Differential case-marking in Ecuadorian Siona

Case, J.T.

CHAPTER 5

A variationist study: DCM in a narrative sample

For the present chapter, I put on my *variationist* hat in order to ascertain the actual distribution of case-marking alternatives across a corpus of naturalistic Siona data — i.e., referred to as the Siona narrative dataset. Therefore, this chapter contributes quantitative data, which complement the qualitative description established elsewhere in this dissertation. More concretely, the actual PROMINENT-marking rates for various Siona DCM RELATIONS is ascertained, while the application of statistical techniques probes the activity of plausible TRIGGERS for Siona DCM effects. The statistical models themselves are reported in full, alongside lifted illustrative examples, in Appendix B, which accompanies the discussion developed in this chapter. Ultimately, this quantitative chapter is a welcome contribution to analyzing the REALIZATIONAL DOMAIN, as a component of the holistic characterization of DCM patterns in the language, as pertains to the set of active TRIGGERS, which drive case-marking patterns in actual usage.

This chapter is organized as follows: Section 5.1 outlines the Siona narrative sample developed for this chapter. Section 5.2 outlines the methodological approach adopted to assemble and code the narrative sample. This section also introduces the variationist techniques utilized to establish the statical models developed to analyze the sample. These models are spelled out in full in Appendix B, which accompanies the present chapter. Section 5.3 discusses the quantitative results on a grammatical relation-by-grammatical relation basis, considering DCM patterns for argumental and for spatial DCM patterns in turn across the narrative sample. Section 5.4 closes the present chapter by integrating the quantitative data into the broader characterization of Siona DCM purported in this dissertation.

5.1 The Siona narrative dataset

The Siona narrative dataset analyzed in this chapter, and Appendix B, comprises five narratives from the ELDP documentary corpus (Bruil 2012),[1] as enumerated in Table 5.1. On the one hand, the length of each narrative is measured in "Time", corresponding to the duration of the raw video recording in the archival materials; whereas "Word count" is determined based upon the lifted transcriptions, which form the basis for the token-based analysis reported here. Each narrative is attributed an English title for convenience, although the analysis purported here is performed upon the pooled dataset:

STORY TITLE [Recording code in Bruil 2012]	TIME (HH:MM:SS)	WORD COUNT	NOMINAL TOKENS *N* (% of total words)
Siona Genesis story [20151023orocr001]	21:33	1714	376 (21.94%)
The demon kwẽẽwahti [20151112oespa001]	15:48	1291	327 (25.33%)
The mother of jaguars [20151001oolpi001]	14:12	981	251 (25.59%)
The mother of the moon [20151112orapi001]	16:16	1350	331 (24.52%)
The children of Baina [20150811sfryi001/2]	33:25	2742	629 (22.94%)
POOLED TOTAL	**1:41:04**	**8078**	**1914 (23.69%)**

Table 5.1: Composition of the Siona narrative dataset

The corpus outlined in Table B.1 is designed in order to provide a balanced and representative sample of the traditional oral folktale genre (cf. Biber et al. 1998, Jones and Waller 2015). In addition to the semi-spontaneous nature of these narratives, the included folktales are appropriate for token-based, statistical inquiry because they implicate a set of characters, which are tracked across longer stretches of discourse. As such, this genre is an ideal testing ground for certain plausible DCM TRIGGER candidates, which have been suggested elsewhere — e.g., specificity and topicality.

To ensure the development of a balanced narrative sample, several factors were taken into consideration. Firstly, each story is told by a different native-speaker narrator: including two middle-aged males, one elderly male, one middle-aged female, and one elderly female.[2] Secondly, the narrative sample is fairly balanced as con-

[1] Project ID IPF0211SG0067 [https://www.elararchive.org/dk0184]. All materials were collected by Martine Bruil between 2010-2015 under affiliation with Leiden University and University of California, Berkeley. Stories are archived as multimedia bundle folders, which may be accessed via the ELAN (EUDICO linguistic annotator) tool: ELAN (Version 6.9) [Computer software]. (2024). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan.

[2] Given the degree of intermarriage with neighbouring communities, it is not always the case that all speakers are equally confident in their Siona language competency. This fact was taken seriously as Prof. Bruil and I determined which narratives to include in this corpus

cerns story length, as shown in Table B.1, with a mean word count of 1615 word forms. This ensures that no narrator is represented disproportionately in the pooled dataset. Additionally, the sample size utilized in this analysis is in keeping with other manually-coded quantitative (DCM) studies in the literature (e.g., Shain and Tonhauser 2010; Tonhauser and Colijn 2010; Bruil 2014; Stenzel 2015; Grzech 2016). Crucially, this sample size is also suitable for the application of statistical techniques (i.e., the *variable*-rule model), whereas smaller datasets do not yield informative results. More methodological details are provided in the following section.

5.2 Methodology: The variationist program

The present chapter adopts a *variationist* methodology for collecting and testing quantitative data, based upon the Siona narrative dataset outlined in Table B.1 above. The family of variationist techniques is unified by the assumption that natural language contains *linguistic variables*, which are components of language which exhibit structured variation in their distribution (cf. Tagliamonte 2006, 2012). The conception of DCM adopted in this dissertation is compatible with the conventionalized notion of *linguistic variable* — i.e., "the phenomenon whereby not every argument bearing a given grammatical relation is encoded with the same case-marker" (lifted from (2) in Chapter 2). For instance, consider the two variant forms of the spatial GOAL-argument, *airo* (jungle), shown in (1) — i.e., the PLAIN zero-marked variant arises in (1a), lifted from *The mother of jaguars* narrative; whereas the PROMINENT *-na* marked variant arises in (1b) from *The children of Baina* narrative:

(1) a. *. . . ba'iquëbi dos dia jesǫ mo'seña baja'i ba'iquëbi airo saiyë caëña*
ba'i-ki-bi dos dia hesõ mo'se-ña ba-ha'i
be-CL:M-SBJ two day how_many day-PL.INAN be-3S.M.PST.ASS
ba'i-ki-bi **ai-ro** sai-jɨ kaa-i-ña
be-CL:M-SBJ big-CL:LOC go-N3S.PRS.ASS say-2/3S.M.PST.N.ASS-REP
'. . . he was (there) two days, or however many days it was, he wanted to go to the forest (to go hunting).' [NAT*: 20151112oespa001.076-77]

b. *. . . airona sani tǫmejë huahuaroquë'rona sanitǫmereña*
ai-ro-na sani + tõme-hɨ wawaro-kɨ'ro-na
big-CL:LOC-GOAL go + fall-PL.PRS.DEP cockroach-DERIV.LOC-GOAL
sani + tõme-re-ña
go + fall-N2/3S.PST.N.ASS-REP
'. . . (they) landed in the forest, they landed where the COCKroaches live.' [NAT*: 20150811sfryi001.352-356]

The proportion of PROMINENT-marking may be established in a systematic way. Firstly, and more straightforwardly, it determines the rate at which PROMINENT-marking arises in the narrative sample — i.e., simply referred to as proportional evidence. Secondly, it is not immediately clear why PROMINENT *-na* marking arises in (1b), but not in (1a), for instance. This chapter develops statistical models in order to ascertain which of a series of predictor variables drive the selection of PROMINENT-markers for a given grammatical relation, each of which is conceived as a plausible DCM TRIGGER candidate — i.e., contributing statistical evidence.

sample.

This analysis is performed on each grammatical relation displaying principled DCM alternations, as described in the *preliminary description*, and reiterated in Table 5.2:

	INANIMATE		ANIMATE (∧ ↑INAN)	
	PLAIN	PROMINENT	PLAIN	PROMINENT
Argumental DCM:				
SUBJECT (S)	-∅	*-bi*	-∅	*-bi*
DIRECT OBJECT (P)	-∅	*-re*	*-re*	*-ni*
INDIRECT OBJECT (R)	*-re* (*-na*)		*-re*	*-ni*
Obligatory case-marking:	*-bi* (Instrument); *-hã're* (Companion)			
Spatial DCM:	INANIMATE		↑INANIMATE	
LOCATION (L)	-∅	*-re*	*-re*	*-ni*
[*some L] GOAL (G)	-∅	*-na*	*-na*	
Obligatory case-marking:	*-bi* (Source); *-hã'ã* (Path, Limit)			

Table 5.2: Siona DCM paradigm, PLAIN-PROMINENT mappings for corpus annotation

For the remainder of this chapter, each DCM RELATION is identified by the shorthand identifier in Table 5.2 — e.g., S-tokens refer to those categorized as SUBJECT nominal tokens. The analysis addresses two research questions for each tested grammatical relation: (i) how often PROMINENT alternatives are selected in the sample, and (ii) what TRIGGERS factor into the selection of PROMINENT, over PLAIN, forms in a given context. Both proportional and statistical evidence are welcome quantitative contributions to the characterization of Siona DCM.

The remainder of this section outlines the how variationist techniques were applied to ascertain this quantitative evidence against the Siona narrative dataset. Section 5.2.1 outlines how the pool of nominal tokens is assembled, and introduces the DCM TRIGGER candidates, which were converted into factors (i.e., predictor variables) for the sake of statistical analysis. Section 5.2.2 introduces the *variable rule* model, which is the statistical technique applied in the study reported in this chapter.

5.2.1 From tales to tokens: Coding the narrative sample

The variationist methodology adopted in this chapter is applied against the narrative sample, which comprises 1914 nominal tokens in its raw form. These tokens within the pooled set are subsequently sorted into their respective grammatical relation categories, which are enumerated in Table 5.2. Once sorted on the basis of grammatical relation, the case-marking alternative (incl. zero-marking) found on each token is recorded. In post hoc fashion, for each individual token, the 'PROMINENT-marking status' category is determined, which is the dependent variable for this study.

Compare the instances of P-tokens reported in (2). Each of the highlighted tokens in these instances bears the case-marker *-re*. However, per the PLAIN-PROMINENT mapping summarized in Table 5.2, the inanimate P in (2a) belongs to the PROMINENT-

marked category; whereas its animate counterpart in (2b) is in the PLAIN-marked category accordingly:

(2) a. [**Context:** The children of Baina have killed the mother of the household and put her in the stew, but the father is not yet aware.]
. . . ñajëna yë'ë biada'care cusi'i caëña
ñaa-hi-na ji'i **bia** **+ da'ka-re** ũhku-si'-i
see-PL.PRS.DEP-DS 1S aji_pepper + liquid-N.SBJ drink-FUT-N3S
kaa-i-ña
say-2/3S.M.PST.N.ASS-REP
'. . . (they) watched (him) as (he) said "I want (some) soup." '
[NAT*: 20150811sfryi001.264]

b. *jaje yo'co aco'ë tsidohuëre yureta'a quëaja'cua'ia'ë ja'quëre careña*
hãhẽ jo'-ko ãh-ko'i **tsĩ-dowi-re**
like_that do-S.F.PRS.DEP eat-3S.F.PST.ASS child.COL-PL.AN-N.SBJ
jure-tã'ã kia-hã'-ko-wa'i-a'-i ha'-ki-re
now-CNT.EXP tell-PRP-CL:F-PL.AN-COP-N3S parent-CL:M-N.SBJ
kaa-re-ña
say-N2/3S.PST.N.ASS-REP
' "(She) did this and ate the children, now we will tell father", (they) said.'
[NAT*: 20151001oolpi001.067]

The examples in (1) and (2) suffice to demonstrate how tokens are identified, extracted, and sorted. Previous descriptive work, assembled in the *preliminary description* in Chapter 3, permits the identification of grammatical relations for a given token and the determination of its PROMINENT-marking status. Once these initial stages are undertaken, the proportional evidence may be calculated in a straightforward way.

The next stages lay the groundwork for the implementation of statistical techniques — i.e., the *variable-rule* model, introduced in Section 5.2.2. The initial step concerns the systematic coding of nominal tokens on the basis of factors, or, conceptually, trigger candidates. In the variationist literature, this is typically referred to as *circumscribing* the independent variables (Sankoff 1988; Tagliamonte 2006, 2012; etc.). An elaborate overview of the coding protocols is provided in Section B.1 of Appendix B; whereas, this section is concerned with motivating the independent variables selected for the statistical analysis reported in this chapter.

The independent variables tested in this chapter correspond to well-established, cross-linguistic TRIGGERS for DCM patterns, discussed in Chapter 2. Conceptually, each of these variables is a DCM TRIGGER CANDIDATE, viewed as a plausible motivation for selecting PROMINENT-marking in a given context. Table 5.3 enumerates and provides a working definition for each TRIGGER CANDIDATE tested in this chapter, whereas details regarding their operationalization for coding purposes are given in the appropriate places in Appendix B:[3]

[3]These variables are assembled into groups by convention, these add conceptual coherence to the discussion found in this chapter.

