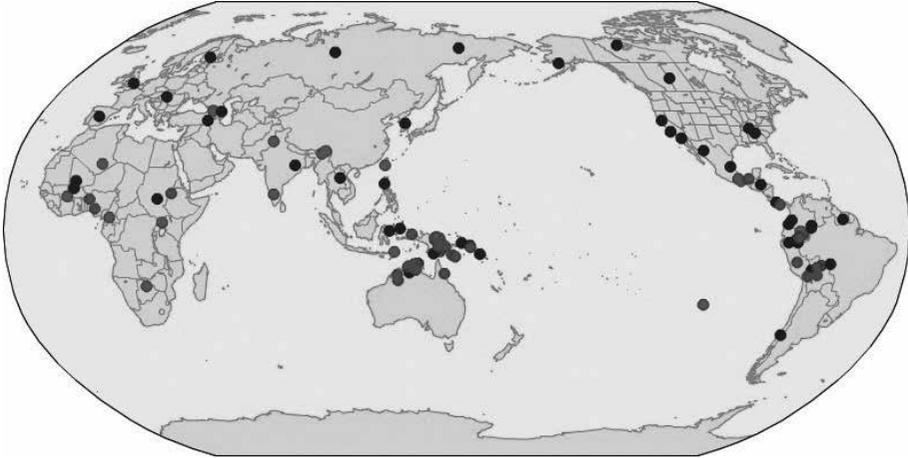
*i-rree-ma biirndi magaadja arree-wagi.*  
 3I-1SG.SBJ-get.FUT money that.IV 1SG.SBJ-go.back  
 'If/When I get money, I will go back there.'

(34)Gaagudju (Harvey 2002: 372)

*ø-ng-goro-garraa-ri arr-geenma-ri = ni.*  
 3I-1SG.SBJ.IRR-see-AUX-PST 1SG.SBJ.say.IRR-PST = 3SG.M.IND.OBJ  
 'If I had seen him, I would have told him.'

As Map 4 demonstrates, non-specialized devices are the most common type (45/93 = 48.38%; indicated by blue dots). These are attested in all macro-areas. However, they seem to be more frequent in North America (11/45 = 24.44%), Eurasia (10/93 = 22.22%), and South America (10/45 = 22.22%). The second most frequent type is that of paratactic counterfactual conditionals (28/93 = 30.10%; indicated by green dots). Interestingly, this type of clause-linking strategy shows clear areal skewings in that they can be found mainly in two macro-areas, viz. Australia (7/28 = 25%) and Papunesia (12/28 = 42.85%), in particular in languages from Papua New Guinea. Note that paratactic counterfactual conditionals are completely absent from Eurasia. The third type, and the least common device, is that of specialized devices (20/93 = 21.50%; indicated by red dots). They are attested in all macro-areas, but do not

seem to show any areal clusters. Note that we removed all languages with unknown clause-linking devices (n = 14) in order to explore the cross-linguistic distribution of clause-linking devices used to express counterfactual conditional constructions.



**Map 4.** Clause-linking devices used in the encoding of counterfactual conditionals.

MACRO-AREA	SPECIALIZED	NON-SPECIALIZED	PARATAXIS
Africa	5	4	3
Australia	0	3	7
Eurasia	5	10	0
North America	2	11	2
South America	3	10	4
Papunesia	5	7	12
Total	20	45	28

**Table 7.** Distribution of clause-linking devices used in the encoding of counterfactual conditionals per macro-area.

### *3. Study 2: Statistical analyses*

In study 2, we perform two statistical analyses. The first aims to uncover the variables that impact whether protases and apodoses are encoded via symmetrical or asymmetrical patterns. The second tests which, if any, TAM markers are distinctively associated with the protasis or apodosis across the languages in our sample. Prior to the analyses, we reduced the sample of languages. In particular, we removed all languages for which it has not been possible to determine whether the linking device is specialized (i.e. devices that are only used to encode counterfactual conditionals) or non-specialized (i.e. devices that encode counterfactual conditionals and other semantic types of conditionals, e.g. real, generic, and hypothetical) ( $n=14$ ). These languages account for approximately 7% of the sample. We further removed those languages with systems that lacked TAM marking on the apodosis ( $n=5$ ) or that allow nominalization of the protasis ( $n=2$ ). These trims were necessary given issues of data sparsity. The final sample consisted of 86 languages.

