Universiteit
Leiden
The Netherlands

Differential case-marking in Ecuadorian Siona

Case, J.T.

Citation
Case, J. T. (2026, June 30). *Differential case-marking in Ecuadorian Siona*. *LOT dissertation series*. LOT publishers, Amsterdam, the Netherlands. Retrieved from https://hdl.handle.net/1887/4307357

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4307357

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 1

Introduction

Case-marking morphology is commonly recruited in natural language to indicate the role of a noun in the sentence (Nichols 1986, 1992; Haspelmath 2010; Nichols and Bickel 2013). The classical Latin examples in (1) are a case in point, where the form of each nominal is determined by its role in relation to the predicate. Following terminology in the Latin descriptive tradition, the SUBJECT argument takes the *nominative* form, whereas the DIRECT OBJECT takes the *accusative* form:

(1) Latin (*lat* — INDO-EUROPEAN, ITALIC) [Haegeman 1991, 143]

a. Caeser$^{(\text{SUBJECT})}$ Belgas$^{(\text{OBJECT})}$ vinc-it
C.NOM Belgians.ACC beat-3SG
'Caesar beats the Belgians.'

b. Belgae$^{(\text{SUBJECT})}$ Caeserem$^{(\text{OBJECT})}$ tem-ent
Belgians.NOM C.ACC fear-3PL
'The Belgians fear Caesar.'

Many languages utilize a case-marking strategy like (1) to encode grammatical relations, and various semantic relations (Simpson 2023). In fact, Iggesen (2013) finds that just over half of the world's languages exhibit some form of case-marking.[1] Case-marking languages vary widely as to the morphosyntactic expression of case-markers, and the number of case-marking categories: ranging from 'two-term systems' to languages with rich inventories exceeding a dozen categories (Blake 2001; Arkadiev 2008). Setting aside discrepancies in inventory size, languages map these categories onto a relatively stable set of relations (cf. Haspelmath 2010 and references therein).

[1] According to Iggesen (2013, *WALS*, chapter 49), 164 languages in his diverse 264-specimen sample (62.12%) have at least two case categories. The remaining 100 languages use another strategy to encode grammatical relations.

This dissertation is concerned with a particular case-marking phenomenon whereby not all nouns with a given relation are marked in the same way. Instead, factors beyond the grammatical or semantic relation itself determine the distribution of certain morphological case-markers without affecting the grammaticality of the sentence: e.g., the animacy status or the specificity status of the referent. This phenomenon is generally called DIFFERENTIAL CASE MARKING (henceforth DCM) (Moravcsik 1978; Bossong 1985; Aissen 1999, 2003). An illustrative instance of a DIRECT OBJECT-oriented DCM pattern is demonstrated in the minimal pair from the Turkic language Sakha in (2). Overt case-marking is found on the specific OBJECT in (2b), but not on its non-specific counterpart in (2a):

(2) Sakha (*sah* — TURKIC)
[Vinokurova 2005, cited in Baker 2015, 1, ex. (1c); 4, ex. (5)]

a. Erel **kinige** atyylas-ta
E.NOM book buy-PST.3SS
'Erel bought a book/books.'

b. Erel **kinige-ni** atyylas-ta
E.NOM book-ACC buy-PST.3SS
'Erel bought the book.'

In a recent survey, Sinnemäki (2014) finds that case-marking languages are more likely to exhibit DCM than obligatory case-marking like that shown for Latin in (1).

DCM patterns are not restricted to DIRECT OBJECTS, as in the Sakha pattern shown in (2). Such patterns are also attested on other core arguments, on spatial arguments, and even on certain oblique arguments (e.g., Chappell and Verstraete 2019; Haspelmath 2021). In fact, a single language may display several DCM patterns, each with their own characteristics. DCM has garnered significant scholarly interest both in typological studies and in detailed, language-particular descriptions (Blake 2001; Klein and de Swart 2011; Baker 2015). A more detailed review and discussion of each of these aspects is undertaken in Chapter 2.

1.1 DCM patterns attested in Siona

This dissertation provides an in-depth description and analysis of the various DCM patterns found in the Ecuadorian variety of the Siona language [*snn* – Western Tukanoan, henceforth Siona]. The analysis developed in these pages complements and expands significantly upon the initial descriptive generalizations regarding the distribution of Siona case-markers sketched in Bruil (2014, §4.4).[2] The present dissertation identifies five primary DCM patterns in the language as shown in Table 1.1 — i.e., labeled the *principled* DCM alternations in Chapter 3. Alongside a brief description of each grammatical relation displaying DCM, the table below provides a relation label for each DCM pattern, used as a short-hand convention throughout this dissertation:

[2]Specific dimensions of Siona DCM patterns are expanded upon in subsequent work: e.g., Case and Jeretič 2021.

Grammatical relation label (used in this dissertation)	Traditional Latin label	Brief prose description
Subject (S)	*nominative*	The single argument of an intransitive clause and the Subject-argument of the (di)transitive clause (i.e., S=A).
Direct object (P)	*accusative*	The direct Object argument in the (di)transitive clause (i.e., P=T).
Indirect object (R)	*dative*	The indirect Object argument in the ditransitive clause. Other dative-like arguments are identified separately.
Static location (L)	*locative*	The static location implicated in the basic locative construction (X is (at) Y), and with other predicates which do not implicate motion.
Spatial goal (G)	*allative*	The dynamic location implicated in directed motion predicates (and change-of-posture predicates), i.e., GOAL.

