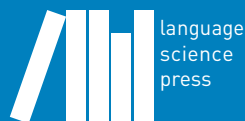


Information structure in Isthmus Zapotec narrative and conversation

Juan José Bueno Holle

Topics at the Grammar-Discourse Inter-
face 99



Series information: ./langsci/seriesinfo/tgdi-info.tex not found!

Information structure in Isthmus Zapotec narrative and conversation

Juan José Bueno Holle

Juan José Bueno Holle. 2018. *Information structure in Isthmus Zapotec narrative and conversation* (Topics at the Grammar-Discourse Interface 99). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/000>

© 2018, Juan José Bueno Holle

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: no digital ISBN

no print ISBNs!

ISSN: 2567-3335

no DOI

Source code available from www.github.com/langsci/000

Collaborative reading: paperhive.org/documents/remote?type=langsci&id=000

Cover and concept of design: Ulrike Harbort

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}TeX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin

no logo

Contents

List of Abbreviations Used in Glosses	v
Orthographic conventions	vii
1 Introduction	1
1.1 Motivation and objectives	1
1.2 Ethnographic setting	4
1.3 Previous work on the language	6
1.4 Methods	7
1.4.1 Corpus creation	8
1.4.2 A discourse corpus	9
1.5 Organization of the study	10
2 Background: the basic grammatical structures of ZAI	13
2.1 The segmental and tonal inventory	13
2.1.1 ZAI segmental inventory	13
2.1.2 The tonal system	15
2.2 The structural function of prosody in ZAI	16
2.2.1 Tones, VQMs and stress	16
2.2.2 Previous studies on Zapotec prosody	18
2.2.3 Prosodic properties of intonation units in ZAI	20
2.2.4 Prosody in ZAI information structure: some initial remarks	22
2.3 Clause structure and constituent order correlations in ZAI	23
2.3.1 Verbal morphology	23
2.3.2 Constituent order correlations	25
2.3.3 Summary of constituent order correlations	34
2.3.4 The pre-verbal position and rigidity in verb-initial syntax	35
2.3.5 The pre-verbal position in ZAI	36
2.4 Summary and research questions	39

3	Preferred Argument Structure and the pragmatic status of nominal forms in ZAI	41
3.1	Preferred Argument Structure in ZAI	41
3.1.1	Data and Methodology	41
3.1.2	Evidence for PAS in ZAI	43
3.1.3	PAS and the notion of Accessibility	47
3.1.4	Accessibility and the introduction of new referents . . .	53
3.1.5	Accessibility and Co-reference	54
3.1.6	LNPs and salience	57
3.1.7	Pronouns and salience	60
3.1.8	Episode boundaries	65
3.1.9	Summary	67
3.2	Nominal forms and the pragmatic status of referents	69
4	Nominal forms in discourse: the alternation of third-person singular pronouns	75
4.1	Syntactic constraints on the overt versus zero alternation	75
4.1.1	Reflexives	75
4.1.2	Dependent clauses	76
4.1.3	Adverbial clauses	78
4.2	The overt versus zero alternation in Pear Story monologue . . .	79
4.3	The overt versus zero form in conversation	82
4.4	Summary and conclusions	85
5	Focus structures in ZAI	87
5.1	Focus structure	88
5.1.1	Predicate focus	88
5.1.2	Sentence focus	92
5.1.3	Argument focus	93
5.1.4	The use of NGA in argument focus	96
5.1.5	Van Valin's (1999) typology of focus structure	100
5.2	Focus structures in discourse: predicate focus plus argument focus	103
5.3	Summary and conclusions	107
6	Topic relations in ZAI	109
6.1	Topic constructions	109
6.1.1	Presentational constructions	110
6.1.2	Topic-comment	111
6.1.3	Identificational constructions	112

6.1.4	Topicalization	115
6.1.5	Detached or LA-marked constructions	117
6.2	Topic relations and the LA particle in discourse	118
6.3	Summary and conclusions	125
7	Conclusions and avenues for further research	127
7.1	Nominal forms and cognitive status	128
7.2	Topic and focus constructions	129
7.3	The LA discourse particle	131
	Appendix A	133
	Appendix B	143

3 Preferred Argument Structure and the pragmatic status of nominal forms in ZAI

In the study of information structure, it is necessary to make a distinction between a) the pragmatic states of the referents of individual sentence constituents in the minds of the speech participants, and b) the pragmatic relations established between these referents and propositions. The focus of this chapter is on the first of these. I will turn to the issue of topic and focus relations in Chapters 5 and 6.

3.1 Preferred Argument Structure in ZAI

This section is concerned with the relationship between the realization of nominal forms and the syntactic role in which they appear. I will frame the analysis using Du Bois's theory of Preferred Argument Structure (Du Bois 1987; Du Bois et al. 2003; Du Bois 2003a,b), with two main goals in mind: 1) to observe the types, frequencies, and syntactic distributions of the nominal forms used by ZAI speakers to satisfy their discursive goals, and 2) to evaluate the capacity of Preferred Argument Structure to account for the patterns observed.

3.1.1 Data and Methodology

The data for this section are made up of narratives elicited from seven ZAI-Spanish bilingual adults between the ages of 25 and 45. To ensure comparability across this and Du Bois and others' work, I asked the consultants to view the Pear film, a short 7-minute film designed for cross-linguistic comparison (Chafe 1980), and then to afterward retell the plot of the story.¹

¹The four main characters in the Pear film are (the abbreviations follow Chafe 1980): Bike boy, Bike girl, Pear man, and the Three boys. The outline of the Pear Story is reproduced here from Chafe (1980: xiii-xiv) for convenience:

The film begins with a man picking pears on a ladder in a tree. He descends the ladder, kneels, and dumps the pears from the pocket of an apron he is wearing into one of three baskets below

I administered the seven interviews and recorded the narratives in Juchitán. At the time of the interviews I had enough knowledge of the language to carry on basic conversations. The speakers I interviewed were all citizens of Juchitán who I saw and spoke to in Isthmus Zapotec on a daily basis and who made regular attempts to help me listen to and understand normal everyday speech. Therefore, although the situation was somewhat unnatural given my lack of native fluency in the language, I do not think this necessarily compromised the naturalness of the recorded narratives. I later transcribed the narratives myself and corroborated my transcriptions with a native ZAI speaker (not one of the seven participants).