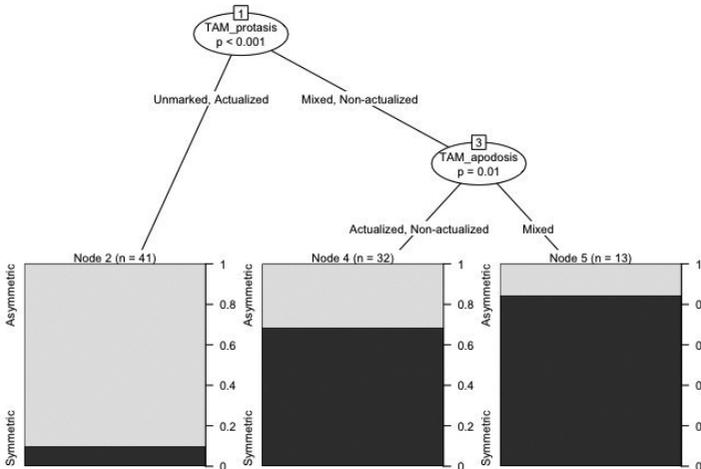
#### *3.1. Classification and Regression Tree (CART) analysis<sup>12</sup>*

Our first goal is to discover which variables predict the presence of symmetrical or asymmetrical systems for counterfactuals cross-linguistically. For this purpose, we use a technique from machine learning known as Classification and Regression Tree (henceforth **CART**) analysis (our task is one of classification). We have selected this analysis for several reasons. First, we are dealing with a relatively small sample of labeled entities (in this case, languages). Second, we have several categorical predictor variables, each with several levels. Third, many of the cells in the cross-tabulated predictor space are sparsely populated or contain zeroes. That is, we do not have enough observations of many of the variable combinations to make reliable estimates of their behavior with respect to our dependent variable. All of these facts create problems for more common methods of classification, such as binary logistic regression.<sup>13</sup> We therefore select the non-parametric classification algorithm known as CART. CART analysis involves the recursive binary partitioning of a dataset based on which predictor variable is most strongly associated with the outcome variable. Associations are weighted using significance tests against the null hypothesis that the predictor and outcome variables are unrelated. At each potential decision point in the tree, all predictors are considered, and the resulting set of  $p$ -values are corrected for multiple comparison. Here we apply the Bonferroni correction. As the partitions must be binary, the levels of each categorical variable used for each split are divided into

two groups. Partitioning is stopped when all corrected  $p$ -values are greater than the significance threshold (here,  $\alpha = .05$ ).

For this analysis, we included four predictors: macroarea, TAM-marking on the apodosis and protasis (respectively), and clause-linking strategy. The resulting model achieved 85% classification accuracy. Simply guessing the most frequent symmetry label yields a performance of 57% (24% poorer than our model). Sampling randomly based on the true distribution (i.e., sometimes guessing the less frequent outcome in proportion to the observed distribution; baseline =  $p_{\text{symmetrical}}^2 + p_{\text{asymmetrical}}^2$ ) yields a performance of 51% (30% poorer than our model).

The classification tree is presented in Figure 1. The highest-level split was made using the TAM marking on the protasis. This finding alone is interesting, as it suggests that morphological (a)symmetry depends most strongly on the properties of the protasis rather than the apodosis of the counterfactual conditional construction. In particular, languages with unmarked or actualized protases are reliably distinguished from those with mixed or non-actualized protases ( $p < .001$ ). The former group contains almost exclusively asymmetric languages (bar graph for node 2). For languages with mixed or non-actualized protases, the TAM of the apodosis further helped to predict (a)symmetry. Actualized and non-actualized apodoses were reliably distinguished from mixed apodoses ( $p < .05$ ). Both groups of languages overwhelmingly prefer symmetric marking (bar graphs for nodes four and five).