Table 1.1: Short-hand labels for Siona DCM RELATIONS with brief description

Siona displays straightforward nominative-accusative alignment, like all Tukanoan languages do (Barnes 1999, 2006; Gomez-Imbert 2011; Stenzel 2013d). However, Siona is rather unusual in that it displays DCM on *all* core argumental grammatical relations (i.e., S, P, and R), and on certain relations in the spatial domain (i.e., L and G). A subset of these DCM patterns formally resemble the Sakha pattern shown in (2), where overt case-marking is not strictly required to encode a particular grammatical relation. For instance, the elicited minimal pair in (3) illustrates what might be called the optional *nominative*-marking pattern in Siona:

(3) a. *jaiye joro aide'oji iño*
hai-je **horo** ai + de'o-hi ihño
big-CL:GEN flower big + become-3S.M.PRS.ASS here
'Many flowers grow here.'

b. [**Context:** The speaker conveys that many kinds of flowers grow in this particular location, although other plants (trees, grass) do not.]
jaiye jorobi aide'oji iño
hai-je **horo #(-bi)** ai + de'o-hi ihño
big-CL:GEN flower-SBJ big + become-3S.M.PRS.ASS here
'Many FLOwers grow here (i.e., not something else).'
[VOL: 20230622ejabi001.012a-c]

Unlike the Sakha pattern, where case-marking is associated with a specific reading, overt case-marking in (3b) evokes an interpretative effect whereby the referent is contrasted with salient alternative referents in the context. This effect is related to information structure, where the speaker indicates which referents merit the listener's attention as the discussion or the narrative unfolds. This dissertation labels this the *emphatic contrast* reading, following similar effects described in the Tibeto-Burman optional ergative literature (e.g., Tournadre 1995; LaPolla 1995). A comparable system of case-marking is described for the Yalaku language of Papua New Guinea by Aikhenvald (2015), under the label 'highlighted participant case'. The emphatic contrast effect is observed across each of the DCM patterns enumerated in Table 1.1. Consider the spatial DCM pattern shown in (4) — zero-marking on the locative (L)

argument in (4a) produces a plain reading, whereas the overt *-re* marking in (4b) produces the emphatic contrast reading:

(4) a. *yë' yo'jei Canada ba'iji*
ji̵'i̵ jo'he-i̵ **Canada** ba'i-hi
1SG younger.sibling-CL:M C-N.SBJ live-3S.M.PRS.ASS
'My younger brother lives <u>in Canada</u>.'

b. [**Context:** If you pull out a map of Ecuador to show me where your brother lives, except he does not live in Ecuador.]
yë' yo'jei Canadare ba'iji
ji̵'i̵ jo'he-i̵ **Canada #(-re)** ba'i-hi
1SG younger.sibling-CL:M C-N.SBJ live-3S.M.PRS.ASS
'My younger brother lives <u>in CAnada</u> (i.e., not some other place).'
[VOL/SUG: 20230619elupa002.011a-b]

The minimal pairs in (3) and (4) demonstrate the information structure-driven usage of overt case-marking across various grammatical relations. This particular function of case-marking is not acknowledged in the previous literature, but is shown to be one of the primary drivers of Siona DCM alternations in this dissertation. This is not a particularly common trigger for DCM in the typological literature — see the discussions in Chapter 2 and Chapter 4.

In addition to the information structure-driven patterns of DCM demonstrated above, this dissertation argues for their interaction with an animacy-based SUBJECT vs. non-SUBJECT divide in the grammar. Put simply, animate non-SUBJECT noun phrases require overt case-marking, whereas inanimates do not. The result is demonstrated for DIRECT OBJECT-oriented DCM in (5) and (6). In the former case, concerning an animate DIRECT OBJECT, a different type of formal DCM is attested where two overt case-markers compete, producing comparable interpretative effects to those noted above. Zero-marking on animate nouns is judged as ungrammatical. Conversely, inanimate DIRECT OBJECTS, like that shown in (6), display the familiar zero- vs. overt case-marking alternation:

(5) a. *huajë gajeire co'esi' cayë*
wahi̵ + gahe-i̵ *(-re) ko'e-si'i kaa-ji̵
new + friend-CL:M-N.SBJ look_for-FUT.N3S say-N3S.PRS.ASS
'I want to look for <u>a new friend</u>.'

b. [**Context:** In a discussion about buying new animals for the house, the speaker would rather find a new friend.]
huajë gajeini co'esi' cayë
wahi̵ + gahe-i̵ *(-ni)/(#-re) ko'e-si'i kaa-ji̵
new + friend-CL:M-N.SBJ2 look_for-FUT.N3S say-N3S.PRS.ASS
'I want to look for <u>a new FRIEND</u>.' [VOL/SUG: 20230619elupa002.018a-c]

(6) a. *huajë jachohuë huerosi' cayë*
wahi̵ + hahcho-wi̵ wero-si'i kaa-ji̵
new + shoot-CL:CONT buy-FUT.N3S say-N3S.PRS.ASS
'I want to buy <u>a new rifle</u>.'

b. [**Context:** At the store, the salesperson tries to sell you an ax.]
huajë jachohuëre huerosi' cayë
wahɨ + hahcho-wɨ #(-re) wero-si'i kaa-jɨ
new + shoot-CL:CONT buy-FUT.N3S say-N3S.PRS.ASS
'I want to buy a new RIfle (i.e., not something else).'
[VOL/SUG: 20230619elupa002.020a-b]