As mentioned in §2, I use the “intonation unit” (Chafe 1994) as the basis for the transcription as well as for the analysis below. I understand intonation unit

the tree. He removes a bandana from around his neck and wipes off one of the pears. Then he returns to the ladder and climbs back into the tree. Toward the end of this sequence we hear the sound of a goat, and when the picker is back in the tree a man approaches with a goat on a leash. As they pass by the baskets of pears, the goat strains toward them, but is pulled past by the man and the two of them disappear in the distance.

We see another close-up of the picker at this work, and then we see a boy approaching on a bicycle. He coasts in toward the baskets, stops, gets off his bike, looks up at the picker, puts down his bike, walks toward the baskets, again looking at the picker, picks up a pear, puts it back down, looks once more at the picker, and lifts up a basket full of pears. He puts the basket down near his bike, lifts up the bike and straddles it, picks up the basket and places it on the rack in front of his handle bars, and rides off. We again see the man continuing to pick pears.

The boy is now riding down the road, and we see a pear fall from the basket on his bike. Then we see a girl on a bicycle approaching from the other direction. As they pass, the boy turns to look at the girl, his hat flies off, and the front wheel of his bike hits a rock. The bike falls over, the basket falls off, and the pears spill out onto the ground. The boy extricates himself from under the bike, and brushes off his leg.

In the meantime we hear what turns out to be the sound of a paddleball, and then we see three boys standing there, looking at the bike boy on the ground. The three pick up the scattered pears and put them back in the basket. The bike boy sets his bike upright, and two of the other boys lift the basket of pears back onto it. The bike boy begins walking his bike in the direction he was going, while the three other boys begin walking off in the other direction. As they walk by the bike boy's hat on the road, the boy with the paddleball sees it, picks it up, turns around, and we hear a loud whistle as he signals to the bike boy. The bike boy stops, takes three pears out of the basket, and holds them out as the other boy approaches with the hat. They exchange the pears and the hat, and bike boy keeps going while the boy with the paddleball runs back to his two companions, to each of whom he hands a pear. They continue on, eating their pears.

The scene now changes back to the tree, where we see the picker again descending the ladder. He looks at the two baskets, where earlier there were three, points at them, backs up against the ladder, shakes his head, and tips up his hat. The three boys are now seen approaching, eating their pears. The picker watches them pass by, and they walk off into the distance.

to mean the stretch of speech occurring between two specific prosodic cues: an initial pause and a final lengthening. The reason for this is that intonation units have been shown to operate as a fundamental unit of cognitive processing, social interaction, and other domains, or in Chafe's words, as representing a single "focus of consciousness" (see also DuBois et al. 1993). Since intonation units tend to correspond very closely with simple clause structure, we will see in the vast majority of the examples below that the intonation unit tends to overlap with a core clause (i.e. a predicate plus its nominal arguments) in such a way that the arguments of a clause core fit within the single intonation contour delimited by the intonation unit.²

This study is based on a total of 346 clauses. The Pear Story was chosen as the method of elicitation because of its conduciveness to cross-linguistic comparison. With the exclusion of first and second person arguments, the analysis concentrates on the variety and distribution of third person forms and involves a quantitative study of the nominal forms, as this allows verification of the recurrent role and quantity tendencies predicted by PAS and observed in the ZAI narratives. Given that there are no other existing linguistic studies of ZAI discourse, and despite a significant amount of poetry and literature published in the language, the claims here are preliminary and leave open the question of possible sociolinguistic variation in terms of variables such as genre or dialect.

3.1.2 Evidence for PAS in ZAI

In his theory of Preferred Argument Structure (PAS), Du Bois (1987; 2003a,b) makes specific correlations between discourse patterns and the form of the "core" arguments of the verb. Based on data from narratives in Sakapultek Maya, an ergative language spoken in Guatemala, Du Bois (1987) proposed the set of four closely related grammatical and pragmatic constraints at work in the distribution of arguments in spoken discourse, shown in Table 3.1:

Along the pragmatic dimension, the One New Argument Constraint reflects the tendency for no more than one core argument in a clause to contain new information. Another constraint, the Given A Constraint, states that this new information (typically expressed by full lexical noun phrases) freely appears in the intransitive subject position (the S role) or the transitive object position (the O role), but not in the transitive subject position (the A role).³ Parallel to this,

²There is, however, an important exception to this tendency in the ZAI data presented here. This is the case of "marked topics" or topicalized NPs set off in a separate preceding intonation unit without a verb, which are analyzed in §3.1.7.2.

³The term "core argument" is used in the sense of Dixon (1979), where A refers to the transitive

Table 3.1: Preferred argument structure constraints (Du Bois 2003a: 34)

	Grammar	Pragmatics
Quantity	Avoid more than one lexical core argument “One Lexical Argument Constraint”	Avoid more than one new core argument “One New Argument Constraint”
Role	Avoid lexical A “Nonlexical A Constraint”	Avoid new A “Given A Constraint”

along the grammatical dimension, the One Lexical Argument Constraint refers to the scarcity of clauses in which more than one core argument is expressed with a full noun phrase, the additional core arguments being expressed with pronouns or zero forms. Finally, the Non-lexical A Constraint reflects the tendency for speakers to freely realize full lexical noun phrases in the intransitive subject position (the S role) or the transitive object position (the O role), but strongly avoid placing them in the transitive subject position (the A role).