**Figure 1.** Results of the CART analysis predicting the (a)symmetry of the counterfactual system across languages.

### 3.2. Contingency analysis

To determine the TAM properties that distinguish apodosis from protasis, we perform a contingency analysis adapted from Gries & Stefanowitsch (2004). This analysis involves a Fisher-Yates exact test computed over a cross-tabulation of TAM marking strategies and type of system. The contingency analysis works by constructing a series of  $2 \times 2$  tables. Each cell contains a frequency. Columns represent the outcome levels (*apodosis* vs *protasis*). Rows represent a given TAM value (e.g. *mixed*) versus all other levels. The direction of any significant results is derived from the difference between observed and expected frequencies (we assume a uniform distribution as the null hypothesis). A positive difference, or over-representation relative to the expected baseline, indicates affiliation; a negative difference, or under-representation, indicates repulsion. The raw data for the analysis are provided in Table 8.

TAM	APODOSIS	PROTASIS
non-actualized	47 (.62)	29 (.38)
unmarked	0 (.00)	15 (1.00)
actualized	19 (.42)	26 (.58)
mixed	20 (.56)	16 (.44)

**Table 8.** Frequency of TAM types per clause type (%).

It is immediately clear that unmarked TAM in this sample appears exclusively on the protasis. Non-actualized TAM markers are roughly twice as likely to occur on the apodosis. Actualized and mixed TAM marking are more evenly distributed across the clause types. Table 9 shows the results of the contingency analysis (only significant relationships are reported).

TAM	$F_{OBS}$ APO	$F_{OBS}$ PRO	$F_{EXP}$ APO	$F_{EXP}$ PRO	$\Delta P_{TYPE \rightarrow TAM}$	$\Delta P_{TAM \rightarrow TYPE}$	$P$	PREF
non-actualized	47	29	38	38	0.21	0.21	<.001	apo
unmarked	0	15	7.5	7.5	0.17	0.55	<.001	pro

**Table 9.** Distinctive affiliation of TAM types to (a)symmetry of the counterfactual.

Table 9 provides several pieces of information. First, we have the observed ( $f_{obs}$ ) and expected ( $f_{exp}$ ) frequencies of the apodosis (apo) and protasis (pro) per TAM type. Next, we have a unidirectional measure of association known as  $\Delta P$  (Ellis 2006), taken both from the clause type to the TAM marker ( $\Delta P_{type \rightarrow TAM}$ ) and from the TAM marker to the

clause type ( $\Delta P_{\text{TAM} \rightarrow \text{type}}$ ).  $\Delta P$  describes the relationship between cues and outcomes. In the present study, cues and outcomes may alternatively be defined as values of the TAM or clause-type variables.  $\Delta P$  equals 0 when the cue is unrelated to the outcome. It approaches 1 as the cue and outcome are positively related (cue predicts presence of the outcome) and -1 as they are negatively related (cue predicts absence of the outcome).<sup>14</sup> Bidirectional relationships are indicated by similar values of  $\Delta P_{\text{type} \rightarrow \text{TAM}}$  and  $\Delta P_{\text{TAM} \rightarrow \text{type}}$ . Finally, we have the  $p$ -value produced by the Fisher-Yates exact test, along with the clause type that is distinctively associated with TAM marker.

First, we see that non-actualized TAM markers are significantly preferred by the apodosis cross-linguistically. As illustrated by the values of  $\Delta P$ , this relationship is largely bidirectional, meaning that non-actualized markers and apodoses are mutually strong cues of one another. Second, we see that unmarked status is strongly associated with protasis (unsurprising given the absence of any languages with unmarked apodoses in the sample). In this case, unmarked status is a much stronger predictor of clause type than the other way around.

### 3.3. Discussion

Morphological (a)symmetry is best predicted by the types of TAM values in the protasis and apodosis. This is to be expected, as (a)symmetry is defined relative to these properties. However, the result has two interesting implications. First, morphological symmetry between clauses is more common for languages with mixed or non-actualized protases. Second, none of the other variables was necessary to achieve a high degree of accuracy in predicting (a)symmetry. However, for certain TAM configurations, one can readily reconstruct the corresponding values of macro-area and clause-linking strategies. For instance, Papunesian languages with actualized apodoses tend to be encoded by juxtaposed clause linkage.