The same *-re* vs. *-ni* alternation shown in (5) is found on Recipients, Experiencers, and Benefactee arguments. For this reason, this dissertation opts for the neutral labeling convention for these case-markers: *-re* (N.SBJ, as NON-SUBJECT) and *-ni* (N.SBJ2), following several Amazonianists (e.g., Aikhenvald 1999, 2007; Farmer 2015; or the related OBJECTIVE label in Zúñiga 2007, *et seq.*).[3] In fact, this dissertation recognizes instances of spatial *-ni*, which are not recognized in earlier descriptions. In the bulk of instances, *-ni* is associated with the emphatic contrast reading as anticipated. This is shown the naturalistic sentences in (7):

(7) a. *më'bi yë'ni ñañe bahuë*
mɨ'ɨ-bi **jɨ'ɨ-ni**$^{(P)}$ ñaa-ñe bãã-wɨ
2SG-SBJ 1SG-N.SBJ2 see-CL:GEN NEG.AUX-N3S.PST.ASS
'YOU didn't see ME (i.e., not vice versa).'
[NAT: 20140925salsu001.032]

b. *... si'ahua'i jaoni bateña*
si'a-wa'i **hã-o-ni**$^{(L)}$ ba-re-ña
all-PL.AN DEM.MED-CL:F-N.SBJ2 be-N2/3S.PST.N.ASS-REP
'... everybody was THERE (they say).' [NAT*: 20151023orocr001.112]

The present work addresses the rich set of DCM patterns found in Siona grammar by establishing the formal patterns at play, and the interpretative effects which are evoked by particular case-markers (or the lack thereof) in a given context. In previous work (e.g., Bruil 2014, §4.4), generalizations are put forth regarding the effect of other possible factors for certain DCM patterns — for example, the specificity-driven usage of *-re* on DIRECT OBJECTS, or ambiguity avoidance-based usage of SUBJECT *-bi* marking. On the assumption that several factors may conspire to account for the actual distribution of case-markers in Siona speech, these hypotheses are scrutinized in this dissertation.

1.2 Research questions and dissertation outline

With the aim of adequately characterizing Siona DCM patterns, and of fitting these patterns into the broader typology, this dissertation addresses the research questions laid out in (8):

(8) a. Which DCM patterns can be identified in Siona? What properties do these DCM patterns have in common, and how do they differ?

[3] Other Tukanoanists, including Bruil 2014, *et seq.*; Johnson and Levinsohn 1990; Stenzel 2008, 2013c; etc. utilize the label OBJ (similar to SBJ for the S-marker), which indicates the dominant marking function of this case-marker. See Stenzel 2008, p. 175, for a brief discussion concerning *-re* and the 'labeling issue'.

b. Which factors determine the selection of case marking alternatives in a given context — such as animacy status, focus, specificity? A combination thereof? How do the active factors interact to determine the DCM patterns in usage?

c. How does Siona DCM compare to other DCM systems described in the Tukanoan literature? In the northwestern Amazonia area? In the broader DCM typology?

In order to adequately address the questions in (8), a diverse methodological program is followed. This dissertation espouses the view that linguistic research is strengthened where diverse, corroborating evidence is made available (e.g., Polinsky 2010; Aihenvald 2018; Davidson 2020; Matthewson 2004, 2022 for recent discussions). I wear several linguistic hats, so to speak, in the different chapters of this dissertation, with the aim of avoiding 'methodological monotheism', following Matthewson (2022, 21), who acknowledges Farrell Ackerman for the term. A general outline of this dissertation is provided below.

Chapter 2 establishes the working definitions and terminology utilized in this work, and navigates certain dimensions of the robust DCM literature in the form of a state-of-the-art. Chapter 3 provides the *preliminary description* of Siona DCM, establishing the empirical facts and framing the various targeted studies presented in the subsequent chapters of this dissertation. Several novel contributions are made here — such as the abstract PLAIN-PROMINENT dichotomy, permitting a paradigmatic analysis, and the description of the noun class-shifting procedure of *promotion*, which forms the basis of the analysis of grammatical animacy classes. The baseline analysis of Siona DCM is established in the *preliminary description* is tested and expanded in the following three content chapters, conceived as independent studies.

For Chapter 4, I put on my *comparativist* hat. This chapter develops a typological study, which serves to situate the baseline Siona DCM patterns laid out in Chapter 3 among other (non-)Tukanoan languages spoken in the putative NWA area. The discussion in this chapter is built around a typological dataset − i.e., the NWA case-marking dataset, assembled in Appendix A. This study gathers various case-marking and DCM facts via the administration of a structural questionnaire (e.g., as found in Krasnoukhova 2012; Birchall 2014), against a modest sample languages spoken in the NWA area. As such, this chapter is not a component of the language-particular analysis developed later in this dissertation per se; instead it contextualizes these language-specific facts, and lays the groundwork for a diachronic analysis of Siona DCM. To date these dimensions of Siona DCM have not received adequate attention.

For Chapter 5, I put on my *variationist* hat. This chapter outlines a quantitative, corpus-driven investigation of DCM patterns in a modest manually-coded Siona narrative sample. Noun tokens are extracted from the sample, categorized on the basis of their case-marking, and are subsequently annotated for animacy, specificity, topicality, and a handful of other properties in order to determine which are significant predictors for Siona DCM patterns. Two types of quantitative data are considered, relating to the proportion of case-marking alternatives in the corpus sample and to the statistical analyses modeled against the extracted tokens (i.e., via the implementation of the *variable-rule* technique). This approach allows a more nuanced view of the DCM phenomenon in Siona, which had previously only been described qualitatively.