Thus, the constraints on role refer to the avoidance of lexical/new transitive subjects and the constraints on quantity refer to the avoidance of more than one lexical/new argument in the same clause. The existence of these constraints has been supported by much empirical cross-linguistic research and this has been accepted by many as evidence that PAS is a universal feature of discourse.

The strong tendency for new and lexical arguments to appear in S and O roles and to avoid the A role, though not a categorical rule, has been shown to occur widely in the spontaneous discourse of many typologically diverse languages (e.g. Hebrew, Sakapultek, Papago, English, Spanish, French, Brazilian Portuguese, Japanese, Achenese, Nepali, Finnish and Mapudungun) and in many genres and contexts (e.g. spoken, written, child interaction) (see Du Bois et al. 2003 and contents therein). That said, however, there are a number of studies that question the validity of PAS and its universality (see e.g. Haig & Schnell 2016 and Brickell & Schnell 2017 for recent, well-structured, and insightful critiques.)

As will be seen in the following discussion, the tendencies predicted by PAS do occur widely in third-person narratives in ZAI. Table 3.2 summarizes the distribution across the core clause of full lexical noun phrases (LNP).

Out of 201 total LNPs in the corpus, only 19 occur in the A role. The pattern

subject, O to the transitive object, and S to the intransitive subject.

Table 3.2: Lexical arguments in core grammatical roles

	A role	S role	O role	Total
LNP	9% (19/201)	26%(52/201)	65% (130/201)	100% (201/201)

of distribution of LNPs obeys the Non-lexical A constraint, as predicted by PAS. The majority of LNPs occur in the O role (65%), followed by the S role (26%) and finally the A role (9%). The rate of lexical mentions in the S role thus falls in between the rate of lexical mentions in the O and A roles. Du Bois (2003b: 37) reports similar patterns found in several other unrelated languages, as seen in Table 3.3:⁴

Table 3.3: A cross-linguistic comparison of lexical arguments in core grammatical roles

Language	A role	S role	O role	Total
Hebrew	8% (18/232)	44%(103/232)	48% (111/232)	100% (232/232)
Sakapultek	5% (11/218)	58% (126/218)	37% (81/218)	100% (218/218)
Papago	10% (37/358)	47% (169/358)	42% (152/358)	100% (358/358)
English	8% (21/257)	35% (90/257)	57% (146/257)	100% (257/257)
Spanish	6% (35/591)	36% (215/591)	58% (341/591)	100% (591/591)
French	5% (32/646)	45% (290/646)	50% (324/646)	100% (646/646)
BrPortuguese	8% ()	39% ()	53% ()	100%
Japanese	7% (48/661)	48% (320/661)	44% (293/661)	100% (661/661)

One possible explanation for the scarcity of lexical As could be the scarcity of A positions that appear in the corpus. This does not appear to be the case, however. Of the 346 total clauses attested, 149 (or 43%) are transitive (or ditransitive) clauses, a fairly common proportion in oral speech.⁵ Table 3.4 shows that when we take the number of lexical As as a proportion of total As, the percentage is still significantly low.

When viewed this way, the percentages also increase slightly for the S and O

⁴The data for Sakapultek, Brazilian Portuguese, English and part of the Hebrew data are from narratives elicited from viewers of the Pear Story (Du Bois 2003a: 62-63). Du Bois does not report the exact number of tokens for Brazilian Portuguese.

⁵“Generally one-third to one-half of clauses are transitive versus two-thirds to one-half intransitive” (Du Bois 2003a: 63-64).

Table 3.4: Proportion of lexical arguments per argument position in ZAI

	A	S	O
percent lexical	13% (19/149)	32% (52/165)	77% (130/168)

roles, but the relative proportion of each with respect to each other remains the same. That is, the PAS pattern is clear: the O role houses the highest proportion of lexical arguments, followed by the S role and finally the A role.

The ZAI data also adhere to the two quantity constraints, the One Lexical Argument constraint and the One New Argument constraint. This is illustrated in Table 3.5. Only 18 of the 149 total transitive clauses (12%) have more than one

Table 3.5: Percent of transitive clauses with 0, 1, and 2 lexical arguments in ZAI

	0	1	2
percent lexical	22% (33/149)	66% (98/149)	12% (18/149)

lexical argument. There are no clauses in the corpus which contain more than one new argument.

Finally, with respect to new mentions, a new referent is introduced in A position only two times in the corpus, thus violating the Given A constraint only twice. This is shown in Table 3.6:

Table 3.6: Proportion of new arguments per grammatical role in ZAI

	A	S	O
percent new	1% (2/149)	11% (18/165)	21% (35/168)

In short, we have thus far seen that the ZAI data patterns as predicted by PAS: lexical and new arguments are avoided in A position and the number of clauses with more than one lexical or new argument are very few. Because the number of new referents introduced and the number of clauses used by each speaker will no doubt vary from speaker to speaker depending on factors such as genre or topic, one important issue related to the frequency of lexical and new As is what

Du Bois terms “information pressure”:

When a number of new protagonists are introduced within the space of a few clauses, the *information pressure* (italics mine) is higher than when fewer protagonists are introduced in the same number of clauses-or when the same number of protagonists are introduced in a longer sequence of clauses. (Du Bois 1987: 834)

As Du Bois notes, the issue is especially relevant because in texts with low information pressure, few new or lexical mentions are likely in any grammatical role. Conversely, in texts with high information pressure, many new or lexical mentions are likely in any role. In this corpus, clauses with no lexical arguments are much less frequent than clauses with one lexical argument, as was shown in Table 3.5.