Trigger candidate (predictor variable)	Working definition
	Sentence-level Triggers
Word order	The nominal token appears before or after the main verbal word in its clause.
Thematic status	The referent of the nominal token is attributed a particular thematic role based upon the predicate: (i) *agentive* or *non-agentive* (s-argument); (ii) *affected* or *unaffected* (p-argument); or (iii) *subcategorized* or *non-subcategorized* (l-argument)
	Referent-level Triggers
Animacy status	The nominal token denotes an animate or inanimate entity.
Specificity status	Pronominal tokens are separated from full nominal tokens. Full nominal tokens refer to specific or non-specific entities.
	Discourse-level Triggers
Contrastivity status	The referent of the nominal token contrasts with another entity (explicit or implicitly) in the narrative.
Topicality status	(1) The referent of the nominal token has a recent, overt mention in the discourse (Look-back topic) (2) The referent of the nominal token persists as a topic in immediately ensuing discourse (Look-ahead topic)

Table 5.3: Working definitions for annotated factors in Siona narrative sample

The first group of trigger candidates tested in this chapter concern sentence-level factors — precise annotation guidelines are provided in Section B.1.1. In certain languages, case-marking is obligatory where the noun is in one sentential position, but variable in another. For instance, certain Eastern Tukanoan languages display such patterns for Direct object-marking, where dcm strictly arises in the immediate pre-verbal position in the sentence (Stenzel 2013c, 333-5). Although these particular constraints on dcm are not observed in Siona or other Western Tukanoan languages, this study tests for any statistical correspondence between the surface position of the nominal token (i.e., either as pre- or post-verbal) and case-marker selection. Put another way, this study measures the impact of deviation from the canonical sov word order on dcm, such that both canonical s- and p-tokens arise pre-verbally.

A second family of trigger candidates concern the impact of thematic properties of an argument on its case-marking. As discussed in Section 2.2.3, in various languages, the 'agentivity' status of the Subject impacts case-marking (Malchukov 2008; Fauconnier 2011); and similar patterns are observed regarding the 'affectedness' status of the Direct object (Næss 2004; von Heusinger and Kaiser 2011; Kizilkaya et al. 2022). Both of these factors are tested systematically in this chapter. This study also pilots the testing of whether the sub-categorization status of spatial arguments affects dcm — i.e., distinguishing implied Locations in a predicate like 'we live in the house', from adjunctive Locations, e.g., 'we are playing in the park'. This chapter adapts the technique used to test for affectedness effects in Spanish *differ-*

ential object marking found in von Heusinger and Kaiser (2011) — the technique is spelled out in detail in Section B.1.1.

The third group of tested trigger candidates are referent-level factors, which form the best-studied set of DCM TRIGGERS in the literature. The *preliminary description* provided in Chapter 3 established unequivocally that Siona (non-SUBJECT) DCM is shaped by animacy-based noun classes, as reflected in Table 5.2 above. Whereas this fact clearly factors into the mapping of *formal* alternatives, the statistical analysis developed in this chapter determines whether notional animacy class is also a relevant factor regarding the proportion of PROMINENT-marking. This aspect of Siona DCM has not been addressed in previous work, although such WEAK TRIGGERS are described for various (non-)Tukanoan languages in the NWA area, as shown in Chapter 4.

An important contribution of the quantitative analysis developed in this chapter is to test whether specificity factors into Siona DCM, which is at best an unsettled issue up to this stage. nominal tokens are sorted into pronominal, specific, and non-specific categories (consult Section B.1.2 for annotation details), and it is determined whether these classes display different PROMINENT-marking rates. This is a dominant TRIGGER according to many descriptions for Tukanoan languages; and, although this dissertation has demonstrated, at least qualitatively, that specificity is not a STRONG TRIGGER for Siona DCM, it remains to be seen whether it has any bearing whatsoever — i.e., as a WEAK TRIGGER for (P-oriented) DCM in the language.

The final group of trigger candidates tested in this chapter pertains to discourse-related TRIGGERS. Firstly, CHAPTER 3 establishes that PROMINENT-marking is related to certain focus effects in Siona DCM, such as QUESTION-ANSWER constructions and what is labeled the *emphatic contrast* reading, which are expounded considerably in Chapter 6. Since there were not a significant number of questions or answers in the narrative sample, these are not directly coded for in the sample. Instead this chapter contributes to this broader discussion by testing whether the contrastivity status of the referent for a nominal token is relevant to the selection of PROMINENT-marking, as inspired by similar text-based techniques (Chafe 1987; Myhill and Xing 1996; Vallejos Yopán 2009; Vallduví 2016, §23.5). More details regarding the precise coding protocols are outlined in Section B.1.3.

This chapter tests for a topicality-related TRIGGER for Siona DCM via the implementation of two corpus-driven approaches adapted from elsewhere in the literature. The first topicality metric used in this study measures 'look back' topicality, as conceived in the seminal works of Givón (1976, 1983, *et seq.*): i.e., a nominal token is more topical if it is mentioned more recently in discourse. A related vein of literature concerns referent activation (e.g., Chafe 1976, 1994; Ariel 2014 — recently summarized in Arnold 2010); on this view this chapter analyzes the recency of (re)activation as a factor for DCM. This chapter opts for the term LOOK-BACK TOPICALITY. Following recent analyses for topicality effects in *differential object marking* (i.e., Chiriacescu and von Heusinger 2010, Shain and Tonhauser 2010), the mean distance between mentions is established for each grammatical relation in the sample, corresponding to the number of intervening main clauses between mentions of a particular referent. A nominal token is topical if it is found below this distance threshold, whereas a nominal token is non-topical if the number of intervening clauses exceeds this threshold. This technique is applied to detect topicality-triggered *differential object marking* in Paraguayan Guaraní (Shain and Tonhauser 2010).

The second topicality metric concerns LOOK-AHEAD TOPICALITY, following the terminological conventions in Givón's work. Put simply, the prediction is that PROMI-

NENT-marking is favoured in instances where the referent of the nominal token persists as an at-issue aboutness topic in immediately ensuing discourse (i.e., the following main clauses in the narrative). This notion includes many instances of topic shift in the discourse; however, some tokens are mentions of referents which are already activated at that moment in discourse, rendering the term *shift* less appropriate. This is compatible with the notion of (re)activation of a given referent as a discourse topic in the activation literature mentioned above. To measure LOOK-AHEAD TOPICALITY, chapter implements a similar technique to that argued to demonstrate significant topicality effects for Romanian *differential object marking* by Chiriacescu and von Heusinger (2010).

In summary, this chapter makes an important contribution to the holistic analysis of Siona DCM undertaken in this dissertation. Conceived as a language which exhibits several multi-dimensional DCM systems, it is expected that several TRIGGERS are active and account for the complex distribution of PLAIN and PROMINENT marking alternatives to varying degrees. The variationist techniques employed in this chapter are appropriate for testing which of these plausible trigger candidates are relevant to Siona DCM, in particular, by refining the characterization of the REALIZATIONAL DOMAIN based upon a naturalistic dataset.

5.2.2 The variable-rule technique: An overview

This chapter employs the *variable rule analysis* technique (henceforth *varbrul*) in order to develop a statistical model to test the trigger candidates identified in Table 5.3, across the relevant grammatical relations in Siona grammar. The *varbrul*-analysis is generally accepted as the gold standard for variationist linguistics, following its inception in the pioneering works of Labov (1966, 1970, *et seq.*; see also Sankoff and Labov 1979; Sankoff 1988). On statistical grounds, this technique produces a linear regression model, based upon a maximum likelihood algorithm. Today several digital tools are available to create *varbrul* models, and the models developed for the Siona narrative sample were generated via the open-source *Rbrul* tool.[4]

This section briefly unpacks the inner-working of the *varbrul* model, how to interpret its results, and how this technique is implemented to ascertain statistical evidence for the present chapter. In fact, the present chapter develops three independent *varbrul* models, analyzing the proportion of PROMINENT-marking across S-oriented, P-oriented, and L-oriented DCM patterns respectively. Whereas this dissertation recognizes principled DCM patterns for the R- and the G-argument, the modest Siona narrative sample consulted here yields insufficient tokens for the development of an informative *varbrul* model for these grammatical relations.[5] Only initial proportional evidence is reported for these relations accordingly. The three full *varbul* models are reported in Appendix B, whereas condensed reports are provided in this chapter in order to focus on the informative results.

[4]Accessible via online manual at http://www.danielezrajohnson.com/rbrul.html. This tool operates within the statistical engine *R* (R Core Team 2021). Alteranative open-source tools exist: e.g., *GoldVarb*, [up-to-date version via http://individual.utoronto.ca/tagliamonte/goldvarb.html]; and *R-varb* (see Palillo 2002).

[5]In statistical parlance the *goodness of fit* metric determines how informative the significance and factor weight measures are for a given linear regression model (i.e., *varbrul*-analysis). Both the number of included tokens and the degree of variation determine the goodness of fit for resulting statistical model. Further details regarding these statistical matters are unpacked in Section B.1 of Appendix B.

On conceptual grounds, the *varbrul* technique is applied to a coded token-based dataset, where the concerned linguistic variables display structured variation — i.e., where the PLAIN marking alternative varies with the PROMINENT-marker. This variation is structured, since these forms are not in free variation. Each token is attributed a series of property values, based upon a series of pre-selected predictor variables, called factors, which correspond to the trigger candidates in Table 5.3. So long as factors are not implicationally related to one another, several factor groups may be tested simultaneously on a given dataset — e.g., if definite tokens are *always* specific, it is inappropriate to test for two independent factors. Each factor represents a hypothesis of the type "S-oriented DCM is driven by *animacy status*", where the *varbrul* technique executes hypothesis testing on a factor-by-factor basis, based upon the dataset at hand.

On statistical grounds, the output yielded by the *varbrul* model reveals three pieces of statistical evidence for each tested factor (Tagliamonte 2006, 2012). Firstly, it addresses the hypothesis directly as to whether a given factor is statistically significant in the dataset, generating a p-value accordingly. A significant result is yielded where the observed proportion of alternatives is determined not to have arisen by chance. By convention, a p-value of < 0.001 is considered to be highly significant, and a p-value between $0.001 - 0.05$ is significant. If the p-value is > 0.05, then it is insignificant. For the sake of presentation; the condensed *varbrul* reports in this chapter collapse all significant factors, and insignificant effects are acknowledged and set aside.

Secondly, where a factor is significant, a FACTOR WEIGHT value is calculated for each factor-internal category. For instance, concerning the word-order factor, a weight value between 0 and 1 is determined for the pre-verbal category and for the post-verbal category on the basis of the proportion of PROMINENT-marking, category frequency and variability metrics. FACTOR WEIGHT is calculated by establishing the difference between the highest and lowest values within the factor-internal categories, which is often referred to as the *Range* value (Tagliamonte 2012). The *Range* value reflects the effect size for a given factor, where consequential differences in proportion are recorded across the internal categories. For the sake of the discussion in this chapter, numerical *Range* values are interpreted as effect sizes as shown in Table 5.4 — full *Range* values are reported in the complete models in Appendix B:

EFFECT SIZE	FACTOR WEIGHT RANGE	*e.g., calculation from varbrul*
LARGE EFFECT SIZE	*Range*: > 50	CAT_1 (0.71) − CAT_2 (0.15) = 56
MODERATE EFFECT SIZE	*Range*: $20 - 50$	CAT_1 (0.56) − CAT_3 (0.23) = 33
SMALL EFFECT SIZE	*Range*: < 20	CAT_1 (0.48) − CAT_3 (0.39) = 9

Table 5.4: Conversion chart: From *varbrul Range* values to effect sizes

The third contribution of the *varbrul* model is related to the second. On the view that FACTOR WEIGHT RANGE, as shown in Table 5.4, corresponds to the notion of TRIGGER STRENGTH in multi-dimensional DCM (Klein and de Swart 2011), a ranking of significant factors may be established for each tested grammatical relation. In addition to converting *Range* values to EFFECT SIZE categories for the discussion in this chapter, the *varbrul* results also permit an internal ranking of, for example, which factors with an MODERATE EFFECT SIZE have a larger impact than the others. Significant factors are represented in their order of EFFECT SIZE in this chapter to

reflect their relative import on the distribution of PROMINENT-marking in the Siona narrative dataset.

In summary, this chapter provides both proportional evidence and different types of statistical evidence. The latter family of evidence ultimately demonstrates which tested trigger candidates are significant, and how big of an impact the significant CANDIDATES have on the observed DCM patterns. For the remainder of this chapter, the descriptive term, TRIGGER CANDIDATE is favoured over the technical term factor.

5.3 Results: DCM in the narrative dataset

This section reports the quantitative evidence yielded by the variationist analysis of the Siona narrative dataset. Firstly, the discussion opens with a consideration of the global distribution of case-marking across all nominal tokens in the corpus in Section 5.3.1 − i.e., considering all 1914 nominal tokens identified in the pooled sample. This initial analysis concerns both grammatical relations with DCM patterns and those without. The ensuing sections report and interpret the statistical evidence based upon the *varbrul* results for each tested DCM RELATION in the dataset in particular: In the first place, Section 5.3.2 considers the results for S-oriented DCM patterns. Next, Section 5.3.3 considers the P-oriented DCM patterns. Finally, the results for L-oriented DCM are reported in Section 5.3.4.

5.3.1 Distribution of marking alternatives in the sample

Table 5.5 categorizes all 1914 nouns tokens in the Siona narrative dataset based upon the intersection of grammatical relation and surface case-marking. The term 'marking alternatives' is used to include all overt case-marking *and* zero-marking for the purposes of the present discussion. This proportion of marking alternatives arising in the full annotated sample is referred to as the global case-marking distribution in this chapter. It is important to note that the raw case-marking figures reported here do not reflect proportional facts per se, given that case-markers are not attributed PLAIN or PROMINENT status in the global marking dataset. Nonetheless, the mapping of included tokens on the basis of PROMINENCE STATUS is crucial to the statistical evidence considered in the ensuing sections.

The patterns of case-marking alternatives laid out in Table 5.5 do not present any real challenges to the generalizations in the *preliminary description* in Chapter 3. Nearly all instances of obligatory case-marking, and principled and accidental DCM patterns, arise with an expected marking alternative according to their grammatical relation (and animacy class, where appropriate). Therefore, the primary contribution of this chapter pertains to the *variation* between the available marking alternatives in the narrative sample. A few initial observations are in order before moving to a discussion of the proportional and statistical results.