For Chapter 6, I put my *descriptivist* hat back on and complement the *preliminary description* by reporting a range of data, collected via the implementation of a

controlled elicitation-driven research program. In particular, the evidence presented in this chapter highlights the role of Siona DCM in a host of contexts traditionally associated with the expression of focus (cf. van der Wal 2015, 2016, etc.). This chapter refines the description of predictable question-answer case-matching patterns and (emphatic) contrastive uses of case-marking recognized in preceding chapters in the dissertation with data characterizing the nature of information structural DCM effects with a degree of precision lacking from earlier accounts.

Chapter 7 concludes this dissertation. On the one hand, this chapter unifies the diverse array of evidence presented in the preceding chapters and amends the *preliminary description* for Siona DCM accordingly. Additionally, this chapter enumerates a number of promising directions for how the description in this work may be furthered with future research, and for how these phenomena might be analyzed.

The remainder of this introductory chapter is organized as follows: Section 1.3 introduces the Siona language and the people who speak it. That section positions Ecuadorian Siona within the Tukanoan genealogy and within the Amazon area. Section 1.4 outlines the glossing and orthographical conventions utilized for representing Siona data in this dissertation.

1.3 The Siona people and their language

Ecuadorian Siona, or by its endonym *bãĩ̃kohka* (the language of the people), is an endangered language spoken in the province of Sucumbíos in Ecuador. Today between two and three hundred individuals speak this variety with varying degrees of fluency (Bruil, p.c.; Mejeant 2001; Fabre 2005[4]). Bruil (2014, pp. 4-5) recognizes six small Siona communities in Ecuador, found along two converging river systems: Puerto Bolívar and Tarabëaya[5] are found in the Cuyabeno Wildlife Reservation along the Cuyabeno river to the north, and Orehuëaya, Aboquëhuira, Bi̱'a̱ña, and Sototsiaya dot the Aguarico river to the south. All of my own original fieldwork is collected in the community of Sototsiaya, supplemented by data collected online and during a four-month visit by Jamil Biaguaje to the University of Ottawa (September-December 2024). Certain other corpus materials scrutinized in this dissertation, collected by Martine Bruil, were gathered in the community of Puerto Bolívar. The map in Figure 1.1 identifies the two communities where the data for this dissertation was collected.

[4] *Diccionario etnolingüístico y guía bibliográfica de los pueblos indígenas sudamericanos* is freely consultable at http://www.ling.fi/DICCIONARIO.htm

[5] For these place-names, I utilize the native orthography conventions as used by community members. Section 1.4.3 lays out the two spelling conventions in detail.

Figure 1.1: Locations of the two Ecuadorian Siona communities where the data used here was obtained (adapted from Bruil and Stewart 2022)

There are two main motivations for which the community of Sototsiaya is well-suited as the primary site for data collection: Firstly, I am well-acquainted with several people and friends within the community, since first being introduced to them in 2018 by Martine. Secondly, this is the community where Siona enjoys its greatest degree of vitality, spoken by most community members and transmitted cross-generationally to the present day. This is in sharp contrast with other Siona communities where, in most cases, a shift to Spanish is all but complete.

1.3.1 Siona in its genealogical context

The Siona language belongs to the smaller Western branch of the Tukanoan language family. This is a medium-sized family comprising roughly two-dozen living languages, straddling the border regions between modern-day Ecuador, Colombia, Peru, and Brazil. The family is divided into two geographically-compact branches. While the members of the Western branch are strewn primarily within the Putumayo-Napo watershed, the Eastern branch is confined within the expansive Upper Río Negro (also called the Vaupés) cultural area. It is likely that the geographical distribution of each branch has remained relatively stable since before contact with European explorers and settlers (Chacon 2013, 2014). As concerns the Siona and related Western Tukanoan groups, united under the label *encabellados* (long-haired), it is well-

established that these were the dominant populations in and around these same fluvial zones in the early contact and evangelization contexts. Siona became elevated to the status of lingua franca in this area over the seventeenth- and eighteenth-centuries (cf. Wheeler 1970, 194-197; Vickers 1976, 1981; Cipolletti 1999; Carrera de la Red 2020, 2021).

Several internal classifications have been proposed for the Tukanoan family, as based upon comparative phonological and lexical datasets (e.g., Waltz and Wheeler 1972; Barnes 2006; Gomez-Imbert 2011). While the two-way division between Western and Eastern languages has been maintained, the more fine-grained classification is far from settled. There is now consensus that the putative Central Tukanoan branch, suggested in previous decades, ought to be abandoned (cf. Mason 1950; Waltz and Wheeler 1972; Malone 1987; Barnes 1999, 2006). Stenzel (2013d) provides a recent overview:

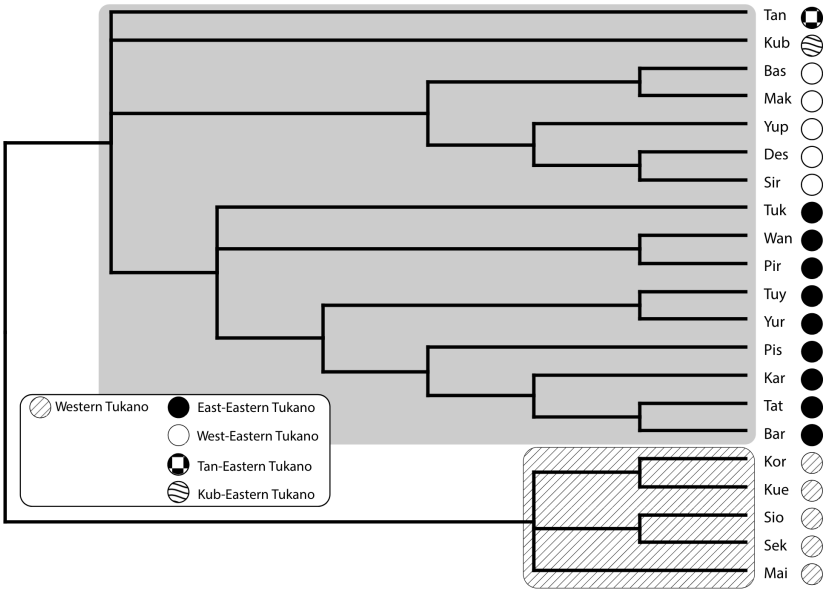

Figure 1.2: Proposed phylogenetic tree model for Tukanoan language family (Chacon and List 2015, 198, Figure 10)

Figure 1.2 reproduces the Tukanoan phylogenetic classification from Chacon and List (2015).[6] One can see that diversification is high in the Eastern branch, whereas the Western branch is more homogeneous. Setting aside the now-extinct language Kueretú (Kue), all specialists agree that Siona and Sekoya are intimately linked on the grounds of lexicostatistics and phonological innovations. A minor controversy exists as to

[6]The short-hand language names used in Figure 1.2 are as follows: Tan (Tanimuka), Kub (Kubeo), Bas (Barasano), Mak (Makuna), Yup (Yupua), Des (Desano), Sir (Siriano), Tuk (Tukano), Wan (Wanano), Pir (Piratapuyo), Tuy (Tuyuka), Yur (Yurutí), Pis (Pisamira), Kar (Karapana), Tat (Tatuyo), Bar (Bará), Kor (Koreguaje), Kue (Kueretú), Sio (Siona), Sek (Sekoya), Mai (Máíhɨ̄kì).

whether Koreguaje is more closely related to Siona-Sekoya varieties (cf. Wheeler 1992; Barnes 1999, 2006; Chacon 2014), or whether Máíhɨ̃kì is closer to these varieties and Koreguaje forms the out-group (Skilton 2013).[7] The former proposition, that Koreguaje is a closer phylogenetic relative to Siona than to Máíhɨ̃kì, is certainly borne out on the basis of the case-grammar facts laid out in this dissertation.

To elaborate slightly on this claim, this dissertation provides further support for claims regarding the development of the Western Tukanoan branch as follows. Based on attestation-type data in contemporary descriptive materials, Bruil (2018, 158-159) proposes that various components of the verbal paradigms can be reconstructed to three distinct stages. My schematization in Table 1.2 expands the original claims made by Bruil (2018):[8]

Shared innovation→	Proto-WT *Interrogative* s*-agree*	> Proto-Kor-Sio-Sek *Dependent verb paradigm*; ***Rich case system***	> Proto-Sio-Sek *Reportative -ña*
Siona (Sio)	✓	✓	✓
Sekoya (Sek)	✓	✓	✓
Koreguaje (Kor)	✓	✓	x
Máíhɨ̃kì (Mai)	✓	x	x

Table 1.2: Developmental chronology of Western Tukanoan (adapting Bruil 2018)

On the basis of this table, the attested DCM phenomena in Ecuadorian Siona analyzed in this dissertation ought to be reconstructed to a Proto-Koreguaje-Siona-Sekoya stage, after the initial split between the Western and Eastern branches. The set of cognate case-markers *-ni*, *-na*, and *-bi* could be reconstructed to the stage labeled *Rich case system* in Table 1.2. By extension, the Máíhɨ̃kì language has a strikingly simple case grammar compared to other Western Tukanoan languages − this comparison is fleshed out in Chapter 4.

On the basis of several shared innovations and a degree of mutual intelligibility (cf. Wheeler 1970; Vickers 1976), the synchronically-attested Siona and Sekoya varieties are generally conceived as a dialect continuum (Bruil 2014, 11-12). On the basis of case-marking facts, Koreguaje is remarkably similar to the Siona-Sekoya languages; although there are good grammatical and phonological reasons to exclude this language from the proposed continuum (Bruil 2014, 2018, 2019, *p.c.*).

An additional dimension is relevant to the establishment of Siona-Sekoya dialectology. A Peruvian variety of Sekoya is spoken in the vicinity of the Middle Putumayo, to the east of Ecuadorian Siona. This variety has begun to receive greater attention (e.g., Schwarz 2014; Vallejos and Schwarz 2016; Vallejos and Brown 2021), although a complete description of the case-marking facts is lacking at this time. The most crucial distinction between this variety and the others for the present purposes concerns the absence of the case-marker *-ni* (Rosa Vallejos, *p.c.*). This finding has implication for the diachronic development of the rich case-marking system found in the other Siona-Sekoya languages and in Koreguaje.

[7] Skilton (2013) bases her claims exclusively on phonological innovations.
[8] WT stands for 'Western Tukanoan'.