The specific TAM affinities of protasis and apodosis are instructive about the general semantics of the counterfactual conditional construction. For example, protases tend to be morphologically unmarked whereas apodoses tend to occur with non-actualized morphology. While an unmarked clause offers no information about its relationship to reality, clauses marked with non-actualized morphology explicitly assert the non-reality of the corresponding state of affairs. Moreover, the (a)symmetry of the overall system is best discriminated by splitting unmarked and actualized from mixed or non-actualized protases (Figure 1). When the protasis is unmarked or actualized, the outcome is almost certainly an asymmetric system. Conversely, when the protasis may occur with

mixed or non-actualized morphology, the likelihood that the overall system will be symmetric increases dramatically. Therefore, (a)symmetry seems to be a product of protases that behave like apodoses rather than the other way around. In other words, symmetrical systems tend to be those that treat the entire counterfactual conditional construction as ungrounded or hypothetical. Asymmetrical systems tend to be those which afford special status (either grounded or unmarked) to the protasis while leaving the apodosis non-actualized. The former kind of system is consistent with the overall meaning of the counterfactual construction. The latter system is not, in principle, though it does follow a certain logic. Protases serve as the background against which apodoses are evaluated. Marking them as actualized grounds them conceptually, hence treating them as a sort of given. Even though neither situation is asserted to have occurred in reality, the situation encoded by the apodosis is treated as contingent on a world in which whatever is expressed in the protasis did in fact occur. Actualized morphology thus anchors the protasis to an imagined world as a precedent for the apodosis.

#### *4. General discussion*

This paper set out to describe the cross-linguistic diversity of counterfactual conditionals by taking into account three parameters, viz. the symmetric and asymmetric morphological patterns of counterfactual conditionals, the range of TAM values that tend to appear in the protasis and apodosis in counterfactual conditional constructions, and the range of clause-linking devices used in counterfactual conditionals.

Through two statistical analyses, we find that morphological (a)symmetry depends most strongly on the properties of the protasis rather than the apodosis of the counterfactual conditional construction. In particular, languages with unmarked or actualized protases contain almost exclusively asymmetric languages. Regarding languages with mixed or non-actualized protases, they overwhelmingly prefer symmetric marking. Another finding is that non-actualized TAM markers are significantly preferred by the apodosis cross-linguistically, while the unmarked status is strongly associated with protasis.

After having explored counterfactual conditionals by taking into account a genetically and areally balanced sample, the following step is to explore particular large genera for which we could only take into account one language (e.g. Oceanic). This will enable us to explore internal diversity and try to come up with more fine-grained typological generalizations.

### *Abbreviations*

1, 2, 3 = first, second, third person; ABL = ablative; ABS = absolutive; ACC = accusative; ACT = active; AD = adessive; ANIM = animate; ANT = anterior; AOR = aoristic; ART = article; ASP = aspect; ASR = assertive; ASSOC = associative; ATN = focus of attention; ATT = attenuative; ATTR = attributive; AUX = auxiliary; AV = actor voice; BP = body part; CAUS = causative; CF = counterfactual; CFP = clause-final particle; CIRC = circumstantial; CNN = connective; COM = communal aspect; COMMIT = comitative; COMP = complementizer; COMPL = completive; COND = conditional; CONJ = conjunction; CONNEG = connegative; CONS = consequential; COP = copula; CVB = converb; DAT = dative; DECL = declarative; DEF = definite; DEIC = deictic; DEM = demonstrative; DET = determiner; DES = desiderative; DIR = directional; DIST = distal; DISTR = distributive; DR = bivalent direct; DS = different subject; DU = dual; DUR = durative; DYNM = dynamic; EMPH = emphatic; EP = epenthetic; ERG = ergative; ES = echo subject; EV = evidential; EVENT = eventive; EX = extended; EXCL = exclusive; F = feminine; FACT = factual; FIN = finite; FOC = focus; FRUST = frustrative; FUT = future; G = general; GEN = genitive; GEND = gender; HAB = habitual; HYP = hypothetical; I = agreement prefix of agreement pattern one; IGN = ignorative; IMAG = imaginative; IMP = imperative; IMPER = impersonal; IMPERF = imperfect; INAN = inanimate; INCH = inchoative; INCL = inclusive; IND = indicative; INDF = indefinite; INF = infinitive; INFR = inferential; INH = inherent; INS = instrumental; INT = intentional; INTR = intransitive; IPD = impeditive; IPFV = imperfective; IRR = irrealis; LIM = limiter; LK = linker; LOC = locative; LOCZR = localizer; M = masculine; MID = middle; MIN = minimal number; MOD = modal; MV = medial verb; NB = notable information; NEG = negative; NMLZ = nominalizer; NOM = nominative; NPPR = personal pronominal; NR = near; OBL = oblique; OBJ = object; OPT = optative; PASS = passive; PAT = patient; PERF = perfect; PFV = perfective; PL = plural; POL = polarity; POSS = possessive; POST = postposition; POT = potential; PRES = present; PRET = preterite; PROG = progressive; PROSP = prospective; PST = past; PTCP = participle; PUNCT = punctual; QUAL = qualitative predication; RDP = reduplication; REAL = realis; REC = recent; REF = referential; REFL = reflexive; REGR = regressive; REIT = reiterative; REL = relativizer; REM = remote; REP = reportative; RES = resultative; RLD = realized; RSG = resigned; SBJ = subject; SEQ = sequential; SG = singular; SIM = simultaneous; SS = same subject; SUB = subordinator; SUBJ = subjunctive; SV = serial verb; TEMP = temporal; TERM = terminative; THEM = thematic; TOP = topic; TRANS = transitive; UAUGM = unit augmented number; UNACC = unaccusative; UNSPEC = unspecified; VAL = validational; VER = veridical.