	-∅	*-bi*	*-re*	*-ni*	*-na*	*-hã'ã*	*-hã're*	**Total (%marked)**
Subject (s)	618 (75.27%)	203 (24.73%)	0	0	0	0	0	**821 (24.73%)**
Direct object (p)	209 (43.63%)	0	224 (46.76%)	46 (9.6%)	0	0	0	**479 (56.37%)**
Indirect object (r)	0	0	53 (75.71%)	17 (24.29%)	0	0	0	**70 (100%)**
Point-of-Contact	2 (7.14%)	0	3 (10.71%)	1 (3.57%)	21 (75%)	1 (3.57%)	0	**28 (92.86%)**
Place-of-Putting	1 (10%)	0	2 (20%)	0	6 (60%)	1 (10%)	0	**10 (90%)**
Instrument	0	**29**	0	0	0	0	0	**29 (100%)**
Companion	0	0	0	0	0	0	**2**	**2 (100%)**
Static location (l)	65 (40.12%)	0	53 (32.72%)	22 (13.58%)	22 (13.58%)	1 (0.62%)	0	**162 (59.88%)**
Goal (g)	44 (34.38%)	0	2 (1.56%)	0	74 (57.81%)	8 (6.25%)	0	**128 (65.63%)**
Source	0	**25**	0	0	0	0	0	**25 (100%)**
Path/Limit	0	0	0	0	0	**21**	0	**21 (100%)**
Time	130 (91.55%)	7 (4.93%)	0	0	3 (2.11%)	1 (0.7%)	0	**142 (7.75%)**
(*N*)	*-∅* (1070)	*-bi* (264)	*-re* (333)	*-ni* (86)	*-na* (126)	*-hã'ã* (33)	*-hã're* (2)	**Total (1914 tokens)**

Table 5.5: Raw distribution of nominal tokens in Siona narrative dataset, by case-marking and grammatical relation

Firstly, concerning argumental case-marking, R-tokens are never zero-marked in the sample, but rather receive *-re* or *-ni* marking, as anticipated based upon the description in Section 3.2.3. There are no instances of inanimate R-tokens in this dataset, such that *-re* and *-ni* correspond to PLAIN- and PROMINENT-alternatives accordingly. On this basis, a PROMINENT-marking rate of 24.29% is observed for R-oriented DCM, which aligns with other argumental DCM patterns in the sample.[6] As regards Point-of-Contact and Place-of-Putting arguments, described in Section 3.2.4 — setting aside a few outliers in the sample, these tokens are marked with *-na*.[7] All attested INSTRUMENT and COMPANION tokens bear their respective obligatory case-markers.

Turning to the spatiotemporal case-marking facts outlined in the global raw data in Table 5.5, a few observations are in order. Firstly, setting aside the full statistical analysis of L-marking for Section 5.3.4, an interesting pattern may already emerges based on the distributional figures alone. As pertains to G-oriented DCM, ignoring a few outlier tokens marked with the PATH-marker, *-hã'ã*, PROMINENT *-na* is selected at a rate of 62.71%.[8] This is markedly higher than the PROMINENT-marking rate observed for R-marking above. In fact, as will be established in this chapter, this discrepancy represents a broader, novel generalization: Argumental DCM displays PROMINENT-marking at a far lower rate than spatial DCM.

Another set of observations emerge from Table 5.5 as concern case-marking and temporal tokens. This nominal category does not arise in the *preliminary description* in Chapter 3, since Siona generally does not exhibit temporal case-marking. This generalization figured in the typological discussion in Chapter 4. The distributional facts considered in this chapter confirm this generalization. (3) displays two typical instances of zero-marked TIME tokens from the narrative sample:

(3) a. *i̲o̲ neato ñami ti̲'a̲ni a̲i̲sicobi a̲nichaoni i̲recato a̲coña*
ĩ-o **nea-to** **ñami** tĩ'ã-ni ãĩ-sih-ko-bi ãni
PRO3-CL:F be_dark-CL:LOC night arrive-SS eat-COMP-CL:F-SBJ eat
+ chao-ni ĩ-i̲-re=kato ãh-ko-ña
+ finish-SS 3PRO-CL:M-N.SBJ=TOP eat-2/3S.F.PST.N.ASS-REP
'She arrived later in the night, and she (lit. the one who had eaten), finished up (the other one) and then ate HIM.'
[NAT*: 20151001oolpi001.092-093]

b. *bacoña i̲o̲ de̲baocato. ja̲macarebacato ë̲jë̲re baoña bacona . . .*
bah-ko-ña ĩ-o dẽbao=kato
be-2/3S.F.PST.N.ASS-REP 3PRO-CL:F vulture(spec)=TOP
hãmahka-reba=kato ĩhĩ-re baa-o-ña
then-INTENS=TOP husband-N.SBJ have-2/3S.F.PST.N.ASS-REP

[6] As indicated above, the sub-sample of R-tokens in the Siona narrative dataset does not meet the threshold for quantity (of internal categories) to merit the application of the *varbrul* technique, following the same coding protocols as the tested DCM RELATIONS analyzed in this chapter. Several examples of coded R-tokens are provided in Section B.2.4 of Appendix B as a matter of reference.

[7] Some typical and outlier Point-of-Contact tokens are illustrated in Section B.2.4.

[8] Note that these baseline figures do not take into account the fact that a small handful of promoted G-tokens (i.e., $N = 8$) are invariable, and ought to be excluded from any *varbrul* model. On this basis, the sub-sample of G-tokens is even smaller than anticipated. Some illustrative G-tokens are lifted in Section B.2.4 in Appendix B.

bah-ko-na
live-S.F.PST.DEP-DS
'There was the Dẽbao. At that time (she) lived on and had a husband...'
[NAT*: 20151112orapi001.005-007]

A handful of exceptionally case-marked TIME tokens are identified in the sample, which reflect certain sub-types of temporal nouns which have not been discussed elsewhere. These concern the few instances of temporal *-hã'ã* marking and *-bi* marking noted in Table 5.5.[9] These appear to reflect two fringe categories of nominals, which likely correspond to temporal extensions of spatial concepts — i.e., what might be labeled, the temporal Start-Point and End-Point relations, aligning with the spatial SOURCE and PATH relations accordingly. As such, these would correspond to instances of obligatory, semantic case-marking, although further work is required to properly analyze the encoding and interpretation of these temporal nominals. An illustrative instance of each is lifted from *The mother of jaguars* narrative in (4):

(4) a. *... yureta'a jamacabi te'e cocabi cañu'u caquë yequë ëjacoca, oracoca, yecua'i goasi'ahua'i*
jure-tã'ã **hãmahka-bi** te'e kohka-bi kaa-ñu'ũ
now-CNT.EXP then-SRC one language-INST speak-HORT
kaa-kɨ jehk-ɨ ɨha + kohka ora + kohka
say-S.M.PRS.DEP other-CL:M foreign + language kichwa + language
jehk-wa'i goasi'a-wa'i
other-PL.AN any- PL.AN
'... and now, from then on, he wanted them to speak but one language; but one (was speaking) Spanish (lit. foreigner language), and others yet the Kichwa language.' [NAT*: 20151001oolpi001.353-354]

b. *... yureta'a goade'ohuesëyë. jamacaja'a baja'i Baina caquë ba'isico*
jure-tã'ã goade'o + wehsɨ-jɨ **hãmahka-hã'ã**
now-CNT.EXP worsen + not_know-N3S.PRS.ASS then-LIM
ba-ha'i Baina kaa-kɨ ba'i-sih-ko
be-3S.M.PST.ASS Baina say-S.M.PRS.DEP be-COMP-CL:F
'... and now (we) don't know (our own language). Up to here, this was what Baina is said to have done.' [NAT*: 20151001oolpi001.396-398]

Besides these few observations, case-marking in the narrative sample aligns with the generalizations laid out in Chapter 3 in this dissertation. The remainder of this discussion undertakes a targeted discussion of the DCM patterns analyzed for S-, P-, and L-arguments in the sample, focusing on both proportional and statistical evidence yielded in this analysis. Each of the ensuing sub-section outlines the grammatical relation-particular sub-sample utilized to develop the statistical models, and then navigates the results of each model.

[9] Regarding the few instances of TIME tokens marked with *-na*, there is a handful of temporal expression which occur with this marker, although its semantic contribution is more opaque. For example, items like *jure-**na*** (now) or *neato-**na*** (tomorrow at dawn) have a more emphatic reading, although this is less likely in the common Siona greeting, *ñami-**na** ñañu'u* (let's see (each other) tomorrow). An instance from the sample is shown in (52) in Appendix B.

5.3.2 S-marking in the Siona narrative dataset

This section is concerned with analyzing patterns of S-oriented DCM which arise in the Siona narrative dataset. As demonstrated in Table 5.5, this is the best-represented grammatical relation in the dataset — i.e., accounting for 821 of the total 1914 nominal tokens identified in the pooled sample. Roughly 35% of main clauses in the dataset arise with an overt S-argument, which is comparable to the rate noted in work with texts in the Eastern Tukanoan language, Wanano [*gvc*] (Stenzel 2015, 222).[10] Nonetheless, a first order of business is to revise the sub-sample particular to S-arguments for the sake of statistical inquiry, such that only those tokens which exhibit the PLAIN zero-marking vs. PROMINENT *-bi* marking alternation are included. This procedure of *cleaning* the data, in order to eliminate *invariable* tokens, is a precursor to performing variationist analyses on any token-based sample (Tagliamonte 2006, 2012). For this chapter, the sub-samples particular to each tested grammatical relation are cleaned prior to implementing the *varbrul* technique to produce statistical evidence.

This section is organized as follows: Section 5.3.2.1 outlines two classes of S-tokens which are excluded from the statistical analysis, and which serves to refine the description of S-oriented DCM accordingly. Section 5.3.2.2, reports and discusses the proportional and statistical evidence results of the *varbrul* analysis.

5.3.2.1 Excluding invariable S-tokens

This chapter recognizes two classes of excluded S-tokens, which do not display the principled PLAIN-PROMINENT alternation, defining S-oriented DCM per **Section 3.1** of the *preliminary description* in this dissertation. The first class pertains to S-arguments participating in a particular dependent-clause construction, which we label SLDC (i.e., *subject-like dependent construction*) in other work (Bruil and Case 2023).[11] The second class of excluded S-token concerns those which are marked with the contrastive topic marker, *=kato*, whose impact on case-marking is discussed in greater detail in the next chapter. Each of these classes is discussed in turn in what follows.

Turning firstly to the SLDC construction, which we identify in a recent talk (Bruil and Case 2023); the crucial fact for the present purposes pertains to the fact that the implicated S-argument displays obligatory *-bi* marking. As such, this construction suppresses the principled DCM alternation characteristic of the S-argument. The SLDC construction refers to a clause-chaining strategy, which is commonplace in Siona narratives, where a de-verbal nominalization serves as the S-argument of a main clause. The referent is interpreted as the S-argument of the de-verbalized event and of the main clause event — i.e., a sort of same-subject interpretation arises. A canonical example is shown in (5), as lifted from *The mother of jaguars* narrative, where the S-argument in (5b) is excluded for the current purposes since zero-marking simply never arises in this construction. This chapter eliminates 48 such tokens from the S-marking sub-sample for statistical analysis:

[10] Consult Table B.9, which provides a more detailed breakdown regarding the distribution of overt and covert S-arguments found in the sample.

[11] In Bruil and Case 2023, we note an instance of PROMINENT *ni*-marking on an SLDC instance with an EXPERIENCER SUBJECT. So far only one instance is attested and requires further investigation.

(5) a. *jɑ̲maca i̲jɛ̲ bainaje̲ nëcaquë ñaëña . . .* [12]
hãmahka [ĩ-i-hẽ Baina-hẽ$_i$] nɨhka-kɨ
then 3PRO-CL:M-ADD B-ADD stand-S.M.DEP.PRS
ñaa-ɨ-ña
look-2/3S.M.PST.N.ASS-REP
'From there he, Baina$_i$, was also standing there, looking out . . . '

b. *. . . nëcaquëbi jaiñë jujujɛ̲̈'ɛ̲̈ ame yëiñë caëña*
nɨhka-kɨ-bi$_i$ hai-ñɨ huhu-hĩ'ɨ̃ a-me jɨi-ñɨ
stand-CL:M-SBJ big-CL:TREE grow-IMPER COP-? cotton-CL:TREE
kaa-ɨ-ña
say-2/3S.M.PST.N.ASS-REP
'. . . (he)$_i$ was standing there (lit. the standing one (M)) and said "grow, grow" and the giant kapok tree grew (they say).'
[NAT*: 20151001oolpi001.092-093]

The second category of excluded S-tokens concerns those arising with the contrastive topic clitic *=kato* (glossed TOP). There turns out to be a general co-occurrence restriction between this nominal discourse marker and the PROMINENT case-marker *-bi*, which is supported by elicitation-based evidence in Section 6.4.2 of the following chapter.[13] (6) illustrates a contiguous stretch of discourse with two S-tokens from *The demon kwɛ̃ɛ̃wahti* narrative. The narrator marks the S-argument in (6b) with the contrastive topic marker *=kato*, which never co-occurs with *-bi* marking on S-nominals. This chapter eliminates 46 such S-tokens from the statistical analysis:

(6) a. *nocabo jɛ̲̈o̲saia baquë i̲ i̲o ëjɛ̲̈maca . . .*
nohka-bo hɨ̃õ + sai-a bah-kɨ [**ĩ-ɨ**
banana-CL:ENCLOS clean + go-NEG be-2/3S.M.PST.N.ASS 3PRO-CL:M
ĩ-o **ɨ̃hĩ-mahka**$_i$]
3PRO-CL:F husband-DIM.SG
'He, her husband$_i$ went to tend to the banana plantation . . . '

b. *. . . i̲omacacato a̲o̲re sëoco ñucoña*
ĩ-o-mahka=kato$_k$ ãõ-re sɨo-ko
3PRO-CL:F-DIM.SG=TOP cassava-N.SBJ fry-S.F.PRS.DEP
ñuh-ko-ña
sit-2/3S.F.PST.N.ASS-REP
'. . . as for her, she$_k$ was sitting, frying casSAva.'
[NAT*: 20151112oespa001.164-165]

On the basis of the two excluded categories of S-token established above, this chapter develops proportional and statistical evidence for S-oriented DCM against a sub-sample which consists of 727 included, *variable* S-tokens.