It is an open question, of course, whether this is a generalizable fact about ZAI narrative discourse. If we calculate the “Information Pressure Quotient” (IPQ) for the ZAI data, defined as the total number of new mentions divided by the total number of clauses, we end up with an IPQ of 0.159 (55/346). This IPQ is very similar to the one reported by Du Bois (1987: 834) for Sakapultek Maya, which translates to approximately one new introduction every 6.5 clauses. More likely, however, given the variation in the number of clauses per individual narrative (as high as 74 for one speaker and as low as 24 for another), the degree of information pressure will differ depending on factors such as the genre, the topic, and the individual speaker. We would expect a different corpus with a different degree of information pressure to show a different proportion of clauses with one or zero lexical arguments. Crucially though, due to the two Quantity constraints, we would not expect a higher proportion of clauses with two lexical or new arguments.

Based on the quantitative data reviewed thus far and summarized in Tables 3.2-3.6, it appears that ZAI speakers conform closely to the PAS constraints proposed by Du Bois. But given the amount of cross-linguistic data that has been collected in support of the same discourse tendencies (see Table 3.3 as well as the studies in Du Bois et al. (2003), this does not come as a surprise. The question I would like to pursue in the next section is: *Why?*

3.1.3 PAS and the notion of Accessibility

One of the important insights of PAS, then, has been that there is a cross-linguistic tendency for new and lexical arguments to avoid the A role and to appear most

consistently in the S and O roles. Conversely, there is a tendency for old or given arguments to occur more commonly in the A and S roles.

The question of what the underlying mechanisms are that might be responsible for the PAS patterns observed cross-linguistically is formulated succinctly by Haspelmath (2006: 910). He argues that while the majority of the research supporting PAS assumes the constraints in (3.1) as given, few of the existing studies question whether those constraints do not ultimately reflect other more basic linguistic and cognitive mechanisms underlying discourse.

Haspelmath points out two main issues with PAS. Most critically, he shows that there is a very close relationship between the constraints referring to lexical arguments and those referring to new arguments: *new arguments tend to be coded with full lexical forms* (a connection that was also noted by Du Bois himself (1987: 829-830)). In Haspelmath's view, then, the four constraints could potentially be reduced to just one Quantity constraint and one Role constraint.

Second, Haspelmath raises the important question of whether the Quantity tendencies do not follow straightforwardly from the Role tendencies. That is, if speakers avoid new and/or lexical As, they automatically avoid clauses with two new or lexical core arguments, because there are maximally two core arguments (A and O). Based on this, Haspelmath (2006: 911) suggests that "it may well be that the quantity maxims can be dispensed with, that is, the universally observable quantity tendencies are reducible to whatever explains the [Given A and Non-lexical A constraints]".

So, what might explain the Given A and Non-lexical A constraints? These two constraints can arguably be based on the strong correlation between the A role, animacy and topicality. Since animates tend to be topical and topical entities tend to be coded with non-lexical forms, the two constraints can be shown to be the result of more fundamental properties of discourse, without the need for any independent maxims.

This is one of the main impulses behind a study by Everett (2009), who takes up Haspelmath's main criticisms and argues in favor of an explanation of the deeper generalizations behind the four PAS constraints. In particular, he argues, based on data from English and Portuguese, that the inherent tendency for the A role to be dissociated with lexical and new mentions is motivated by the tendency of the A role to be filled by human referents which are inherently more topical, and for the S and O roles to be filled by non-human referents which are less topical. The data in the following table show the same holds for the ZAI data, at least as far as the A and O roles are concerned:

Although the percentage of human referents in the S role is very high, the point

Table 3.7: Percent human referents per core grammatical role

	A role	S role	O role
percent human	99% (147/149)	88% (146/165)	32% (53/168)

made by Everett still holds: As tend to be topical and represented anaphorically since they typically represent humans (and “humans like to talk about humans”). Os should tend to be new and represented more frequently by lexical arguments since they typically refer to non-humans, which are generally non-topical. Ss represent a middle ground in that they present relatively less human referents than As (and therefore more lexical and new arguments), but more than Os. In other words, for Everett, the observed patterns in the proportion of lexical As, Ss, and Os can be reduced to the factor of human-ness.⁶

Here, I build on the arguments made by Haspelmath (2006) and Everett (2009) (as well as Haig & Schnell 2016 and Brickell & Schnell 2017) and claim that the underlying reasons for the PAS patterns observed cross-linguistically are related to basic discourse-functional factors such as topicality and animacy. In contrast to those authors, I propose a different mechanism responsible for the PAS patterns, that is, that the fundamental mechanism driving the avoidance of new and lexical As in discourse can be shown to be one of accessibility (Ariel 1990; 2001). On the view developed here, the fact that lexical and new referents tend to correlate with grammatical roles in certain predictable ways is due to the degree of accessibility of the referents that appear in the respective grammatical roles.

In the rest of this chapter, I explore the idea that, because new referents are (almost) always coded using lexical arguments, these tendencies can be accounted for based on Ariel’s scalar notion of accessibility (Ariel 1990; 2001): As tend to be highly topical and hence highly accessible and thus rarely new and are rarely coded with full lexical forms; Os tend to be relatively non-topical and hence inaccessible, frequently the locus of introduction for new referents, and thus are often coded using full lexical forms; Ss, frequently topical but also often the stage for new referents, form somewhat of a middle ground.

Ariel’s scalar notion of accessibility is based on the premise that a nominal expression is best characterized as an instruction for the addressee to retrieve a piece of information from either the physical world or the discourse context

⁶See also Haig & Schnell 2016 and Brickell & Schnell 2017 for further empirical, cross-linguistic study and discussion in this direction.

by indicating how accessible or salient the particular piece of information is to the addressee at that particular point in the discourse. From the perspective of accessibility, “nominal expressions are actually accessibility markers” (Ariel 2001: 31).