1.3.2 Siona in its areal context

Siona is spoken at the northwestern-most extremities of Amazonia, one of the most linguistically diverse areas in the world (Dixon and Aikhenvald 1999; Epps and Michael 2017; van Gijn and Muysken 2020). Traditionally, the top-level areal division in the study of South American languages opposes the highland (Andean) languages to those of the lowlands. In recent decades many specialists have sought criteria to identify smaller and more cohesive language areas in the lowlands (Muysken 2012; van Gijn 2014, 2017, 2020; van Gijn et al. 2023). Epps and Michael (2017) suggest the following micro-areas for lowland South America:[9]

Figure 1.3: Linguistic micro-areas across lowland South America (Epps and Michael 2017, Map 32.1)

Identifying areas is useful for delimiting domains for typological comparison. However, the Tukanoan family as a whole does not fall neatly within any one of the micro-areas proposed in the literature. On the one hand, Eastern Tukanoan languages are firmly established within the Vaupés culture-linguistic area (Aikhenvald 1999, 2001, 2003; Epps 2007; Stenzel 2008, 2013c; Zúñiga 2007; etc.). On the other hand, the status of

[9]Epps and Michael 2017 utilize the term *contact zones* for regions where lexical and grammatical traits are diffused (piggybacking on cultural diffusion). The suggested contact zones align neatly with the linguistic micro-areas presumed here.

the Western Tukanoan languages is less clear. They are not recognized in the putative Putumayo-Caquetá region in Epps and Michael (2017, §32.2.2), who enumerate the Witotoan family, Boran, Resígaro (Arawakan) and Andoke (isolate).[10] However, a plausible contact scenario between Máíhĩkì and Witotoan languages is proposed by Jolkesky (2016). Most Western Tukanoan languages are spoken further to the west of this area. Given such uncertainty regarding linguistic micro-areas, this dissertation uses a broad label, northwestern Amazonia (henceforth the NWA area), for the large geographical expanse between the Ecuadorian-Colombian-Peruvian foothills to the west and the Vaupés region to the east (following our convention in van Gijn et al. 2023).

Ecuadorian Siona in particular is spoken alongside several non-Tukanoan NWA languages within the Aguarico-Upper Napo fluvial zones − i.e., Shiwiar (Chicham), A'ingae (isolate), Wao Tededo (isolate, per Adelaar and Muysken 2004), and lowland Kichwa. Some preliminary evidence is suggestive of contact-induced lexical and grammatical patterns between Siona, Shiwiar and Kichwa (Bruil 2015a; Kohlberger 2020; Piispanen 2021; van Gijn et al. 2023), and perhaps A'ingae. Many questions remain as to patterns of contact-induced grammatical convergence, including DCM patterns, some of which are addressed in Chapter 4 of this dissertation. In any case, there is no question as to whether Siona and Sekoya communities have experienced cultural contact with Wao Tedede and A'ingae speakers for centuries (e.g., Vickers 1976; Wasserstrom 2014; de Carvalho 2022; Fischer and Hengeveld 2023).

1.4 The representation of Siona data

This section describes how Siona examples are reported in this dissertation. Section 1.4.1 considers how Siona examples are categorized on the basis of the data source and the technique employed for their collection. Then the glossing conventions followed in this dissertation are detailed in Section 1.4.2, followed by a discussion of orthographic conventions in Section 1.4.3.

1.4.1 A taxonomy of Siona example types

Siona examples used in this dissertation fall into two primary categories: NATURALISTIC data is extracted from free-flowing, semi-spontaneous Siona texts from Bruil (2012), and ELICITED data collected in the context of targeted interviews conducted with Siona speakers.

The NATURALISTIC examples are identified with an example-typing tag − i.e., NAT. This type tag precedes the metadata identifier formatted following the conventions of Bruil (2012, 2014, 2015b, 2018, 2019, *et seq.*). A subset of NAT examples reported in this dissertation are identified with an asterisk (i.e., NAT*), which indicates that the example is lifted from the narrative sample assembled for the quantitative analysis in Chapter 5, and Appendix B. (9) provides the full example identifiers for two text-based examples as an illustration:

(9) a. [NAT: 20100629slicr001.001]
b. [NAT*: 20151023orocr001.324]

[10]The putative Putumayo-Caquetá micro-area corresponds to the *People of the Centre* cultural zone (Echeverri 1997; Seifart 2005, 2007, 2011; Seifart et al. 2009; Wojtylak 2021).

The metadata tags shown in (9) are identical to those used to label the documentary raw data bundles in the ELAR database, i.e., where the original file bundles are archived.[11] Each identifier recognizes a number of details regarding the recording session as follows:[12]

- The date of data collection is recorded in YYYYMMDD format ;
- The session type identifier: (e) elicitation session, (s) staged session − including narrations, and (o) observed events ;
- The first two characters of first and last name of the (lead) consultant of the session − e.g., Ligia Criollo [licr] in (9a) ;
- Session number, for instance if there are multiple sessions with the same speaker on a given date ;
- Line number − based upon lines in transcription files provided in data bundles in Bruil 2012, as adapted upon normalization. E.g., (9b) is taken from the 324th line in the transcription of the narrative analyzed in Appendix B.

The same metadata conventions are utilized to label elicited examples. All of my elicited data is identified by the session type identifier (e); however, these data points are further compartmentalized into two sub-categories, inspired by Bochnak and Matthewson 2015. On the one hand, elicited data which is produced entirely by the speaker is labeled with the shorthand example-typing tag VOL (for 'volunteered' data). For instance, this tag is attributed to examples produced in the context of a translation task. Although the researcher provides a prompt in the communication language, i.e., Spanish, the speaker is responsible for producing the Siona sentence which forms the data point under consideration.