[12] Baina is the primary benevolent cultural hero in Siona (and Sekoya) mythology and he is ubiquitous in the folktale genre, as discussed in various places in the Vickers (1976) ethnography.

[13] See Table B.10 in the Appendix for a full breakdown regarding the distribution of *=kato* and other discourse particles in the narrative sample.

5.3.2.2 Exploring the *varbrul* results for S-oriented DCM

Table 5.6 provides the condensed *varbrul* report based upon the S-marking sub-sample ($N = 727$), which corresponds to the full *varbrul* report in Table B.6, found in Section B.2.1 of Appendix B. In this sub-sample PROMINENT *-bi* arises on 21.18% of these tokens — i.e., which will be referred to as the 'global PROMINENT-marking rate' in this chapter. This global rate acts as proportional evidence for the present discussion, which is established prior to categorizing the dataset of the basis of TRIGGER CANDIDATE categories. Statistical evidence is gleaned on the basis of TRIGGER CANDIDATE-internal categories, and their respective PROMINENT-marking rates, as displayed in Table 5.6, for significant factors:

TRIGGER CANDIDATE	*N* [% PROMINENT]
GLOBAL S-MARKING RATE	727 [21.18%]
CONTRASTIVITY STATUS	(LARGE EFFECT SIZE)
Contrastive S	83 [95.18%]
Non-contrastive S	644 [11.6%]
SPECIFICITY STATUS	(MODERATE EFFECT SIZE)
Pronominal S	346 [17.6%]
Specific S	273 [26.4%]
Non-specific S	108 [19.4%]
WORD ORDER	(MODERATE EFFECT SIZE)
Pre-verbal S	554 [24%]
Post-verbal S	173 [12.1%]

Table 5.6: Significant factors for S-oriented DCM in *varbrul* sub-sample

As a point of departure, Table 5.6 demonstrates that three tested trigger candidates are highly significant in the S-marking sub-sample (i.e., $p =< 0.001$). By extension, this chapter provides statistical support that the non-significant trigger candidates do *not* factor into Siona S-oriented DCM patterns (i.e., $p => 0.05$). Therefore, these quantitative findings generally complement observation made in earlier descriptions (Chapter 3 in this dissertation; Bruil 2014, §4.4): Neither AGENTIVITY STATUS nor ANIMACY STATUS are active S-marking TRIGGERS in the language.[14] This chapter also finds that neither tested topicality metric is relevant to these patterns — i.e., deviating from topic-driven S-marking patterns with the cognate *-pi* described for the related language, Koreguaje (Cook and Levinsohn 1985, §1).

Turning to the active TRIGGERS in the S-marking sub-sample, the only tested TRIGGER CANDIDATE with a large statistical effect concerns CONTRASTIVITY STATUS. Table 5.6 indicates the nearly all contrastive S-tokens are marked with *-bi*; whereas, of

[14] A handful of examples of both PLAIN zero-marking and PROMINENT *-bi* marking across these non-significant categories are provided in Section B.2.1 in Appendix B.

the 644 non-contrastive S-tokens, only 11.6% of tokens receive PROMINENT *-bi* marking. A clear instance of contrastive S-marking is demonstrated in (7), from *The mother of the moon* narrative, where two established personified bird characters go out to check their traps for food. In this instance, the narrator describes the activities undertaken by the *pɨ'pɨri* (vulture) character in (7a), which happens to be marked with *=kato*, and contrasts these activities directly with those undertaken by the *tãkaro-**bi*** (black vulture, species) character in (7b):

(7) a. *pë'përi ba'iji iñore mëabecato ba'i ñabi. jo'ni ja̲maca i̲o̲maca ba'iona . . .*
pɨ'pɨri$_i$ ba'i-hi [ihño-re mɨabe=kato] ba'i-ɨ
vulture be-3S.M.PRS.ASS here-N.SBJ above=TOP be-S.M.PRS.DEP
ñaa-bi yo'-ni hãmahka **ĩ-o-mahka**$_i$ ba'i-o-na
see-3S.M.PST.ASS do-SS then 3PRO-CL:F-DIM.SG be-S.F.PST.DEP-DS
'The VULture$_i$ was up HERE, looking out. Doing this, as she$_i$ was (there)...'

b. *. . . ja̲re ta̲carobi ña baco. ñani nëcadojaiyo ñani*
hãre **tãhkaro-bi**$_k$ ñaa-a bah-ko ñaa-ni
like_that black_vulture-SBJ see-NEG be-2/3S.F.PST.N.ASS look-SS
nɨhka + dohai-o ñaa-ni
be_standing + wander-S.F.PST.DEP look-SS
'. . . likewise the BLACK vulture$_k$ was looking out. (She$_k$) wandered on foot (i.e., on the ground) to go on the look out . . . '

[NAT*: 20151112orapi001.174-176]

This chapter reports statistical evidence for a very strong correlation between CONTRASTIVITY STATUS and the selection of the PROMINENT-marker, *-bi*; however, this does not mean that PROMINENT-marking is blocked on non-contrastive tokens. In fact, a non-trivial set of naturalistic, non-contrastive S-tokens arise with PROMINENT *-bi* marking. It is also relevant to note that the *varbrul* model, reported in Table 5.6, yields no statistically significant intersective effects; which is taken to indicate that the other noted significant effects shed light on the factors which drive the selection of non-contrastive *-bi* marking. More concretely, the analysis demonstrates that two tested trigger candidates display moderate effect sizes — i.e., SPECIFICITY STATUS and WORD ORDER. Each is discussed in turn in what follows.

Beginning with SPECIFICITY STATUS, this analysis finds a meaningful discrepancy regarding the proportion of PROMINENT *-bi* across the three specificity categories, as conceived here: pronominal S, specific full noun S, and non-specific full noun S. In particular, specific full noun S-tokens bear PROMINENT-marking at an above-average rate, compared with the other categories.[15] This subtle effect is unanticipated — given what we know based upon the *preliminary description* in Chapter 3, specificity is not expected to factor into S-oriented DCM patterns.

For instance, unlike non-SUBJECT DCM, pronominal S is not obligatorily case-marked. (8) juxtaposes two post-verbal personal pronoun S-tokens from the sample.

[15]By design, this statistical study conflates specific indefinites with definites — for the reasons laid out in Section B.1.2 of Appendix B. Impressionistically, the noted discrepancy across these sub-categories is subtle. It is not clear whether the corpus-based variationist approach adopted here is appropriate to study definiteness and its, possibly subtle, effects on case-marking. A more targeted semantically-driven study is likely in order.

In (8a) from *Siona Genesis story* the pronoun is zero-marked; whereas the corresponding S-pronoun from *The demon kwẽẽwahti* narrative in (8b) bears PROMINENT *-bi* marking. In its context, the latter S-token is coded as non-contrastive:

(8) a. *. . . cajë saijëna be'teiña i̱*
kaa-hɨ̱ sai-hɨ̱-na be'te-ɨ̱-ña
say-PL.PRS.DEP go-PL.PRS.DEP-DS follow-2/3S.M.PST.N.ASS-REP
ĩ-ɨ̱
3PRO-CL:M
'. . . (they) spoke and left, and he followed.'
[NAT*: 20151023orocr001.108]

b. *i̱o ëjë cañeje̱ caoña i̱obi*
[ĩ-o ɨ̃hɨ̃ kaa-je-hẽ] kaa-o-ña
3PRO-CL:F husband say-CL:GEN-SIM say-2/3S.F.PST.N.ASS-REP
ĩ-o-bi
3PRO-CL:F-SBJ
'She (i.e., the *kwẽẽwahti* demon) spoke like her husband speaks.'
[NAT*: 20151112oespa001.170]

Besides pronouns, I am unaware of any claims regarding *-bi* as a specificity marker for S-nominals, either in earlier descriptions of Ecuadorian Siona (Bruil 2014, §4.4.1), or for related languages with cognate S-markers, as discussed in Chapter 4.[16] Rather, as discussed at length in the preceding chapter, specificity is claimed to be an active TRIGGER in P-oriented DCM in many Western and Eastern Tukanoan languages. As will be shown, the quantitative analysis in this chapter arrives at the surprising finding that PROMINENT-marking is favoured on specific full noun tokens for both S- and P-oriented DCM patterns, with a compatible MODERATE-SIZE EFFECT across both tested argumental grammatical relations.

The weakest significant effect for S-oriented DCM reported in Table 5.6 concerns WORD ORDER. Although Siona displays a dominant SOV word order, as laid out in Chapter 3, the order of arguments is relatively flexible. This chapter finds statistical evidence that the proportion of PROMINENT-marking on the S-argument is affected by its surface position, relative to the inflected verbal word in the clause. Most S-tokens in the sample arise in a pre-verbal position — i.e., 554 tokens, following the dominant pattern and arising with *-bi* marking at roughly the global marking rate.[17] However, PROMINENT *-bi* marking only arises on 12.1% of the 173 remaining S-tokens in post-verbal position.

To my knowledge, the correlation between word order and S-oriented DCM has not been established in previous research, for Ecuadorian Siona, or for other Western Tukanoan languages. It is not immediately clear how to interpret the correlation between word order and DCM. Nonetheless it is certainly possible that certain factors which are relevant to determining PROMINENT-marking also factor into variable sentential word order patterns.[18] Some hypotheses for this correlation are laid out in

[16] A handful of examples of S-tokens falling within each specificity category is provided in Section B.2.1 of Appendix B.

[17] Table B.9 in Appendix B demonstrates that 76.2% of overt S-tokens in the sample, in main clauses, arise in pre-verbal position.

[18] It may also be relevant to note that most arguments are simply dropped in naturalistic speech. An adequate variationist account of Siona word order effects likely accounts for argument-dropping in some way.

the final section of this chapter, which merit refinement via other techniques. Interestingly, however, this chapter detects a similar WORD ORDER effect, as a TRIGGER CANDIDATE for P-oriented DCM, which points to another commonality across argumental DCM patterns which has previously gone unnoticed. The following section explores the proportional and statistical evidence for P-oriented DCM in the dataset.

5.3.3 P-marking in the Siona narrative dataset

This section is concerned with analyzing patterns of P-oriented DCM which arise in the Siona narrative dataset. Table 5.5 identifies 479 P-tokens in the sample. However, as is established in detail in Section 3.2 of the *preliminary description*, the determination of the proportion of PROMINENT-marking for this sub-sample may not be ascertained by a straightforward mapping of particular surface case-marker onto PLAIN and PROMINENT categories. Following the terminological conventions laid out in Chapter 2, Siona P-oriented DCM exhibits the formal HYBRID-type, where three formal alternatives are mapped onto PLAIN and PROMINENT categories, as determined by an animacy-based noun class split — i.e., zero-marking, *-re*, and *-ni*. On this basis, Figure 5.1 schematizes the distribution of surface case-marking alternatives across animate and inanimate P-tokens in the full Siona narrative dataset:

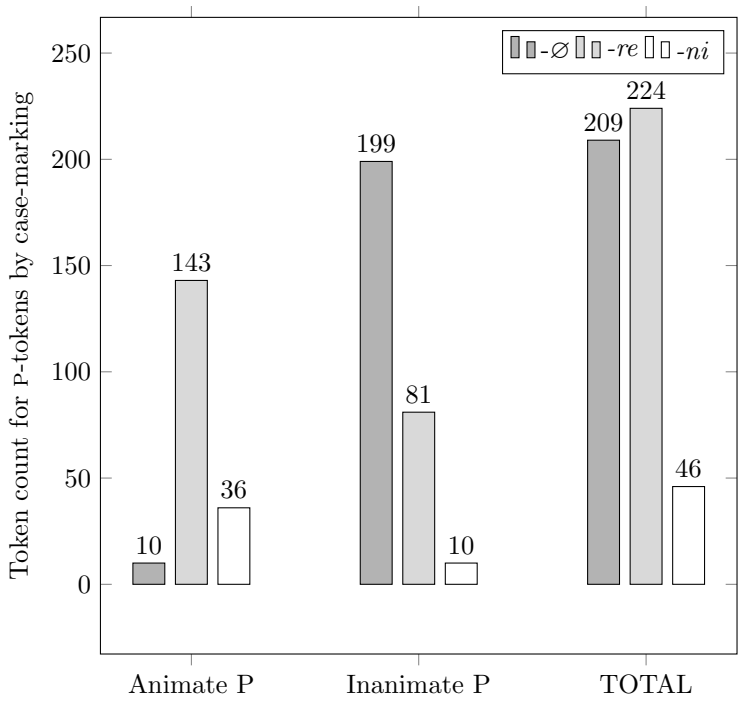

Figure 5.1: P case-marking alternatives across notional animacy categories

Firstly, the small subset of 10 inanimate P-tokens with *-ni* marking are anticipated, given the phenomenon of promotion detailed in Section 3.2.2. Predictably, promoted

inanimate nominals pattern with the animate noun class for the purposes of case-marking. (9) lays out the post hoc conversion process, from surface case-marking to PROMINENT-marking status, which acts as the dependent variable for the analysis developed here — as such, promoted P-tokens are not excluded from the sub-sample:

(9) a. Animate P ⇒ PLAIN *-re* vs. PROMINENT *-ni*
b. ↑Inanimate P ⇒ PLAIN *-re* vs. PROMINENT *-ni*
c. Inanimate P ⇒ PLAIN *-∅* vs. PROMINENT *-re*

Promoted P-tokens display the same principled, *variable* DCM pattern analyzed for canonical P-tokens. (10) provides two instances of promoted pronominal P-tokens: An instance of PROMINENT *-ni* marking from *The children of Baina* narrative is shown in (10a); whereas (10b) from the *Siona Genesis story* demonstrates PLAIN *-re* marking:

(10) a. *... yo'quëna i̲o̲ bë'cacoje̲ caoña iconi gu̲ijë'ë̲ caoña. neo garabëre*
jo'-ki-na ĩ-o bɨ'ka-ko-hẽ
do-S.M.PRS.DEP-DS 3PRO-CL:F parent-CL:F-ADD
kaa-o-ña ↑**ih-ko-ni** gũĩ-hɨ̃'ɨ̃
say-2/3S.F.PST.N.ASS-REP DEM.PROX-CL:F-N.SBJ2 bite-IMPER
kaa-o-ña nee-o gara-bɨ-re
say-2/3S.F.PST.N.ASS-REP make-3S.F.PST.ASS charcoal-CL:RND-N.SBJ
'When (he) did this, her mother said, "bite into THIS!", she had made charcoal.' [NAT*: 20150811sfryi001.092-093]

b. [**Context:** All of the animals in the forest gather to fell a great tree.]
icore cue̲ni je̲u̲ñu'u̲ cue̲ni je̲o̲na ja̲o̲bi ta̲itota̲'a̲ tsiaya ba'ija̲'coa
↑**ih-ko-re** kwẽẽ-ni hẽõ-ñu'ũ kwẽẽ-ni
DEM.PROX-CL:F-N.SBJ take_down-SS leave_behind-HORT take_down-SS
hẽõ-i-na ↑hã-o-bi tãi-to-tã'ã
leave_behind-S.F.PST.DEP-DS DEM.MED-CL:F-SBJ fall-COND-CNT.EXP
tsia-ja ba'i-hã'-ko-a
river-CL:RIV be-PRP-CL:F-COP.3S.F
' "Let's take it (i.e., the tree) down, when (it) comes down, it (lit. she) will be the river." ' [NAT*: 20151023orocr001.251-253]

To confirm, as shown in Figure 5.1, promoted tokens are coded as belonging to the notional inanimate class for the sake of testing ANIMACY STATUS as a TRIGGER CANDIDATE. Both concerned tokens in (10) have inanimate referents, independently from the coding of the PROMINENT-marking status for a particular token.

Whereas this chapter includes all instances of promoted P-tokens (*N=16*); there are two small classes of zero-marked animate P-tokens, identified in Figure 5.1, which are eliminated for the sake of statistical analysis in the present chapter. Section 5.3.3.1 briefly discusses these sets of tokens. Section 5.3.3.2 explores the proportional and statistical evidence produced against the P-marking sub-sample.

5.3.3.1 Excluding invariable and outlier P-tokens

Figure 5.1 identifies ten zero-marked animate P, whose case-marking deviate from the dominant, principled DCM patterns, and which are eliminated from this analysis.

This set of tokens is divided evenly into two sub-categories, each of which informs the broader characterization of P-oriented DCM.

Firstly, as alluded to in Chapter 4, the class of coordinated P-tokens is *invariable* since they simply resist case-marking altogether and are excluded on this basis. A few such instances arise in the sample (N=3), where nominals are coordinated via juxtaposition.[19] Consider (11) lifted from the *Siona Genesis* narration where the corresponding non-conjoined animate P-tokens would require case-marking if they had surfaced independently:

(11) *yo'oquë sai uye naso huasaiya baquë . . .*
jo'o-kɨ̶ sai-ɨ̶ **uje** **nahso** wa'i + sai-a
do-S.M.PRS.DEP go-S.M.PRS.DEP turkey wolly_monkey kill + go-NEG
bah-kɨ̶
be-N2/3S.PST.N.ASS
'. . . doing (this), (he) went out to hunt turkey, monkey(, etc.).'
[NAT*: 20151023orocr001.324]

The second set of excluded P-tokens (N=7) concern the small class of lexical generics (e.g., *hoja* (dogs, domesticated animals), *tsĩ̃ĩ* (children)), whose case-marking facts require further investigation. These nouns do not *require* case-marking, like regular animate nouns in non-SUBJECT positions; albeit they readily accept both *-re* and *-ni* elsewhere. Therefore, following best practices in variationist research, these few unclear instances are eliminated as outliers for the present purposes.[20,21] (12) juxtaposes P-tokens of the zero-marked lexical generic *tsĩ̃ĩ* with the canonical animate *mamakɨ̶-re* (child (M)) in a single episode of *The demon kwẽ̃ẽwahti* narrative:

(12) a. *cani jɑ̱maca nasobi mëco'ë tsɨ̱ gachaco ɨ̱oje. tsɨ̱ hueo yo'ye ba'io. . .*
kaa-ni hãmahka nahso-bi mɨ̶h-ko'ɨ̶ **tsĩ̃ĩ̃**
say-SS then wolly_monkey-SBJ climb-3S.F.PST.ASS child
gahcha-ko ĩ̃-o-hẽ̃ **tsĩ̃ĩ̃** wee-o
put_down-3S.F.PRS.ASS 3PRO-CL:F-ADD child carry-S.F.PRS.DEP
jo'-je ba'i-o
do-CL:GEN be-2/3S.F.PRS.N.ASS
'After (she) said "then the monkey went up (there)", she put down the child$_i$ – she was also carrying a child$_i$.'

[19]The same blocking effect arises where speakers realize coordination with the overt conjunction *kwĩ'ne* (also), which is commonplace in elicited data. Simply put, the presence of coordination bleeds the obligatory case-marking requirement found on animate non-S arguments. This generalization is corroborated by elicited data in my field notes.

[20]In other corpus-driven work, we discuss lexical generics in slightly greater detail (Bruil & Case, *forthcoming*). On the basis of preliminary fieldwork, this pattern is only observed for a small class of nouns, like those mentioned above; whereas generic animates otherwise pattern with animate nouns in general for case-marking purposes (e.g., *wa'i* (fish); *bĩ'ã* (birds)).

[21]One hypothesis: the textual instances which do arise with these generics appear to be in immediately pre-verbal position, and may belong to a separate construction of pseudo-noun incorporation of the DIRECT OBJECT — i.e., not DCM, as discussed by Driemel 2023 for a variety of languages.

b. *. . . jetere hueco io mamaquëre hueco*
hehte-re wee-ko [**ĩ-o mama-ki-re**]
back-N.SBJ carry-3S.F.PRS.DEP PRO3-CL:F child-CL:M-N.SBJ
wee-ko
carry-3S.F.PRS.ASS
'(She) is carrying her child$_i$ on (her) back.'
[NAT*: 20151112oespa001.102-103]

The handful of eliminated tokens discussed here hint at certain questions which remain to be addressed in future research, regarding classes of P-arguments which deviate from the principled DCM outlined in Section 3.2.1. The resulting sub-sample for P-marking consists of 469 P-tokens, classified as bearing their appropriate PLAIN- or PROMINENT-alternative accordingly. The following section considers the proportional and statistical evidence which is yielded against this P-marking sub-sample.

5.3.3.2 Exploring the *varbrul* results for P-oriented DCM

In addition to the relative complexity of P-oriented DCM on formal grounds (HYBRID-type DCM), Chapter 3 recognizes an array of plausible TRIGGERS for the observed DCM patterns, which are not necessarily relevant to S-oriented DCM. For instance, previous descriptions indicate that both specificity- and discourse-related factors conspire to influence case-marking of the P-argument (i.e., Bruil 2014, §4.4.2). This chapter provides statistical evidence which confirms that P-oriented DCM is more complex in this regard. The condensed *varbrul* model in Table 5.7, corresponding to the full report in Table B.7 in Appendix B, demonstrates that more trigger candidates are active in P-oriented DCM, as compared to the S-marking patterns analyzed in the previous section.

Table 5.7 strictly considers significant factors, corresponding to active trigger candidates, in the sample; however, it is also relevant to note that this *varbrul* model provides evidence *against* two tested trigger candidates — i.e., AFFECTEDNESS STATUS and LOOK-BACK TOPICALITY, as coded in this chapter yield insignificant effects (i.e., $p => 0.05$).[22] Therefore, on the one hand, it is not the case that the degree of affectedness of the P-argument affects DCM. On the other hand, regarding topicality, this analysis finds that LOOK-BACK TOPICALITY STATUS — i.e., how recently the referent of a given token is mentioned in the narrative ('look back' topicality per Givón 1976, *et seq.*) is *not* relevant to P-oriented DCM. Conversely, the LOOK-AHEAD TOPICALITY metric is siginificant in the sample, as discussed further below:

[22]The reader is referred to Section B.2.2 in Appendix B for a handful of examples of pertaining to these inacive factors.

TRIGGER CANDIDATE	*N* [% PROMINENT]
GLOBAL P-MARKING RATE	469 [24.27%]
CONTRASTIVITY STATUS	(LARGE EFFECT SIZE)
Contrastive P	33 [87.9%]
Non-contrastive P	436 [19.6%]
LOOK-AHEAD TOPICALITY STATUS	(MODERATE EFFECT SIZE)
Local topical P	45 [51.1%]
Local non-topical P	424 [21.5%]
SPECIFICITY STATUS	(MODERATE EFFECT SIZE)
Pronominal P	106 [20.2%]
Specific P	162 [32.7%]
Non-specific P	201 [19.6%]
WORD ORDER	(MODERATE EFFECT SIZE)
Pre-verbal P	383 [26%]
Post-verbal P	86 [19.7%]
ANIMACY STATUS	(SMALL EFFECT SIZE)
Animate P	179 [19.7%]
Inanimate P	290 [27.2%]

Table 5.7: Significant factors for P-oriented DCM in *varbrul* sub-sample

Firstly, as pertains to the proportional facts, Table 5.7 recognizes a global PROMINENT-marking of 24.27% on the P-marking sub-sample in the Siona narrative dataset. This global marking rate is comparable to the 21.18% marking rate observed in the S-marking sub-sample. This analysis finds that five of the seven tested factors have significant effects on the distribution of PROMINENT-marking, which includes all three which are active in the S-marking sub-sample, in addition to two other effects: including LOOK-AHEAD TOPICALITY STATUS and ANIMACY STATUS. The relative impact of each of these TRIGGERS in the sample is outlined in what follows.

Firstly, as was found with S-oriented DCM patterns in the dataset, CONTRASTIVITY STATUS stands alone as the only independent variable with a large effect size (i.e., a highly significant result, $p =< 0.001$). The small class of contrastive P-tokens are PROMINENT-marked at an especially high rate, as compared with non-contrastive P-tokens.[23] The sentences in (13) demonstrate clear PROMINENT-marking on contrastive

[23]Interestingly, non-contrastive P is marked at roughly the global marking rate, unlike non-contrastive S. There may be several reasons for this fact; however, there are clearly more factors, which impact P-marking generally. The FACTOR WEIGHT (i.e., *varbrul Range*) is higher in the S-marking sub-sample, compared to P-marking, albeit both clearly have large effect sizes in their respective analyses. It is not clear at this stage how much weight, if any, to place on this statistical discrepancy.

P-tokens: (13a) illustrates an instance of *-ni* marking from *The mother of the moon* narrative, where the human referent is contrasted with the other members of the husband's family. (13b) provides an example, from *The children of Baina* narrative, where *-re* marking contrasts one type of meat with other types of food:[24]

(13) a. [**Context:** A recently deceased husband transforms into a wasp to bring his family to the Upper World to join him in the afterlife.]
. . . ja ñacona. ja ioni jëasi'i cani huajëconi jëasi'i cani yo'quëna . . .
ñaa-ko-na hã **ĩ-o-ni** hɨa-si'-i
see-S.F.PRS.DEP-DS DISC.PART 3PRO-CL:F-N.SBJ2 bring_across-FUT-N3S
kaa-ni **wahɨ-ko-ni** hɨa-si'-i kaa-ni yo'-kɨ-na
say-SS alive-CL:F-N.SBJ2 bring_across-FUT-N3S say-SS do-S.M.PRS.DEP-DS
'. . . when (she) saw (this), oh, (he) wanted to bring HER, the living woman, (to the other side). (He) did (this) and then. . . '
[NAT*: 20151112orapi001.081-082]

b. [**Context:** In order to ascend and rejoin their father in the sky, the children of Baina must drink yage.]
. . . hueo yë'cato soquëtene hua'ire aicoa'ë caco jaëhua'ini. iocato goeye baco
wee-o jɨ'ɨ=kato **sõhkɨtẽne + wa'i-re**
lie_in_hammock-S.F.PST.DEP 1S=TOP catfish(spec) + meat-N.SBJ
ãĩ-ko-a'-ɨ kaa-ko hã-ɨ-wa'i-ni
eat-PRP-CL:F-COP.3S.F say-S.F.PRS.DEP DEM.MED-CL:M-PL.AN-N.SBJ2
ĩ-o=kato gwee-je bãã-ko
3PRO-CL:F=TOP not_want-CL:GEN NEG.AUX-S.F.PRS.DEP
'. . . (she) lay (there, in the hammock) and said to them, "As for ME, I eat (only) CATfish meat". She didn't want (i.e., yage).'
[NAT*: 20151001oolpi001.065]

The second statistical effect noted in the P-marking sub-sample concerns another discourse-level TRIGGER CANDIDATE, LOOK-AHEAD TOPICALITY STATUS (i.e., $p =< 0.001$). Unlike the inactive LOOK-BACK TOPICALITY STATUS, which distinguishes tokens with recently mentioned referents, from those with a more distantly mentioned referent; there is statistical evidence for an active LOOK-AHEAD TOPICALITY STATUS effect, demonstrating that P-tokens whose referents persist in immediately ensuing lines of the narrative favour PROMINENT-marking. This medium-sized effect is markedly distinct from that noted for CONTRASTIVITY STATUS above, although, taken together, these demonstrate that the selection of PROMINENT-marking is largely driven by discourse-related TRIGGERS.