How do nominal expressions indicate different degrees of accessibility? Ariel (2001: 32) claims that “the more informative, rigid, and unattenuated an expression is, the lower the degree of accessibility it codes, and vice versa, the less informative, rigid, and more attenuated an expression is, the higher the degree of accessibility it codes”. In other words, different nominal expressions have different discourse functions because they are marked for different degrees of accessibility: less attenuated nominal expressions such as LNPs signal less highly accessible or less salient referents while attenuated expressions such as pronouns or zeros signal more highly accessible or more salient referents.

The possible link between Du Bois’ theory of PAS and Ariel’s Accessibility theory has been mentioned sporadically by the authors themselves, but to my mind has not been sufficiently developed. For example, Ariel (2001: 67) states:

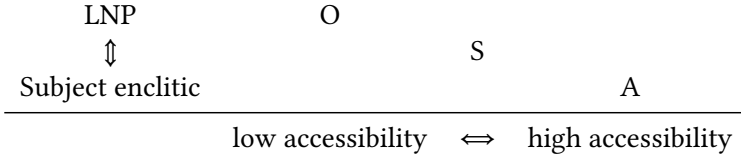
If the motivation [Du Bois] proposes for ergative and accusative markings is based on the lexical versus nonlexical distinction, then it is probably based on the consistently high degree of accessibility of agents versus the inconsistent degree of accessibility associated with intransitive subjects and objects, rather than on the given-new distinction between them.

In later work, Du Bois (2006: 194) remarks that he “started thinking about PAS in terms of accessibility theory and, more specifically, the notion of topicality or salience in terms of high versus low accessibility.” To my knowledge, however, this claim has not yet been forcefully stated in the literature: no detailed studies exist which explore the possibility that the deeper generalization behind the distribution of new and lexical arguments in the A versus the S and O roles is this: accessibility and the cognitive costs associated with different types of nominal expressions.

One goal, then, is to draw a firm connection between degree of accessibility, the forms of nominal expressions and the three core grammatical roles, S, A, and O. In short, the link between PAS and Ariel’s notion of accessibility is this: the O role tends to house low accessible referents that are coded with more linguistic material such as LNPs. The A role tends to house highly accessible referents that are coded with less linguistic material such as zeros. The tendencies for the S role will be found somewhere between these two poles, tending more towards the O role in the marking of new information, but more towards the A role in

contexts of topic continuity, i.e. the marking of topical or human elements. Therefore, I propose that the PAS tendencies can be represented graphically in terms of accessibility in the following way:

(1) Accessibility and PAS



Importantly, Ariel emphasizes that often more than one factor acts simultaneously to affect the degree of accessibility– and thus the form– of nominal expressions. Several of the main factors involved are listed in (2):

(2) Main factors involved in assessing the degree of accessibility⁷ (Ariel 1990; 2001)

- a. Number of previous mentions, i.e. number of new vs. old mentions
- b. Grammatical role, i.e. subject versus non-subject
- c. Animacy
- d. Degree of discourse salience or topicality, i.e. topics vs. non-topics
- e. Recency of mention
- f. Paragraph and frame boundaries, i.e. paragraph-initial positions such as episode boundaries

I have already discussed several of these factors above: two (number of new mentions and grammatical role) are directly mentioned in the PAS constraints, and two (animacy and topicality) are factors that have been suggested (Haspelmath 2006; Everett 2009) to be fundamental in motivating those constraints. The remaining two factors (recency of mention and episode boundaries) are taken up in §3.1.6 and §3.1.8.

⁷This list is not an exhaustive one. For example, in later work, Ariel emphasizes the role that phonetic and intonational cues might play in marking the degree of accessibility of a referent. She mentions Mithun (1995) who shows how the same accessibility marker, a definite NP, can encode different degrees of accessibility through prosodic cues: low degrees of accessibility are encoded by definite NPs which occur in separate intonation units, slightly higher degrees of accessibility are encoded by definite NPs which are not separated by any intonational cues, and high degrees of accessibility are encoded by definite NPs that occur in the more given syntactic position (in Central Pomo) with a specific, unmarked intonation (Ariel 2001: 50).

In the remainder of the chapter, I analyze these accessibility factors with respect to the ZAI data and show that all of the factors, subsumed under the notion of accessibility, not only condition the forms of nominal expressions but also restrict their distribution to specific grammatical roles. I explore the extent to which the gradable notion of accessibility can be shown to underlie the PAS patterns in ZAI, by asking the following questions:

- What types of accessibility markers occur in the corpus in each of the three grammatical roles?
- What are the main accessibility factors involved in determining the distribution of nominal expressions across the three roles?
- To what extent can the notion of accessibility, as a notion that encompasses at least the factors listed above in (2), sufficiently account for the patterns found in the ZAI data?

To answer these questions, each argument in the Pear Story corpus was coded for the following five factors:

(3) Coding scheme

- a. Form of reference: lexical, pronominal, or zero
- b. Core grammatical role: S, A, or O
- c. Animacy: human vs. non-human
- d. Level of salience: New, Previous subject, Active, Old (see 4 for details)
- e. Appearance at episode boundaries

This coding scheme includes each of the accessibility factors listed in (2). It is based on the coding scheme used by Arnold (2003) in her study of constraints on reference form in Mapudungan, but it differs in my formulation of the category Active (see (4) below) and in the inclusion of two categories: animacy and appearance at episode boundaries. To simplify the quantitative analysis, only matrix clauses were included in counting the number of referents that occurred in each of the three roles. Since one focus of this study is the distribution of zero versus overt third person reference forms, I did not want to include cases where either type of mention was disallowed. A more detailed identification of the conditions under which one or other form is used is discussed in §4. For the purposes of the PAS study, however, subordinate and relative clauses, which were very infrequent, were excluded. Finally, given the special nature of “presentational” or

“sentence focus” constructions (“out of the blue” constructions; cf. §6.1.1; §5.1.2) that typically appear at the beginnings of narratives, I have excluded these from the analysis as well. In the majority of cases, the speakers began the narrative with a transitive clause containing a LNP in both the A and the O role (e.g. *cuchugube ti rigola pera* ‘A man is/was picking pears’). Since these types of constructions were not found in other parts of the Pear Story corpus, they are excluded from the analysis (except, of course, in the relevant sections dealing with the introduction of new referents) as they would otherwise have inaccurately biased the data.