### *Acknowledgements*

Many thanks to Peter Arkadiev, Marianne Mithun, Bernard Comrie, and two anonymous reviewers for their comments. Any errors remain our responsibility.

## Notes

<sup>1</sup> As correctly pointed by one reviewer, the notion of state of affairs is a useful concept in that it is used unambiguously as a hyponym of different classes of predicates such as situations, actions, events, and processes (see Dik 1997:105).

<sup>2</sup> “[A] genus is a [maximal] group of languages whose relatedness is fairly obvious without systematic comparative analysis” (Dryer 2013, slightly modified).

<sup>3</sup> The present study is in line with other typological research that has also adopted the genealogical and areal stratification proposed by Dryer (2013) without following the procedure(s) he adopts in Dryer (1989). Some of these typological studies can be found in Miestamo (2005) and Shagal (2019), to name but a few.

<sup>4</sup> The Genus-Macroarea sampling method involves different samples or levels of sampling: the Genus Sample, the Core Sample, the Restricted Sample, and the Extended Sample. Their selection depends on the type of research question(s).

<sup>5</sup> Haspelmath (2010: 664) explains that comparative concepts are concepts created by comparative linguists for the specific purpose of cross-linguistic comparison. They are based on universal conceptual-semantic concepts and universal formal concepts. As pointed out by one reviewer, it should be noted that comparative concepts were developed much earlier, and have been used by typologists for at least the past three decades (e.g. Stassen 1985: 14). This approach implies that any language should have a means to encode particular conceptual state of affairs, though not necessarily a dedicated or a grammaticalized one.

<sup>6</sup> It is important to bear in mind that the literature on complex sentence constructions is vast. However, only a few studies have provided an explicit definition of what a complex sentence construction is.

<sup>7</sup> We refer the interested reader to Van Valin & LaPolla (1997) on subordination, coordination and co-subordination, and Yuasa & Sadock (2002) on pseudo-subordination.

<sup>8</sup> In a similar fashion, bound adverbial subordinators also operate in subordinate clauses with various properties. However, in comparison to free adverbial subordinators, it is not infrequent to observe bound adverbial subordinators operating in clauses with properties similar to those found in main clauses. In this regard, Hetterle (2015: 108) mentions that fully inflected verbs are not as rare as one might suspect. In her typological study, she mentions that in 38 of the 164 constructions with a bound adverbial subordinator, the verb of such a construction is fully inflected and identical to a main clause verb.

<sup>9</sup> The term *para-hypotaxis* is used by Romance linguists to refer to sentences containing a dependent clause with the main clause introduced by a coordinative conjunction. According to Bertinetto & Ciucci (2012) this term was traditionally considered as an idiosyncratic feature of Old Romance languages.

<sup>10</sup> Note that asymmetric counterfactual conditionals may also include instances in which the protasis is deprived of TAM marking while the apodosis appears with particular TAM values (e.g. Warekena, Macushi, Pashto, and Kam, among others; cf. Qian 2016: 158) or instances in which the protasis appears with particular TAM values and the apodosis is deprived of TAM marking (e.g. Savosavo and Yimas; cf. Qian 2016: 158).