Below are two canonical instances of PROMINENT-marked P-tokens, which are non-contrastive local topics, per the definitions assumed in this chapter. On the one hand, (14) demonstrates an instance from *The children of Baina* narrative, where PROMINENT *-ni* is found on the P-argument, which persists as at-issue argument in the following sentence. A similar instance is lifted from *The mother of the moon* narrative in (15): PROMINENT *-re* is used where the *pɨ'pɨri wa'tijo* (the vulture's blade) is retained as an argument in the following sentence, important to the narrative development:

[24]Various instances of P-tokens sorted on the basis of all tested discourse-level trigger candidates are provided in Section B.2.2 of Appendix B.

(14) a. *ja bë'ë tsiayacua'ini tsi'choa baquë . . .*
hã̃ [**bɨ'ɨ tsia-ja-ah-ko-wa'i-ni**] si'cho-a
DISC.PART caiman river-CL:RIV-COP-CL:F-PL.AN-N.SBJ2 gather.TRS-NEG
bah-kɨ
be-2/3S.M.PST.N.ASS
'He gathered the caimans that lived in the (the) river. . . .'

b. *. . . sanijeñaquë caquë saquëña jaona teore cuni tayobi. . .*
sani + hẽ̃ẽ̃ + ñaa-kɨ kaa-kɨ
go + cross + see-S.M.PRS.DEP say-S.M.PRS.DEP
sah-kɨ-ña hã-o-na te'-o-re kũni
go-2/3S.M.PST.N.ASS-REP DEM.MED-CL:F-GOAL one-CL:F-N.SBJ bite
+ tajo-bi
+ tear-3S.M.PST.ASS
'. . . (he) said (he) would go and look, so (he) went there and killed bit one (of them) and tore (them) up.' [NAT*: 20150811sfryi001.427-438]

(15) a. *iconi isisi'i pë'përi hua'tiyore caëña. . .*
↑**ih-ko-ni** ĩhsi-si'-i [**pɨ'pɨri wa'ti-jo-re**]
DEM.PROX-CL:F-N.SBJ2 give-FUT-N3S condor blade-CL:LONG-N.SBJ
kaa-ɨ-ña
say-2/3S.M.PST.N.ASS-REP
'He wanted to give IT (to her), the vulture's sword. . . '

b. *. . . isiña jare isina jamaca caquëna jamaca io go'ihuëocona. . .*
ĩhsi-ɨ-ña hãre ĩhsi-ɨ-na hãmahka
give-2/3S.M.PST.N.ASS-REP like_that give-S.M.PST.DEP-DS then
kaa-kɨ-na hãmahka ĩ-ɨ go'i + wɨo-ko-na
say-S.M.PST.DEP-DS then 3PRO-CL:M return + begin-S.F.PRS.DEP-DS
'. . . (he) gave (it) like that and then, after giving (it), after saying (that), he started to return (to the terrestial world).'
[NAT*: 20151112orapi001.290-294]

Besides these two discourse-level trigger candidates, the next largest effect sizes are found for SPECIFICITY STATUS (i.e., a highly significant factor, $p =< 0.001$) and WORD ORDER (i.e., a significant factor, $p = 0.001$-0.05); demonstrating highly compatible findings to those discussed for the S-marking sub-sample in Section 5.3.2. On the one hand, as regards SPECIFICITY STATUS, the same pattern emerges as discussed above, such that specific full noun P-tokens are marked at a slightly higher rate than their non-specific and pronominal counterparts. This finding indicates that, although several classes of pronouns undergo promotion, which has implications for surface case-marking (see Section 3.2); a stable distribution of PLAIN and PROMINENT alternatives is observed across all types of pronominal tokens in the sample — example (10), shown earlier in this section, demonstrates instances of PLAIN *-re* and PROMINENT *-ni* on promoted pronoun *ihko* (this one/it — lit. this one (F)).

On the other hand, concerning WORD ORDER effects, the same pattern is noted as found with S-marking, although the effect on P-marking is slightly less robust. Given the dominant SOV word order, pre-verbal P-tokens arise more frequently than post-verbal P, as anticipated. More importantly, PROMINENT-marking is found at a higher

rate for the former category, than for the latter.[25] Further remarks about possible explanations for this pattern, particular to argumental DCM patterns in this dataset, are set aside for the discussion section of this chapter in Section 5.4.

There is one more statistically significant TRIGGER CANDIDATE in the P-marking sub-sample, as demonstrated in Table 5.7: i.e., ANIMACY STATUS. The *varbrul* model finds that inanimate P is slightly more likely to be PROMINENT-marked than animate P. The proportional discrepancy across animacy categories is nearly identical to that found for word order categories, despite differences in effect size.[26] To my view, there are two main ways this statistical effect can be interpreted.

Firstly, a *direct* interpretation takes the notional animacy status of a token's referent to influence the proportion of PROMINENT-marking, along the lines of the effect that specificity status is shown to have. Of course, this type of WEAK TRIGGER is described for several related Tukanoan languages, discussed in Chapter 4. Nevertheless these other Tukanoan languages display the more robust cross-linguistic pattern in which the animate noun class favours overt (i.e., PROMINENT-)marking (Moravcsik 1978; Bossong 1984, 1985; Aissen 1999, 2003; Sinnemäki 2014; etc.); whereas the opposite pattern is detected in the P-marking sub-sample for Siona analyzed in this chapter.

A second interpretation of this finding is possible, which I consider to be *indirect* since the mapping of case-marking alternatives to abstract PLAIN and PROMINENT categories, motivated in Chapter 3 is, in a sense, suspended. The idea is that this slight proportional discrepancy across inanimate and animate categories, not found in S-marking, has to do with the fact that canonical inanimate P displays OPTIONAL-type DCM (i.e., PLAIN zero-marking vs. PROMINENT *-re* marking), whereas animate P displays ALTERNATING DCM (i.e., PLAIN *-re* vs. PROMINENT *-ni*).[27] In light of this formal split, it is conceivable that inanimate P-tokens are PROMINENT-marked in some instances in order to avoid zero-marking, for the sake of disambiguation (see Section 2.2.3.2). I return to this hypothesis in the discussion section of this chapter.

Ultimately, even if we are to accept the *indirect* interpretation of this final finding, it is relevant to recall that the effect of ANIMACY STATUS on the proportion of PROMINENT-marking is a subtle one. This fact is not taken for granted a priori. Although there is evidence for uniting PROMINENT *-re* and *-ni* as a single abstract category shown elsewhere in this dissertation — e.g., QUESTION-ANSWER pairs in Chapter 3; this does not indicate that these two markers necessarily display compatible distributions in a naturalistic sample. As such, this chapter is viewed as supporting the analysis of Siona principled DCM patterns as pairs of PLAIN and PROMINENT forms, which goes beyond a simple formal mapping of case-marking forms. Taken together with the previous section on S-marking patterns in the sample, the quantitative analyses considered thus far have demonstrated that a conspiracy of discourse-related TRIG-

[25]Consult Table B.9 in Appendix B for a more detailed breakdown regarding argument-dropping and word order patterns in the narrative sample.

[26]The ANIMACY STATUS effect detected in the P-marking sub-sample is a significant effect (p=0.001−0.05) — i.e., not a *highly* significant statistical effect as found with the other tested trigger candidates in this analysis (p=< 0.001). This fact is not relevant to the discussion at hand, although it further supports the notion that ANIMACY STATUS has only a slight effect on the distribution of PROMINENT-marking in the sample. Crucially, no intersective effects between ANIMACY STATUS and other tested factors are found in the dataset. This finding merits further discussion.

[27]For all intents and purposes, promotion is irrelevant, since it accounts for only a small subset of P-tokens in the sub-sample (i.e., *N=16*).

GERS, SPECIFICITY STATUS and WORD ORDER factor into argumental DCM in Siona narrations. The following section reviews the results of the L-marking sub-sample.

5.3.4 L-marking in the Siona narrative dataset

The final *varbrul* model developed for the present chapter concerns the STATIC LOCATION, the L-argument. According to the raw distributional facts laid out in Table 5.5 at the opening of this results section, there are a total of 162 L-tokens identified in the Siona narrative dataset. Although the set of L-tokens displays a range of surface case-marking patterns, these all conform to the description of spatial DCM spelled out in the *preliminary description* in Chapter 3 — especially Section 3.3. For this reason, as is the case with the P-marking sub-sample in discussed in the previous section, the mapping from surface case-marking to PROMINENT-marking status has a few steps.

Firstly, the dominant pattern for L-oriented DCM is syncretic with the P-oriented pattern. Canonical, non-promoted L-tokens arise with either PLAIN zero-marking, or with PROMINENT *-re* marking. (16) displays instances of canonical L-tokens: (16a) provides a zero-marked, complex L-argument from *The demon kwẽẽwahti* narration; whereas the noun, *ma'ã-**r**e* (on the path) bears PROMINENT *-re* marking in the instance in (16b), lifted from *The children of Baina* narrative:

(16) a. *mëabebi jaro mëabe tuina naso yai cha'caquë cha'caquë daquëña*
mɨabe-bi **hã-ro** **mɨabe** tui-ɨ-na
above-SRC DEM.MED-CL:LOC above sit_on_top-S.M.PST.DEP-DS
nahso + jai cha'ka-kɨ cha'ka-kɨ
wolly_monkey + jaguar jump-S.M.PRS.DEP jump-S.M.PRS.DEP
dah-kɨ-ña
come-2/3S.M.PST.N.ASS-REP
'As (he) sat up there, the monkey jaguar (species) came, jumping and jumping, from above.' [NAT*: 20151112oespa001.115-116]

b. [**Context:** Baina helps his pregnant wife retrieve arari fruit from the trees]
. . . ñacona arari darëse'e baquëña ma'are
ñaa-ko-na arari da'rɨ-si-je
see-S.F.PRS.DEP-DS arari_fruit fall_to_ground-COMP-CL:GEN
bah-kɨ-ña **ma'ã-re**
be-2/3S.M.PST.N.ASS-REP path-N.SBJ
'. . . (she) saw the arari fruits that had fallen, (they) were on the path.' [NAT*: 20150811sfryi001.073]

Another relevant class of L-tokens concerns promoted tokens. Like animate P, promoted L-tokens arise either with PLAIN *-re* marking, as in (17a) from *The children of Baina* narrative; or with PROMINENT *-ni* as in the promoted spatial pronoun in (17b), lifted from a different episode in the same narrative:

(17) a. *. . . ba'ina jo̲o̲re saniti̲'o̲ni abita mëni o̲cuereña*
ba'i-ɨ-na ↑**hã-o-re** sani + tĩ'ã-ni abita
be-S.M.PST.DEP-DS DEM.MED-CL:F-N.SBJ go + find-SS abita_fruit
mɨni + õhkwe-re-ña
go_up + slurp-N2/3S.PST.N.ASS-REP
'. . . (it, the fruit) was up <u>there</u>, so (they) climbed up (there) to find the *abita* fruit and to eat (it) up.' [NAT*: 20150811sfryi001.450-452]

b. *. . . iconi ja'që baquë ? caëña jo̲maca*
↑**ih-ko-ni** ha'-kɨ bah-kɨ
DEM.PROX-CL:F-N.SBJ2 parent-CL:M be-2/3S.M.PST.N.ASS
kaa-ɨ-ña hãmahka
say-2/3S.M.PST.N.ASS-REP then
'. . . then (he) said, "was Dad <u>HERE</u>?" ' [NAT*: 20150811sfryi002.666]

In a similar way to the P-marking sub-sample, when the distinction between canonical L-tokens, as in (16), and promoted L-tokens, as in (17), is taken into account, the PROMINENT-marking status of these tokens may be ascertained. Nonetheless, one final class of L-arguments exists, which do not arise in the context of P-marking, as detailed in Section 3.3.3. These concern L-arguments which display *-na* marking, instead of the typical L-marking patterns outlined above. This quirky case-marking pattern is determined by the predicate — as observed in other languages in the region, per the typological discussion in Chapter 4. (18) demonstrates an instance with PROMINENT *-na*, as found with the the predicate *bɨa-* (stay), arising in the *Siona Genesis story* narrative:

(18) *i̲ jo̲mu ti ja'o daasiquë jo̲re sa'nihuëna bëaëña*
[ĩ-ɨ hãmu ti ha'o daa-sih-kɨ] hãre **sa'niwɨ-na**
3PRO-CL:M armadillo ANA mud bring-COMP-CL:M like_that inside-GOAL
bɨa-ɨ-ña
stay-2/3S.M.PST.N.ASS-REP
'The armadillo, who brought the mud, stayed <u>inside</u> (the river).'
[NAT*: 20151023orocr001.098]