3.1.4 Accessibility and the introduction of new referents

In ZAI, singular indefinite referents are typically introduced using *ti* ‘one’ followed by a noun phrase, as in *ti xcuidi* ‘a [certain] child’ or *ti badunguiiu* ‘a [certain] man’. Plural indefinite referents are introduced with a quantifier such as *cadxi* ‘some’ as in *cadxi cuananaxhi* ‘some fruit’. Referents may also be introduced as a bare (uncountable) noun *bicicleta* ‘bicycle’ or within a possessor phrase such as *lari stibe* ‘his shirt’ (cloth + POSS=3SG).

Since new referents are referents that have not previously been introduced to the discourse, we would expect them to be referred to with the lowest accessibility markers, lexical NPs (LNP). This is indeed the case, as is shown in Table 3.8.

Table 3.8: Distribution of new mentions (all referents) by grammatical role

	A role	S role	O role	Total
Lexical NP (LNP)	4% (2/55)	33% (18/55)	64% (35/55)	100% (55/55)
Dependent pronoun (DPR)	0	0	0	0
Independent pronoun (IPR)	0	0	0	0

All new referents are introduced with a lexical NP. The main tendency is for indefinite NPs of the type *ti badunguiiu* ‘INDEF + man’ to be used to mark previously “unidentifiable” and subsequently “activated” referents (Lambrecht 1994). This occurs in 58% (32/55) of the cases. In the remaining cases (42% (23/55)), NPs

preceded by a quantifier, such as *chonna badunguiuhuiini* ‘three + boys’, or bare NPs, such as *pera* ‘pear’, are used.

As is predicted by PAS, the majority of new referents are introduced in the O role, followed by the S role, while only two new referents in the entire corpus are introduced in the A role. This pattern is expected, as is predicted by the graphic in (1): high accessibility markers such as LNPs tend to occur in the O role while low accessibility markers such as pronouns tend to occur in the A role.

Interestingly, this pattern becomes skewed somewhat if we introduce the factor of animacy and consider only the introduction of human referents. This is shown in Table 3.9. When we factor in animacy, the proportion of new referents

Table 3.9: Distribution of new mentions (human referents) by grammatical role

	A role	S role	O role	Total
LNP	7% (2/28)	54% (15/28)	39% (11/28)	100% (28/28)

introduced in each role changes: now, the majority of new human referents are introduced in the S role, followed by the O role, and to a much lesser extent, the A role. The pattern found in Table 3.9 is due to the fact that human participants tend to be more salient and, hence, more accessible than non-human referents, which allows them to be introduced at a higher rate in the S role.

Furthermore, a referent that is introduced in the S role, as opposed to the O role, marks that referent as subsequently more accessible.⁸ This is perhaps most visible when we consider the types of human referents that were introduced in each role. For example, the most salient human participant in the Pear film around whom the majority of the action occurs is the bike boy, who was introduced exclusively in the S role. Meanwhile, the least salient human participant, the bike girl, was introduced exclusively in the O role.

3.1.5 Accessibility and Co-reference

There are significant differences between the forms speakers use to introduce referents and the forms they use to track the referents through the narrative. Whereas new referents are always introduced using LNPs, the array of nominal forms available for coding non-new referents is wider. In this section, I

⁸Du Bois (1987: 831) argues that the S role acts as a cognitive “staging area”. I come back to this idea below.

present data showing that the nominal expressions ZAI speakers employ correlate with the accessibility factors of animacy, topicality, recency of mention, episode boundaries and, crucially, with grammatical role.⁹ The reason that specific types of nominal forms tend strongly to occur in certain core argument positions is because they mark specific levels of accessibility. In particular, we find that low accessibility markers tend to avoid the A role and to occur most regularly in the O role, conversely, that high accessibility markers tend to avoid the O role and to occur most regularly in the A role. The S role, in contrast, tends to house high accessibility markers in contexts of topic continuity and low accessibility markers in contexts of new or marked information.

In the tracking of already-introduced referents, ZAI speakers have two basic anaphoric strategies available: lexical noun phrases (plus a demonstrative) and pronouns (see §3.1.7 for discussion). One of four demonstrative forms may appear after either a singular or a plural noun. The four-way distinction between proximal (for objects near to the speaker), mesioproximal (for objects near to the addressee), mesiodistal (for objects away from both of both speaker and addressee but rather near), and distal (for objects far away from both) demonstratives is shown in Table 3.10:

Table 3.10: ZAI demonstratives

proximal	<i>ri'</i>
mesioproximal	<i>ca</i>
mesiodistal	<i>rica'</i>
distal	<i>que</i>

Plural noun phrases are additionally marked using the plural marker *ca* as in *ca badunguiiu que* ‘those boys’ (lit. PL + boy + DIST) or with a quantifier, as in *chonna badunguiiu que* ‘those three boys’ (lit. three + boy + DIST).

Table 3.11 shows the distribution of each type of form per grammatical role.

Here we see that when we exclude new referents from the count, referents encoded with LNP + Demonstrative, i.e. a low accessibility marker, still occur most often in the O role (65%) and least often in the A role (12%). Within these, the proximal form is used only two times, the medial form only once, and the

⁹It is important to note, however, that although Ariel considers grammatical role a factor in accessibility marking (see (2b)), she does not make the distinction between subject of transitive verbs (A) and subjects of intransitive verbs (S). However, I believe that this distinction is critical in assessing degrees of accessibility, as we will see below.