<sup>11</sup> As correctly pointed by one reviewer, nominalized verb forms must be considered a subset of actualized and non-actualized patterns because nominalized verb forms can be used for both actualized and non-actualized state of affairs.

<sup>12</sup> The results of the CART analysis were verified by means of a random-forests analysis. We present the CART results because they offer a clearer perspective on how the different predictors are weighted with respect to their reliability in partitioning the data.

<sup>13</sup> Indeed, we attempted logistic regressions with both generalized linear and generalized additive models. While the results are similar to those presented in the CART analysis, the models performed quite poorly.

<sup>14</sup> We only provide the absolute values here because we have just two outcomes: negative values indicate association with protasis, while positive values indicate association with apodosis.

### *Bibliographical References*

- Arkadiev, Peter 2020. Actionality, aspect, tense, and counterfactuality in Kuban Kabardian. *Studia Orientalia Electronica* 8. 5-21.
- Bennett, Charles E. 1908. *A Latin grammar*. Boston / Chicago: Allyn and Bacon.
- Bertinetto, Pier Marco & Ciucci, Luca 2012. Parataxis, hypotaxis and para-hypotaxis in the Zamucoan languages. *Linguistic Discovery* 10. 89-111.
- Bhatt, Rajesh 1998. *CF marking in the modern Indo-Aryan languages*. Paper presented at the University of Konstanz.
- Black, Cheryl A. 1994. *Quiégolani Zapotec syntax*. PhD dissertation. University of California, Santa Cruz.
- Blackings, Mairi & Fabb, Nigel 2003. *A grammar of Ma'di*. Berlin / New York: Mouton de Gruyter.
- Brooks, Joseph 2018. *Realis and irrealis: Chini verb morphology, clause chaining, and discourse*. PhD dissertation. University of California, Santa Barbara.
- Bugenhagen, Robert D. 1995. *A grammar of Mangap-Mbula: An Austronesian language of Papua New Guinea*. Notes on Linguistics. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Chang, Anna Hsiou-chuan 2006. *A reference grammar of Paiwan*. PhD dissertation. Australian National University, Canberra.
- Comrie, Bernard 1986. Conditionals: A typology. In Traugott, Elizabeth; ter Meulen, Alice; Reilly, Judy & Ferguson, Charles (eds.), *On conditionals*. Cambridge: Cambridge University Press. 77-99.
- Craig, Grinevald Colette 1990. *A grammar of Rama*. Report to National Science Foundation.
- Croft, William 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William 2003. *Typology and universals*, 2<sup>nd</sup> ed. Cambridge: Cambridge University Press.
- Dik, Simon 1997. *The theory of functional grammar. The structure of the clause*. Berlin / New York: Mouton de Gruyter
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257-292.
- Dryer, Matthew S. 2013. Genealogical language list. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27. 1-24.
- Foley, William A. 1991. *The Yimas language of Papua New Guinea*. Stanford: Stanford University Press.
- Fox, Greg J. 1979. *Big Nambas grammar*. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- François, Alexandre 2002. *Araki: A disappearing language of Vanuatu*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Gallagher, Steve & Baehr, Pierce 2005. *Bariai grammar sketch*. Data Papers on Papua New Guinea Languages. Ukarumpa: Summer Institute of Linguistics.
- Gast, Volker & Diessel, Holger 2012. The typology of clause linkage: Status quo,