On the other hand, some examples considered in this dissertation involve a Siona sentence produced by the researcher, which is subsequently subjected to judgment or to manipulation by the speaker.[13] This latter type of sentential data is labeled as SUG (i.e., 'suggested' by the researcher). Such suggested-type examples are important for the presentation of negative evidence and data collected in highly-controlled contexts. Nonetheless volunteered and naturalistic examples are used wherever possible. Some instances, particularly in Chapter 6, concern volunteered sentences which are manipulated in the course of elicitation and the adapted sentence is suggested for judgment. Such instances are identified via the combined tag VOL/SUG in recognition of the split role of the consultant with respect to the production of the data in question. (10) provides example identifiers for elicited data points:

(10) a. [VOL: 20240918ejabi001.007b]
b. [SUG: 20190618emapa002.012]

[11] The full documentary dataset is found in the *Endangered Language Archive*: Bruil, Martine. 2012. "Documentation of Ecuadorian Siona". *Endangered Languages Archive*. Handle: http://hdl.handle.net/2196/00-0000-0000-000D-EA53-3.

[12] More details regarding the metadata labeling conventions is outlined in Bruil 2014, 14-17.

[13] The type of judgment, and degree of introspection involved, is a well-studied component of linguistic methodology (Chomsky 1961; Schütze 1996; Davidson 2020; and many others): e.g., grammaticality judgment, truth-conditional judgments, felicity judgment. Other types of introspective judgments provide insight into subtleties, and may fall within either the VOL or SUG category based upon the role of the consultant in gathering the concerned data point.

1.4.2 Glossing conventions: Reporting parsed Siona data

The glossing conventions followed in the present dissertation mostly replicate those found in Bruil (2014, 2018, 2019; and elsewhere). These conventions align with the best practices in the standard Leipzig Glossing Rules.[14] The tiers used in Siona glossed examples are as in (11):

(11) (A) ⇒ [**Context:** . . .]
(B) ⇒ *Transcription . . .*
(C) ⇒ Morpheme-by-morpheme parse: e.g., x-y **z**
(D) ⇒ Interlinear gloss word-affix target_word
(E) ⇒ Suggested English translation '. . . ' [(F) ⇒ Example identifier]

The Context-tier (A) is optional. In the case of certain naturalistic examples, this tier is utilized to contextualize the example in the surrounding text. In certain volunteered examples, this tier reports the suggested contexts in context-conjuring tasks (van der Wal 2016). Otherwise this tier reports speaker comments where they inform the discussion at hand.

The original transcribed sentence in (B) is presented in a legible format for the Siona public − i.e., the native orthography introduced in Section 1.4.3 below. No capitalization or punctuation conventions are used in this tier, following the format used in Bruil (2014, Appendix I). Regarding naturalistic data, this tier lifts the sentence(s) as they appear in the documentary materials. This transcription connects to its idiomatic English translation provided in tier (E), which is written following typical English punctuation conventions.

The interlinear gloss comprises morphologically-parsed Siona items, presented in *linguistic orthography* in tier (C), and the corresponding glossing labels are presented in tier (D). The linguistic orthography is also used to refer to Siona items in-text. The abbreviations used for glossing in tier (D) are presented in the list of abbreviations in the front matter of this dissertation.[15] No quotations or punctuation are found in tiers (C) or (D). (12) illustrates a fully-glossed example from the Appendix B. The material which is most relevant to the argumentation, typically nominal (sub-)constituents, is presented in **boldface** in tier (C) for ease of reference (e.g., the nominal *tsia-**re*** 'egg' in the example).

(12) [**Context:** The wife of Baina, the mythological folk hero of the Sionas, has a final meal before laying the eggs which will contain his children.]
. . . anichaojëna tsiare baoña io
ãni + chao-hi-na **tsia-re** baa-o-ña ĩ-o
eat + finish-PL.PRS.DEP-DS egg-N.SBJ have-2/3S.F.PST.N.ASS-REP 3PRO-CL:F
'. . . once (they) finished eating, she had eggs.'
[NAT*: 20150811sfryi001.112]

Turning now to the suggested English translation in tier (E), presented between single quotes '. . . ', the element corresponding to the **boldface** Siona item(s) is underlined.

[14] Guidelines accessible at https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf

[15] Following LGR conventions, < - > represents affixal concatenation, < = > represents cliticization, while head-head sequences (e.g., compounding) are indicated via < + > between individual elements.

Where relevant, quotations are presented, "...", and all English punctuation conventions are followed. Material which is either covert or tacitly understood is presented in parentheses to form an idiomatic English translation of the Siona original.

The role of DCM for encoding contrast is explored at various points in this dissertation. Where appropriate, focus is presented in the suggested English translation tier (E) as SMALL CAPS, which emulates the prosodic stress placement, encoding focus (or the emphatic contrast reading) in naturalistic English, based upon my own native speaker intuitions. The elicited example in (13) demonstrates how focal status is reported in this dissertation:[16]

(13) [**Context:** A boy comes home to find coffee and tea on the table.]
yë'ë ocore ụcusi'i cayë
jɨ'ɨ **ohko #(-re)** ũhku-si'-i kaa-jɨ
1SG water-N.SBJ drink-FUT-N3S say-N3S.PRS.ASS
'I want to drink WAter.' [VOL/SUG: 20230623ejabi001.024-025]

In addition to demonstrating the presentation of focus, (13) reports a linguistic judgment. These judgments are strictly appropriate for elicited data points, whereas naturalistic data is presumed to be grammatical and pragmatically felicitous in all cases.[17] This dissertation primarily deals with two types of linguistic violations: grammatical violations, indicated with an asterisk (*), and felicity violations, indicated with a hash symbol (#). The usage of these violation symbols aligns with the dominant conventions found in the literature.