Table 3.11: Frequency of forms used for co-reference: LNPs + Demonstrative vs. Pronouns

	A	S	O	Total
LNP + DEM	12% (17/146)	23% (34/146)	65% (95/146)	100% (146/146)
Pronouns	46% (130/281)	40% (113/281)	14% (38/281)	100% (281/281)

distal form zero times. The distal demonstrative is by far the most frequent. Also, as we would also expect, referents encoded with pronouns, i.e. high accessibility markers, occur most often in the A role (46%) and least often in the O role (14%).

On additional piece of data worth commenting on here, however, is the differential rate of lexical mention between the A and S roles that shows up in Table 3.11. It appears that transitive subjects (As) are half as likely to be coded using a LNP than are intransitive subjects (Ss). As we saw in §3.1.2, Du Bois (1987) attributes this tendency to the One Lexical Argument Constraint (the tendency to use only one lexical argument per clause). According to Du Bois, this tendency was in turn due to the fact that As tend to be “given” or salient more often than Ss, resulting in a lower rate of lexical reference. As Arnold (2003: 237) argues, if this were the case, if we hold salience constant, we would expect similar rates of lexical reference for A and S. Table 3.12 appears to show that this is not the case. The categories of salience we distinguish here are (in order of low to high accessibility): New, Old, Active, and Previous Subject (further review and description of these categories will be covered below in the next section, 3.1.6).

Table 3.12: Lexical arguments for A and S at each level of salience

	New	Old	Act	PrS
A	100% (2/2)	43% (6/14)	21% (5/24)	6% (6/109)
S	100% (18/18)	74% (23/31)	27% (6/22)	5% (5/94)

Here, the A role contains a significantly lower rate of lexical reference for the level of salience categorized as “Old” and a slightly lower rate for the level “Active”. Therefore, when salience is held constant, LNPs are still used more for S than for A. For Arnold, this is evidence that the One Lexical Argument Constraint cannot be explained based on discourse factors such as topicality.

From the perspective of accessibility, however, this is not necessarily true. One

of the reasons that the high rate of lexical arguments in the S role in “Old” contexts is that more than 40% (10 out of 23) of the tokens are used to refer to non-human referents. In contrast, only 17% (1 out of 6) of the lexical arguments in the A role in “Old” contexts are used to refer to non-human referents. The data in Table 3.12 thus ignores the tendency for human referents to be more salient and, therefore, more likely to be transitive agents (i.e. the *potentiality of agency scale* (Silverstein 1976) than non-human referents. For this reason, I suspect that the different rates of lexical arguments for S than for A is not due to the One Lexical Argument Constraint, as Arnold (2003) claims, but to the independent tendency for the A role to house human, highly salient and, therefore, highly accessible referents.

At this point, it should be clear from this discussion as well as from Table 3.7 (§3.1.3) and Table 3.9 (§3.1.4) that animacy strongly influences accessibility and, hence, the distribution of nominal expressions per grammatical role. In what follows, I examine the categories of full lexical noun phrases (LNP) and pronouns in more detail with respect to two additional accessibility factors, topicality and recency of mention (both captured under the label ‘salience’).

3.1.6 LNPs and salience

We would expect the two accessibility factors of topicality and recency of mention to correlate in predictable ways with the occurrence of LNPs. The effects of these two factors in the ZAI data can be observed through the use of the coding scheme for salience described in (4).

I use the term salience here in the same sense as Arnold (2003) since it effectively combines two of the factors in (2), recency of mention and topicality. The result is a four level scale:

(4) Salience of discourse referents (adapted from Arnold (2003: 231))

- New** = New: referent is brand new to the text.
- Old** = Old: referent had appeared previously in the text, but not in the previous three clauses.
- Act** = Active: referent was last mentioned as either the object of the previous three clauses, as a subpart of the subject or object in the previous three clauses, or both subject and object of the previous three clauses together.¹⁰

¹⁰This category allows for the distinction between the relative discourse prominence of an antecedent that was mentioned in subject position and an antecedent that was mentioned in a

PrS = Previous subject: referent mentioned as subject of the previous clause.

This scheme allows us to observe how referential forms can be simultaneously affected by several discourse constraints. In particular, distinguishing between these four levels in this way allows us to measure differences in salience between both recency of mention (by comparing “Previous Subject” with “Active” and “Old”) and topicality (by comparing “Previous Subject” with “Active”). I thus assume salience to be a gradable scale (Hopper & Thompson 1980)– where referents can be more or less salient– and for the relative values on this scale to coincide directly with those on the scale of accessibility– where referents can be more or less accessible.

First, with respect to recency of mention, reference to something in the previous three clauses (“PrS” and “Act”) is less likely to be encoded with a LNP than reference to something prior to those three clauses (“New” and “Old”). This is shown in Table 3.13. Of all the LNPs in the corpus, three times as many occurred

Table 3.13: Percent of LNPs and recency of mention

Reference to:	% lexical
Entities in the previous clause or previous three clauses (PrS + Act)	26% (53/201)
Entities prior to three clauses (New + Old)	74% (148/201)

in “New” and “Old” contexts than in “PrS” and “Act” contexts. In other words, more recent mentions are less likely to be coded with a LNP than are less recent mentions.

Second, with respect to topicality, reference to a subject (A or S) in the previous clause or in the previous three clauses (PrS) is less likely to be encoded with a LNP than reference to a non-subject in any of the previous three clauses (Act). This is shown in Table 3.14. Of the LNPs in the corpus, three times as many occurred in “Act” contexts than in “PrS” contexts. That is, more topical referents are less likely to be coded with a LNP than are less topical referents.

non-subject position (Arnold 2003: 226). I have decided to adjust this category slightly from Arnold (2003: 231) formulation to include the previous three clauses (and not only the previous clause), because I think it more accurately describes the patterns observed in the data, particularly the distribution of pronouns and the demonstratives, discussed below.