- challenges, prospects. In Gast, Volker & Diessel, Holger (eds.), *Clause linkage in cross-linguistic perspective: Data-driven approaches to cross-clausal syntax*. Berlin / New York: Mouton de Gruyter. 1-36.
- Givón, Talmy 2001. *Syntax: An introduction: Volume 2*. Amsterdam / Philadelphia: John Benjamins.
- Greenberg, Joseph 1966. *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.
- Gries, Stefan & Stefanowitsch, Anatol 2004. Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9. 97-129.
- Haiman, John 1980. The iconicity of grammar. *Language* 56. 515-540.
- Haiman, John 1983. Paratactic *if*-clauses. *Journal of Pragmatics* 7. 263-281.
- Haiman, John & Kuteva, Tania 2002. The symmetry of counterfactuals. In Bybee, Joan & Noonan, Michael (eds.), *Complex sentences in grammar and discourse*. Amsterdam / Philadelphia: John Benjamins. 101-124.
- Harvey, Mark 2002. *A grammar of Gaagudju*. Berlin / New York: Mouton de Gruyter.
- Haspelmath, Martin 1995. The converb as a cross-linguistically valid category. In Haspelmath, Martin & König, Ekkehard (eds.), *Converbs in cross-linguistic perspective*. Berlin / New York: Mouton de Gruyter. 1-55.
- Haspelmath, Martin 2004. Coordinating constructions: An overview. In Haspelmath, Martin (ed.), *Coordinating constructions*. Amsterdam / Philadelphia: John Benjamins. 3-39.
- Haspelmath, Martin 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86. 663-687.
- Haude, Katharina 2006. *A grammar of Movima*. PhD dissertation. Radboud Universiteit, Nijmegen.
- Heath, Jeffrey & Hantgan, Abbie 2018. *A grammar of Bangime: Language isolate of Mali*. Berlin / Boston: De Gruyter Mouton.
- Hetterle, Katja 2015. *Adverbial clauses in cross-linguistic perspective*. Berlin / Boston: De Gruyter Mouton.
- Hill, Deborah 1992. *Longgu grammar*. PhD dissertation. Australian National University, Canberra.
- Lynch, John 2000. *A grammar of Anejom*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Nichols, Johanna 2011. *Ingush grammar*. Berkeley: University of California Press.
- Karawani, Hadil 2014. *The real, the fake, and the fake fake in counterfactual conditionals, crosslinguistically*. Utrecht: Landelijke Onderzoekschool Taalwetenschap, Netherlands National Graduate School of Linguistics.
- Kenesei, István; Vago, Robert M. & Fenyvesi, Anna 1998. *Hungarian*. London: Routledge.
- Klinken, Catharina Lumien van 1999. *A grammar of the Fehan Dialect of Tetun, an Austronesian language of West Timor*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Kortmann, Bernd 1997. *Adverbial subordination: A typology and history of adverbial subordinators based on European languages*. Berlin / New York: Mouton de Gruyter.
- Lastra, Yolanda 1989. *Otomí de San Andrés Cuexcontitlán*. Mexico: El Colegio de México.

- Lehmann, Christian 1988. Towards a typology of clause linkage. In Haiman, John & Thompson, Sandra A. (eds.), *Clause combining in discourse and grammar*. Amsterdam / Philadelphia: John Benjamins. 181-225.
- Levy, Paulette 1990. *Totonaco de Papantla, Veracruz*. México: Centro de Investigación para la Integración Social.
- Lindstrom, Lamont & Lynch, John 1994. *Kwamera*. München: Lincom.
- Longacre, Robert E. 1985. Sentences as combinations of clauses. In Shopen, Timothy (ed.), *Language typology and syntactic description*. Cambridge: Cambridge University Press. 235-286.
- Mauri, Caterina & Sansò, Andrea 2009. *Irrealis and clause linkage*. Paper presented at the 8<sup>th</sup> Biennial Meeting of the Association of Linguistic Typology, Berkeley.
- Mauri, Caterina & van der Auwera, Johan 2012. Connectives. In Keith, Allan & Jaszczolt, Kasia M. (eds.), *The Cambridge handbook of pragmatics*. Cambridge: Cambridge University Press. 347-402.
- McGregor, William 1990. *A functional grammar of Gooniyandi*. Amsterdam / Philadelphia: John Benjamins.
- Merlan, Francesca C. 1994. *A grammar of Wardaman, a language of the Northern territory of Australia*. Berlin / New York: Mouton de Gruyter.
- Miestamo, Matti 2005. *Standard negation: The negation of declarative verbal main clauses in a typological perspective*. Berlin / New York: Mouton de Gruyter.
- Miestamo, Matti; Bakker, Dik & Arppe, Antti 2016. Sampling for variety. *Linguistic Typology* 20. 233-296.
- Mithun, Marianne 1995. On the relativity of irrealis. In Bybee, Joan & Fleischman, Suzanne (eds.), *Modality in grammar and discourse*. Amsterdam / Philadelphia: John Benjamins. 367-388.
- Nicolle, Steve 2017. Introduction to special issue on conditional constructions in African languages. *Studies in African Linguistics* 46. 1-15.
- Noonan, Michael 1992. *A grammar of Lango*. Berlin / New York: Mouton de Gruyter.
- Olawsky, Knut J. 2006. *A grammar of Urarina*. Berlin / New York: Mouton de Gruyter.
- Olguín Martínez, Jesús 2016. *Adverbial clauses in Veracruz Huasteca Nahuatl from a functional-typological approach*. MA thesis. University of Sonora, Hermosillo.
- Olson, Clif 1992. Gumawana (Amphlett Islands, Papua New Guinea): Grammar sketch and texts. In Ross, Malcolm D. (ed.), *Papers in Austronesian Linguistics* 2. Canberra: Research School of Pacific and Asian Studies, Australian National University. 251-430.
- Overall, Simon 2017. *A grammar of Aguaruna (Iiniá Chicham)*. Berlin / Boston: De Gruyter Mouton.
- Prince, Kilu von 2015. *A grammar of Daakaka*. Berlin / Boston: De Gruyter Mouton.
- Prince, Kilu von 2019. Counterfactuality and past. *Linguistics and Philosophy* 42. 577-615.
- Qian, Yong 2016. *A typology of counterfactual clauses*. PhD dissertation. City University of Hong Kong.
- Sanders, Arden & Sanders, Joy 1994. *Kamasau (Wand Tuan) grammar: Morpheme to sentence*. Ms.
- Seiler, Walter 1985. *Imonda, a Papuan language*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Shagal, Ksenia 2019. *Participles: A typological study*. Berlin / Boston: De Gruyter