PROMINENT *-na* marking as in (18) alternates with PLAIN zero-marking, such that a principled DCM pattern is maintained (consult Section 3.3.3 for further details). By extension, where promoted L-nominals are concerned, zero-marking is blocked, as anticipated, and *-na* marking is forced where one of the appropriate predicates arises. Consider the promoted spatial pronoun in (19) from the same narrative as above, which exhibits obligatory *-na* marking with the predicate *guja-* (bathe):[28]

(19) *. . . go'ini caquë bo̲'quëña baja'i tsiaya jo̲o̲na. jo̲ë̲hua'i guyayë*
go'i-ni kaa-kɨ ba-∼'-kɨ-ña
return-SS say-S.M.PRS.DEP be-RMT.PST-2/3S.M.PST.N.ASS-REP
ba-ha'i tsia-ja ↑**hã-o-na** hã-ɨ-wa'i
be-3S.M.PST.ASS river-CL:RIV DEM.MED-CL:F-GOAL DEM.MED-CL:M-PL.AN

[28]This is one of the predicates which has been tested in elicitations. Speakers do not accept *-re* on the L-argument with the predicate *guja-* (or by extension *-ni* where the L-argument is promoted).

guja-ji̶
bathe-N3S.PRS.ASS
'. . . (he) came back and said, "There is a river. They are bathing THERE." '
[NAT*: 20151023orocr001.240-242]

Given that promoted L-arguments, with an appropriate predicate, bear obligatory *-na* marking, these tokens are *invariable* and ought to be excluded from the statistical analysis. Only 9 such tokens are eliminated from this study, including the pronominal instance shown in (19).[29] All other classes of L-tokens are included. Table 5.8 lays out the post hoc conversion from surface case-marking to PROMINENT-marking status utilized for the categorization of tokens in the L-marking sub-sample:

	PLAIN alternative	PROMINENT alternative
LOCATION $_{(-na)}$	*-∅*	*-re/-na*
↑LOCATION	*-re*	*-ni*
↑LOCATION$_{-na}$	*-na* (excluded token)	

Table 5.8: Mapping from noun class to PLAIN-PROMINENT alternatives for L-tokens

The conversion outlined in Table 5.8 is applied against the 152 included L-tokens, which comprise the L-marking sub-sample for this chapter. The following section explores the proportional and statistical evidence yielded by the *varbrul* analysis.

5.3.4.1 Exploring the *varbrul* results for L-oriented DCM

As a point of departure, Table 5.9 recognizes the significant factors from the *varbrul* model developed for the L-marking sub-sample, corresponding to the full model reported in Table B.8 in Appendix B. However, before assessing these results, a disclaimer is in order: Due to the relatively small number of tokens in this sub-sample, the statistical model developed here is found to represent a moderate fit for the data, at best.[30] For the present purposes, this means that the *statistical* evidence reported here is preliminary and ought to be taken with a grain of salt. A future research program, with a larger, targeted dataset, is in order so as to confirm and refine the results outlined in this section. In any case, these preliminary quantitative findings contribute to the broader description of Siona DCM aspired to in the present chapter.

Firstly, and less controversially, Table 5.9 demonstrates that the appropriate PROMINENT-marking arises in the L-marking sample at a global marking rate of 52.63%. This marking rate is markedly higher than those observed for argumental DCM patterns in their respective sub-samples: i.e., S-tokens displays a PROMINENT-marking rate of 21.18%, and P-tokens, 24.27%. As hinted at in the introduction of

[29]Table 5.5 identifies a single outlier L-token which bears *-hã'ã* (PATH/LIMIT) marking. This token is also excluded from the study per variationist conventions.

[30]On statistical grounds, this is determined by the *input rate* metric, reported in the full model at Table B.8 in Appendix B. Given that this rate is relatively close to 1 (as opposed to 0), the *varbrul* model is deemed to provide *weak* statistical evidence for any of the findings which it yields. It is also likely that the higher global marking rate implies a lower degree of *variability* in the dataset, which necessitates the assessment of further data to detect effect size.

this chapter, this fact points to a previously unnoticed pattern where spatial DCM exhibits PROMINENT-marking at a far higher rate than argumental DCM in the language. I return to this generalization in the discussion in Section 5.4.

Setting aside the proportional discrepancy between argumental and spatial DCM, Table B.8 reports preliminary statistical evidence for remarkably similar effects for the tested trigger candidates in the L-marking sample and discussed for argumental patterns in the previous sections of this chapter. The TRIGGER CANDIDATE, LOOK-AHEAD TOPICALITY, is retained for the sake of this discussion. However, every L-token coded as topical (based upon the *topic persistence* metric) bears PROMINENT-marking in the sub-sample, I take this finding to be relevant for the present discussion. For this reason, this TRIGGER CANDIDATE is indicated separately from the other significant factors in the table below — i.e., in gray text with the symbol '!' :[31]

TRIGGER CANDIDATE	*N* [% PROMINENT]
GLOBAL L-MARKING RATE	152 [52.63%]
CONTRASTIVITY STATUS	(LARGE EFFECT SIZE)
Contrastive L	20 [90%]
Non-contrastive L	132 [47%]
SPECIFICITY STATUS	(MODERATE EFFECT SIZE)
Pronominal L	65 [46.2%]
Specific L	67 [62.7%]
Non-specific L	20 [19.4%]
!LOCAL TOPICALITY STATUS	(INVARIABLE)
Local topical L	8 [100%]
Post-verbal L	144 [50%]

Table 5.9: Significant factors for L-oriented DCM in *varbrul* sub-sample

Firstly, Table B.8 demonstrates that this *varbrul* model did *not* detect an effect for LOOK-BACK TOPICALITY (i.e., $p => 0.05$); which distinguishes tokens with locations mentioned more recently, from those with distant mentions in a given discourse. Likewise, the thematic effect tested for L-oriented DCM, SUBCATEGORIZATION STATUS, is not statistically significant (i.e., $p => 0.05$). This result points to the fact that L-marking is not significantly different for L-arguments which are implied by the predicate, and for those which are not. Finally, this analysis finds that WORD ORDER does not impact L-oriented DCM.[32]

[31] Many variationists would elect to eliminate this factor from the report, since these are technically *invariable* tokens. Given that this fact is determined *a posteriori*, and is not expected based upon the grammar, per Chapter 3, this type of invariability is informative for the present purposes.

[32] Table B.9 in Appendix B outlines the distribution of core and non-core arguments relative to the main, conjugated, verbal word in main clauses in the Siona narrative dataset. Spatial arguments are collapses together with temporal and adjunctive arguments; however, there is

Turning to trigger candidates with active statistical effects in the L-marking sub-sample, CONTRASTIVITY STATUS exhibits a large effect size (i.e., a highly significant result, $p=< 0.001$). (20) illustrates a pair of prototypical contrastive L-tokens from *The demon kwẽ̃wahti* narrative. In this episode, while the demon attacks one man in his home, the shaman protagonist lies in a hammock in the corner of the room. The narrator uses PROMINENT *-ni* on the two promoted L-tokens to contrast the locations of each of the victims:

(20) *a̲coña i̲re ba̲i̲re te'i yequëje̲ baquëña turihuë hueosiconi. ja̲o̲ni u̲quëña yequë*
ãh-ko-ña ĩ-i-re bãĩ-i-re te'-i
eat-2/3S.F.PST.N.ASS-REP 3PRO-CL:M-N.SBJ person-CL:M-N.SBJ one-CL:M
jehk-i-hẽ bah-ki-ña ↑**[turi-wi**
other-CL:M-ADD be-2/3S.M.PST.N.ASS-REP corner-CL:CONT
weo-sih-ko-ni] ↑**hã-o-ni**
lay_down.TRS-COMP-CL:F-N.SBJ2 DEM.MED-CL:F-N.SBJ2
ũh-ki-ña jehk-i
be_lying-2/3S.M.PST.N.ASS-REP other-CL:M
'(She) ate him, the man. Also another man was in the CORner where he laid down. The other one (M, i.e., the protagonist) was lying THERE.'
[NAT*: 20151112oespa001.024-028]

Example (20) resembles the prototypical instances of contrastive S- and P-tokens presented in the preceding sections. In a similar vein, Table B.8 a similar, highly significant effect is detected for SPECIFICITY STATUS (i.e., $p=< 0.001$), resembling such effects in the argumental DCM sub-samples. Specific full noun L-tokens bear PROMINENT-marking at a markedly higher rate than pronominal and non-specific tokens. This TRIGGER CANDIDATE displays a similar medium-sized effect within the L-marking sub-sample. As such, the propensity for specific noun tokens to favour PROMINENT-markers is common to all statistically tested grammatical relations in this chapter.

Returning to LOOK-AHEAD TOPICALITY STATUS, given that all 8 local topical tokens coded in this modest sample bear PROMINENT-marking, many variationists would eliminate this factor from the *varbrul* report — i.e., technically, no significance value may be ascertained. Given that the *varbrul* model developed for the L-marking sub-sample is preliminary anyhow, this is treated as a relevant finding to be refined in future work. This is the same brand of topicality effect which is active in the P-marking sub-sample — i.e., such that L-tokens which refer to locations that are at-issue in immediately ensuing lines of narrative bear PROMINENT-marking. This pattern is illustrated in the following excerpt of *The children of Baina* narrative, lifted in (21):

(21) a. *iñore ba'ico cani sani saëña . . .*
ihño-re ba'i-ko kaa-ni sani + saa-i-ña
there-N.SBJ live-3S.F.PRS.ASS say-SS go + bring-2/3S.M.PST.N.ASS-REP
' "(She) lives THERE", (he) said, and (he) brought (her there). . . '

still a propensity for nominal tokens to arise before the verb in the sentence, as was found with core arguments. The difference is that this does not appear to have implications for PROMINENT-marking on spatial nouns.

b. *. . . cani saëna tɨ̲'a̲o̲ña – tɨ̲'a̲cona jai tsio baquëña joborare huë'e baquëña*
kaa-ni saa-ɨ-na tĩ'ã-o-ña
say-SS bring-S.M.PST.DEP-DS arrive-2/3S.F.PST.N.ASS-REP
tĩ'ã-ko-na hai tsio bah-kɨ-ña
arrive-S.F.PRS.DEP-DS big garden be-2/3S.M.PST.N.ASS-REP
hobo-reba-re wɨ'e bah-kɨ-ña
halfway-INTENS-N.SBJ house be-2/3S.M.PST.N.ASS-REP
'. . . after (he) said that and brought (her) (there), (she) arrived. (She) arrived and there was a big garden, and there was a house in the middle.'
[NAT*: 20150811sfryi001.450-452]

In spite of the need for further data to strengthen the statistical findings presented in this section, the analysis of the L-marking sub-sample of the Siona narrative dataset contributes preliminary quantitative findings to the broader discussion in this chapter. On the one hand, L-oriented DCM in the sample deviates from argumental DCM patterns as regards their global PROMINENT-marking rates. It is the case that spatial arguments tend to bear overt, PROMINENT-marking, in more non-contrastive instances than is noted for argumental DCM. On the other hand, all grammatical relations which have undergone statistical analysis here demonstrate a strong effect of CONTRASTIVITY STATUS, and a medium-sized effect based upon SPECIFICITY STATUS. These findings are integrated into the broader analysis of Siona DCM in the discussion section below.

5.4 Discussion: Siona DCM in usage

This section summarizes the proportional and statistical evidence assembled in the variationist study in this chapter and bridges these findings into the broader discussion in this dissertation. As a point of departure, Table 5.10 provides a summary for the global PROMINENT-marking rates (i.e., proportional evidence) and the effects detected for each tested TRIGGER CANDIDATE in the present analysis (i.e., statistical evidence), as found in the S-, the P-, and the L-marking sub-samples of the Siona narrative dataset. trigger candidates found to be insignificant are identified as displaying $^{!}$NO EFFECT (i.e., $p => 0.05$); whereas significant trigger candidates are identified by their respective effect sizes. A subscript numeral ranks these effects relative to one another for each given grammatical relation (e.g., 1, 2. . . n):[33]

[33] As is the case throughout this chapter, this discussion collapses significant and highly significant factors, as yielded by the *varbrul* models (p-value $= < 0.05$ are treated as significant). For the present purposes, this distinction is not relevant — however, the reader is directed to the full reports in Appendix B for a more fine-grained demonstration of significance values.

Trigger candidate [Global marking rate]	S-marking [21.18%]	P-marking [24.27%]	L-marking [52.63%]
		Sentence-level factors	
Word order	Moderate effect$_3$	Small effect$_4$	!No effect
Thematic status	!No effect	!No effect	!No effect
		Referent-level factors	
Animacy status	!No effect	Small effect$_5$	n/a
Specificity status	Moderate effect$_2$	Moderate effect$_3$	Moderate effect$_2$
		Discourse-level factors	
Contrastivity status	Large effect$_1$	Large effect$_1$	Large effect$_1$
Look-ahead topicality status	!No effect	Moderate effect$_2$	!Invariable
Look-back topicality status	!No effect	!No effect	!No effect

Table 5.10: Summary of proportional and statistical evidence from corpus study

Firstly, Table 5.10 summarizes the proportional evidence for this analysis, based upon global PROMINENT-marking rates. As suggested in the preceding sections, a non-trivial discrepancy is detected regarding the proportion of PROMINENT-marking found with argumental DCM patterns, compared with spatial DCM in the sample. Concretely, the L-marking sub-sample explored in Section 5.3.4 exhibits a global marking rate which is roughly twice as high as those found for the S-marking sub-sample analyzed in Section 5.3.2, or for P-marking in Section 5.3.3.