1.4.3 Orthography matters: Presenting Siona data

There are two orthographical systems for recording Siona: (i) the native orthography, and (ii) the linguistic orthography. Siona speakers are able to read and write in the former system, but are generally unfamiliar with the latter. All preliminary documentary materials are transcribed using the native orthography. Following Bruil (2014, *et seq.*), this dissertation utilizes both orthographies, and opts for the linguistic orthography for in-text reference to Siona words or phrases.

The native orthography (originally *practical orthography*) was developed by American missionaries, Orville Johnson and his wife María Johnson, from the Summer Institute of Linguistics (SIL). This writing system was developed in the late 1950's among speakers of Siona's sister language, Sekoya. These activities were sanctioned by the regional governments across adjacent regions of Ecuador, Colombia and Peru in order to promote evangelization and assimilation among the indigenous communities of the area (cf. Vickers 1976, 1981). This writing system is found in bible translations, children's tales, work-related documentation, and other texts to the present day.

The Spanish influence of the native orthography is evident: e.g., /k/ is written as either <c> when preceding a vowel {/a/,/o/,/u/}, or their nasal counterparts, whereas it is written as <qu> before front vowels {/i/,/e/} (or before the high-mid vowel /ɨ/, written <ë>). In addition to this sixth basic vowel quality, the Johnsons

[16] Note that (13) is built around a (consultant) volunteered sentence, presented with follow-up judgment regarding the absence of *-re*; hence this example is typed as VOL/SUG accordingly.

[17] This dissertation is not concerned with false starts, self-corrections, or other dynamics of naturalistic data which may be integrated into the discussion in future research. Rather, naturalistic data are treated as a window into actual usage patterns, which are relevant to the description of certain aspects of Siona DCM in this work.

introduce the glottal stop, written as an apostrophe <'>, and the sub-litteral bar to represent the nasal quality of the vowel: e.g., <a̲> represents /ã/ (see Johnson and Levinsohn 1990, 23-24). Beyond these additions, the native orthography mirrors spelling and punctuation conventions of Spanish.

Conversely, the linguistic orthography is developed by Bruil (2014) to accomplish a more transparent phonetic representation of the language, favouring IPA conventions. This writing system is used in academic work reporting Siona data. Table 1.3 is replicated from (Bruil, 2014, 132), which maps correspondences between the native and linguistic orthography, and provides the underlying phonemic realization. I direct the reader to Bruil (2014, §3), and to Bruil and Stewart (2022), for a more elaborate discussion of various phonological and phonetic phenomena in Ecuadorian Siona:[18,19]

[18]One minor discrepancy regarding the usage of the linguistic orthography here, compared with Bruil 2014, concerns the representation of <j> and <ñ> allophony. Whereas Bruil collapses these realizations as underlying [j], this dissertation opts for a more transparent representation, without taking a theoretical stance on the issue. For instance, regarding the generic classifier, *-je* and (nasal) *-ñe* are distinct if unified in the glosses (i.e., CL:GEN).

[19]Bruil (2014 and others) suggest that the phonetic realization of the voicing contrast in the language has to do with creaky phonation. This is a crucial factor which distinguishes Ecuadorian Siona from Ecuadorian Sekoya phonology. Wiegertjes (2020) and Bruil and Stewart (2022) provides some initial phonetic analysis of this phenomenon.

Phoneme	Native orthography	Linguistic orthography
[p]	<p>	<p>
[p̰]	<b>	<b>
[t̰]	<t>	<t>
[t̲]	<d>/ <r>	<d>
[k]	<c>/ <qu>	<k>
[k̰]	<g>/ <gü>	<g>
[k^{w}]	<cu>	<kw>
[k̰^{w}]	<gu>/ <gü>	<gw>
[ʔ]	<’ >	<’ >
[s]	<s>	<s>
[s̲] / [ts]	<ts>	<z>
[ʧ]	<ch>	<ch>
[h]	<j>	<h>
*[h] (hC)	C	hC
[m]	<m>	<m>
[n]	<n>	<n>
[ɲ]	<ñ>	<ñ>
[w]	<hu>	<w>
[j]	<y>	<j>
[i]	<i>	<i>
[ĩ]	<i̲>	<ĩ>
[ɨ]	<ë>	<ɨ>
[ɨ̃]	<ë̲>	<ɨ̃>
[u]	<u>	<u>
[ũ]	<u̲>	<ũ>
[e]	<e>	<e>
[ẽ]	<e̲>	<ẽ>
[o]	<o>	<o>
[õ]	<o̲>	<õ>
[a]	<a>	<a>
[ã]	<a̲>	<ã>

Table 1.3: Overview of Siona orthographies - adpated from Bruil 2014, p. 132

*The linguistic orthography recognizes preaspiration arising before the plosive onset at certain syllable boundaries (Bruil 2014, 103-106; van ’t Veer et al. 2025), which are not represented in the native orthography.