Mouton.

- Smyth, H. W. 1920. *Greek grammar for colleges*. New York: American Book Company.
- Stassen, Leon 1985. *Comparison and universal grammar*. Oxford: Basil Blackwell.
- Steele, Susan 1975. Past and irrealis: Just what does it all mean? *International Journal of American Linguistics* 41. 200-217.
- Tharp, Douglas 1996. Sulka grammar essentials. In Clifton, John M. (ed.), *Two non-Austronesian grammars from the islands*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics. 77-179.
- Traugott, Elizabeth C. 1985. On conditionals. In Haiman, John (ed.), *Iconicity in syntax*. Amsterdam / Philadelphia: John Benjamins. 289-307.
- Valin, Robert Jr. van & LaPolla, Randy 1997. *Syntax*. Cambridge: Cambridge University Press.
- Van den Berg, Helma 1995. *A Grammar of Hunzib (with Texts and Lexicon)*. München: Lincom.
- Van de Velde, Mark 2008. *A grammar of Eton*. Berlin / New York: Mouton de Gruyter.
- Verstraete, Jean-Christophe 2014. The role of mood marking in complex sentences: A case study of Australian languages. *Word* 57. 195-236.
- Vries, Lourens de 2004. *A short grammar of Inanwatan: An endangered language of the Bird's head of Papua, Indonesia*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Vuillermet, Marine 2018. Grammatical fear morphemes in Ese Ejja: Making the case for a morphosemantic apprehensional domain. In Ponsonnet, Maïa & Vuillermet, Marine (eds.), *Morphology and emotions across the world's languages*. Special issue of *Studies in language* 42. 256-293.
- Wierzbicka, Anna 1997. Conditionals and counterfactuals: Conceptual primitives and linguistic universals. In Athanasiadou, Angeliki & Dirven, René (eds.), *On conditionals again*. Amsterdam / Philadelphia: John Benjamins. 15-59.
- Wilkins, David P. 1989. *Mparntwe Arrernte (Aranda): Studies in the structure and semantics of grammar*. PhD dissertation. Australian National University, Canberra.
- Wulff, Stefanie; Lester, Nicholas & Martinez Garcia, Maria 2014. That-variation in German and Spanish L2 English. *Language and Cognition* 6. 271-299.
- Xrakovskij, Viktor 2005. Conditional constructions: A theoretical description. In Xrakovskij, Viktor (ed.), *Typology of conditional constructions*. München: Lincom. 3-95.
- Yuasa, Etsuyo & Sadock, Jerry M. 2002. Pseudo-subordination: A mismatch between syntax and semantics. *Journal of Linguistics* 38. 87-111.