As suggested at the opening of Section 5.3, based upon the raw surface case-marking distributions in Table 5.5, further support for the proportional discrepancy between argumental and spatial DCM is found where INDIRECT OBJECT (R) and the spatial GOAL (G) case-marking are taken into account. Each of these grammatical relations displays a principled, PLAIN-PROMINENT DCM pattern in its own right — as laid out in Section 3.2.3 and Section 3.3.2, respectively, in the *preliminary description*. Although the corresponding sub-samples in this analysis were not extensive enough to permit the informative application of the *varbrul* technique, on the basis of the raw distributional facts alone, the division between argumental and spatial DCM regarding PROMINENT-marking rates is borne out. The preliminary proportional figures for these grammatical relations are reiterated in Table 5.11, indicated with the symbol '!' (based upon datasets which are not thoroughly 'cleaned'). These are placed alongside the proportional facts for the tested grammatical relations, and animate and inanimate P-marking rates are presented independently, for the sake of the ensuing discussion:[34]

[34]As regarding the R-tokens in this sample, only animate R is recorded — according to Section 3.2.3, any inanimate R-tokens would reject zero-marking and be excluded in any case. The G-marking figures demonstrated here do *not* account for excluded promoted tokens, which would require *-na* marking in all instances. These rough figures are strictly intended to highlight the discrepancy as a generalization.

Grammatical relation	*N* [% Prominent]
s-marking	727 [21.18%]
p-marking	478 [24.27%]
Animate p	188 [19.68%]
Inanimate p	290 [27.24%]
! r-marking	70 [24.29%]
l-marking	152 [52.63%]
! g-marking	118 [57.8%]

Table 5.11: Proportion of prominent-marking in sample, argumental vs. spatial (! indicates that no corresponding *varbrul* analysis is developed in the current analysis)

The motivation for the above-noted proportional discrepancy is not immediately clear. In elicitation-based research, like that considered in Chapter 3 and Chapter 6, speakers report highly compatible judgments for argumental and spatial dcm: For instance, the use of overt, prominent-marking in out-of-the-blue contexts tends to trigger the *emphatic contrast* reading. Nonetheless, in the naturalistic narrative dataset analyzed in this chapter, spatial nouns differ markedly from argumental nouns on the basis of proportional evidence. Since l- and g-tokens have spatial referents, on conceptual grounds, it is unlikely that their propensity for overt, prominent-marking is driven by pressures to disambiguate them from surrounding (argumental) nominals. At present, the proportional discrepancy summarized in Table 5.11 is taken to be a descriptive generalization, to be more fully unpacked in future research.

The remainder of this discussion considers the statistical evidence put forth in this chapter, and recapitulated in Table 5.10. In light of patterns of overlap regarding the activity of particular trigger candidates, as conceived in this chapter; a discussion by Factor groups (rather than by grammatical relation) is in order. Section 5.4.1 delves into the findings regarding discourse-related Triggers. Then, Section 5.4.2 considers referent-level Triggers. This discussion closes with a consideration of sentence-level Triggers in Section 5.4.3.

5.4.1 Discourse-related Triggers in the dataset

The largest statistical effect found for dcm patterns on s-, p-, and l-oriented nominal tokens concerns Contrastivity status. This analysis adds statistical evidence for contrastivity as a Strong trigger for Siona dcm across-the-board, which is furthered with elicitation-based evidence in the ensuing chapter of this dissertation. To be certain, the inherent connection between prominent-marking and discourse-related Triggers aligns with earlier text-based analyses for dcm in related Siona+ languages, conceived as 'focal', vs. 'plain', case-markers accordingly (Cook and Levinsohn 1985 for Koreguaje [*coe*]; Wheeler 1970, etc., for Colombian Siona, who includes a so-called *emphatic* focal marker category); and the present analysis lends quantitative support to this conception of dcm, which has been adopted from the outset of this dissertation for the analysis of the target variety of Ecuadorian Siona.[35]

[35]In this chapter I am not committed to making a distinction between contrast and focus as the Strong trigger for Siona dcm, since Focus status is not tested here per se. The patterns noted in this quantitative study are largely compatible with those arrived at via

Besides highlighting the uniquely strong effect of CONTRASTIVITY STATUS on the distribution of PROMINENT-marking, this chapter assembles statistical evidence to test for effects related to TOPICALITY STATUS, as recognized in the broader literature. The summary provided in Table 5.10 points to a correlation between topicality and PROMINENT-marking in both the P- and the L-marking sub-samples. As such, the role of topicality in Siona DCM transcends the argumental-spatial divide noted on the basis of proportional evidence, and reinvigorates the SUBJECT vs. non-SUBJECT divide, established in Chapter 3.

Section 5.2.1 (and Section B.1 in Appendix B) lays out how two text-based metrics for testing for topicality effects are implemented in this study: On the one hand, LOOK-BACK TOPICALITY STATUS, based upon the distance between overt mentions, is not significant in the sample. Conversely, what I label LOOK-AHEAD TOPICALITY STATUS, measures (re)activation and persistence of a referent across clauses (related to the effect detected in Romanian *differential object marking*, Chiriacescu and von Heusinger 2010). The latter topicality metric accounts for the second strongest statistical effect in the P-marking sub-sample, and *all* local topical L-tokens are PROMINENT-marked.

A relevant question emerges on the basis of the topicality effects detected in the Siona narrative dataset: Why is LOOK-AHEAD TOPICALITY not active in the S-marking sub-sample? To my view, this discrepancy has to do with the fact that the topicality of the S-argument is encoded via non-DCM mechanics of Siona grammar — i.e., switch-reference markers on dependent verbs, and the related, S-particular, SLDC construction introduced in Section 5.3.2 (see Bruil and Case 2023). As such, strictly non-SUBJECT grammatical relations recruit DCM to signal that their referent is important for the immediately ensuing discourse[36] — i.e., a mechanism to facilitate referential access in the sense of Gernsbacher (1989). It is relevant to note that Cook and Levinsohn (1985) report a similar effect of 'importance to narrative development' as a TRIGGER for 'focal' case-marking in Koreguaje [*coe*]. Though they claim that this effect equally obtains for S-marking via the cognate *-pi* marker, based upon illustrative text-based examples. Nonetheless, as regards non-SUBJECT DCM, highly compatible findings are reported for Siona in this chapter. Discourse-related TRIGGERS outweigh others in Siona DCM.

5.4.2 Referentiality-related Triggers in the dataset

Besides the robust discourse-related DCM effects found here, SPECIFICITY STATUS demonstrates a remarkably stable effect on PROMINENT-marking across each tested sub-sample, independently of other factors. After discourse-related TRIGGERS, the next largest effect in all sub-samples is one where specific full noun tokens are fitted with PROMINENT-marking at a slightly higher rate than do pronominal or non-specific tokens.[37] Whereas this observation is made impressionistically for *-re* marking on inanimate P in Bruil (2014, §4.4.2),[38] the generality of this effect in the Siona narra-

qualitative techniques, where the notion of contrastive focus is at the fore.

[36]It is not clear whether this analysis finds non-SUBJECT PROMINENT-marking to be a bona fide topic shift device. This analysis is perhaps untenable given that only a medium-sized effect is detected in the relevant sub-samples.

[37]As laid out in Section B.1 of Appendix B, this analysis finds no meaningful distinction in the coding of specific indefinite nouns and definite noun phrases, as defined here. This indirectly supports the notion that definiteness is *not* a bona fide category in Siona grammar.

[38]Bruil 2014, §4.4.3, hypothesizes that specificity may also impact the distribution of *-na* marking on the G-argument.

tive dataset is unanticipated. For instance, specificity-triggered SUBJECT-marking is generally taken to be cross-linguistically rare (de Hoop and Lamers 2006; Chappell and Verstraete 2019; cf. Haspelmath 2021 regarding pronoun-full noun distinction). It may be the case that specific nouns are more likely to be important, or for their referents to be contrasted with others; although, crucially, no intersective effects are detected with these factors in the respective *varbrul* models developed for the present analysis. Rather, like the strong contrastivity effect, this subtle specificity effect seems to characterize the PROMINENT-marking category itself, as a component of principled DCM patterns in Siona grammar. This conception is at odds with specificity-related effects noted in the *differential object marking* literature (Moravcsik 1978; Bossong 1985; Aissen 1999, 2003; Sinnemäki 2014; etc.). SPECIFICITY STATUS corresponds to a general WEAK TRIGGER for Siona DCM, which merits further investigation via different techniques.

As regards ANIMACY STATUS, Table 5.10 confirms that no effect is observed in the S-marking sample. This finding aligns with the description for S-oriented DCM laid out in the *preliminary description* (Section 3.1). However, as regards P-oriented DCM, a notably small effect is noted such that inanimate P favours PROMINENT-marking over animate P (reiterated in Table 5.11). In Section 5.3.3.2, it is claimed that this finding likely points to the fact that inanimate P sometimes recruits (PROMINENT) case-marking for the sake of disambiguation, given that the dominant pattern is one where this overt PROMINENT-marker competes with PLAIN zero-marking (disregarding promoted P-tokens: N=16) — i.e., unlike animate P. Of course, the design of this chapter only arrives at this hypothesis indirectly. A targeted study of disambiguation effects in Siona DCM is certainly warranted, since several questions arise which cannot be addressed here: e.g., Are these effects also expected elsewhere where an OPTIONAL-type DCM pattern is attested (i.e., S-marking)? Does disambiguation depend upon the presence of *overt* competitors in the sentence? The role of disambiguation is avoided by design in this dissertation, although it merits further investigation as a component of a holistic description of Siona DCM.

5.4.3 Sentence-level Triggers in the Siona narrative dataset

This chapter tests for two sentence-level trigger candidates: i.e., WORD ORDER and THEMATIC STATUS. Firstly, regarding the latter, the quantitative analysis carried out here finds that there is no correlation between the proportion of PROMINENT-marking and the grammatical relation-appropriate thematic categories. Therefore, as concerns the S-marking sub-sample, the agentive S and the non-agentive S categories demonstrate compatible DCM patterns. The same is found in the P-marking sub-sample, regarding affected and non-affected P categories; and in the L-marking sub-sample as concerns those spatial arguments which are subcategorized for by their predicate, and those which have an adjunctive function. These findings corroborate impressionistic generalizations described for S-oriented DCM and P-oriented DCM in Chapter 3. This finding for L-oriented DCM is a novel one, but fits the following generalization: The thematic properties of the predicate do not have any bearing on DCM patterns found on its arguments in Siona DCM. This generalization matches similar calls for skepticism for thematically-driven DCM patterns in the literature (Fauconnier 2011; Fauconnier and Verstraete 2014).

Turning to the effect of WORD ORDER on Siona DCM, this TRIGGER CANDIDATE is found to be active for both argumental grammatical relations tested in this sample —

i.e., no such effect is detected in the L-marking sub-sample. As regards the S-marking sub-sample, this is a medium-sized effect, but the weakest of the active effects for the tested factors. A small effect size is noted in the P-marking sub-sample, such that only the above-noted effect of ANIMACY STATUS is found to have a smaller effect. These effects are highly similar, such that nouns arising before the main verbal word in the sentence display a higher proportion of PROMINENT-marking. In the general case, this corresponds to PROMINENT-marking being *disfavoured* where nominal tokens arise after the verb. Given the dominant SOV word order in the language, post-verbal argumental noun tokens are always outside of their dominant sentential positions.

To my view, the above-noted subtle word order effect on argumental DCM signals two things about Siona DCM. Firstly, it is conceivable that noun phrases arising outside of their canonical positions might display PROMINENT-marking at an above-average rate given that their grammatical relation is potentially ambiguous. In a sense, the speaker is eliminating one cue for identifying the role of a noun token. The hypothesis is not borne out since the pattern arises in the opposite direction in the present narrative sample.

Another way to interpret the small, yet significant, word order effects is that pre-verbal nominals tend to be those which are more important in one way or another. (Given that there are no intersective effects in the present study, this effect must be rather indirect.) On these grounds, however, it might be expected that similar effects arise for spatial nouns. In any case, a more detailed investigation of the motivations for variation in word order is merited, so that the link between pre-verbal sentential position and PROMINENT case-marking in narration may be better established. Ultimately, word order is taken to be a WEAK TRIGGER for argumental DCM in naturalistic settings, which indicates a fickle connection between DCM and other variable components of the grammar.[39]

In summary, this chapter has shown that a unified account of Siona DCM is tenable, on the basis of a PLAIN-PROMINENT analysis. Certain quirks separate argumental and spatial DCM, particularly on proportional grounds; whereas, in other respects, the SUBJECT vs. non-SUBJECT divide in Siona DCM also has subtle implications for the relevance of particular TRIGGERS. As such, this latter division goes beyond the activity of a formal split based upon animacy classes, as established in Chapter 3, and set aside for the present discussion. Inasmuch as the animacy-conditioned split is a STRONG TRIGGER for Siona DCM, following the conception of multi-dimensional DCM codified by Klein and de Swart (2011); it is different in kind from the STRONG TRIGGER of contrastivity. Whereas the former determines morphological case-marking splits, the latter is the strongest predictor for selecting the appropriate PROMINENT-marking in a given context. The quantitative analysis in this chapter corroborates similar claims made on the basis of elicited data presented in the following chapter.

[39]The conspiracy of these grammatical devices, and probably patterns of argument-dropping, are often viewed as components of discourse-configurationality in grammar. Perhaps a broader investigation along these lines is in order.