Table 3.14: Percent of LNPs and topicality

Reference to:	% lexical
Subject (A or S) of the previous clause or previous three clauses (PrS)	28% (15/53)
Non-subject in any of the previous three clauses (Act)	72% (38/53)

Based on these correlations as well as those we have set up between low degrees of accessibility, LNPs and the O role on one hand and high degrees of accessibility, pronouns and the A role on the other, we would expect recency of mention and topicality to also correlate with grammatical role in the following way: referents that occur in the O role will be less topical and less recent (and coded as “New” or “Old”) and referents that occur the A role will be more topical and recent (and coded as “Previous Subject”). Table 3.15 shows that this pattern indeed holds for the ZAI data.

Table 3.15: Frequency of referents in each category of salience

	A	S	O
New	1% (2/149)	11% (18/165)	21% (35/168)
Old	9% (14/149)	19% (31/165)	44% (74/168)
Act	16% (24/149)	13% (22/165)	30% (50/168)
PrS	74% (109/149)	57% (94/165)	5% (9/168)
Total	100% (149/149)	100% (165/165)	100% (168/168)

Conversely, we would also expect the majority of less topical and less recent arguments, such as those found in “New” and “Old” contexts, to occur in the O role and for the majority of more topical and more recent arguments, such as those found in “Previous Subject” contexts, to occur in the A role. This is also what we find, as shown in Table 3.16. The A role appears specialized for more topical and more recent mentions, while the O role is more specialized for mentions that are less topical and less recent.

Finally, we would predict the tendencies shown in Tables 3.15 and 3.16 to correlate with particular types of nominal expressions. That is, we would predict low accessibility markers such as LNPs to occur most often in contexts categorized as “New” and “Old” and high accessibility markers such as pronouns to occur

Table 3.16: Frequency of referents in each category of salience

	A	S	O	Total
New	4% (2/55)	33% (18/55)	63% (35/55)	100% (55/55)
Old	12% (14/119)	26% (31/119)	62% (74/119)	100% (119/119)
Act	25% (24/96)	23% (22/96)	52% (50/96)	100% (96/96)
PrS	51% (109/212)	44% (94/212)	5% (9/212)	100% (212/212)

most often in “Previous Subject” contexts. As Table 3.17 shows, this is also what we find.

Table 3.17: Type of nominal expression per category of salience

	New	Old	Act	PrS
LNP + DEM	100% (55/55)	88% (93/106)	53% (38/68)	7% (15/202)
PRONOUNS	0% (0/55)	12% (13/106)	47% (30/68)	93% (187/202)
Total	100% (55/55)	100% (106/106)	100% (68/68)	100% (202/202)

The inverse relation that exists between degrees of salience (defined in terms of topicality and recency of mention) and rates of LNPs should be clear: a high degree of salience and accessibility correlates with a low rate of LNPs and a low degree of salience and accessibility correlates with a high rate of LNPs. Further, the relation should also be clear between high rates of LNPs and the O role and low rates of LNPs and the A role. In the next section, I analyze the relation between degrees of salience and the distribution of higher accessibility expressions, i.e. pronouns.

3.1.7 Pronouns and salience

The ZAI pronominal system is summarized in Table ?? . This system does not distinguish between masculine and feminine, or between formal and informal. The third person pronoun differentiates between human, animal, and inanimate. In addition, first person plural distinguishes between inclusive and exclusive.

Although NPs do not get marked for case in ZAI, pronouns do have independent and dependent forms that are sensitive to their grammatical position within the clause. Dependent forms occur immediately after the verb or noun. In all

Table 3.18: The ZAI pronominal system

	Dependent form	Independent form
1SG	<i>-a'</i>	<i>naa</i>
2SG	<i>-lu'</i>	<i>lii</i>
3SG.HUM	<i>-be, -∅</i>	<i>laa-be, laa-∅</i>
3SG.ANIM	<i>-me, -∅</i>	<i>laa-me, laa-∅</i>
3SG.INAN	<i>-ni, -∅</i>	<i>laa-ni, ni, laa-∅</i>
1PL.INCL	<i>-nu</i>	<i>laa-nu</i>
1PL.EXCL	<i>-du</i>	<i>laa-du</i>
2PL	<i>-tu</i>	<i>laa-tu</i>
3PL.HUM	<i>-ca-be, -ca-∅</i>	<i>laa-ca-be, laa-ca-∅</i>
3PL.ANIM	<i>-ca-me, -ca-∅</i>	<i>laa-ca-me, laa-ca-∅</i>
3PL.INAN	<i>-ca-ni, -ca-∅</i>	<i>laa-ca-ni, -cani, laa-ca-∅</i>

other positions, the independent form is used which is comprised of a base form *laa* plus the dependent pronoun. For example, the third person singular pronoun can be realized as an overt form or as a zero form and, when used in object position, before the verb, or in isolation, the pronoun base *laa* carries the dependent pronoun.¹¹ The dependent forms mark already activated referents, i.e. they mark continuing topics. These forms do not mark the same contrasts as the independent forms, which can function as either topical or focal expressions. In a canonical verb-initial clause, pronouns in the S and A roles appear in the dependent form as enclitics on the verb. Pronouns in the O role occur in the independent form after the subject.

In the remainder of this section, I focus on two main distinctions that appear in (3.1.7). First, I compare the distribution in the Pear Story corpus of the overt third-person singular dependent form, *=be*, to that of the zero form, *=∅*. Second, I analyze the distribution of dependent pronouns versus independent pronouns.

3.1.7.1 Distribution of third-person dependent pronouns: overt vs. zero

In simple intransitive (5 – 6) or simple transitive constructions (7 – 8), the choice between the overt or the zero form of the pronominal subject clitic is free (see Marlett & Pickett (1996)):

¹¹The option to use an independent form for the A or S role, as in the case of “marked topics”, does exist. These cases will be discussed in more detail below.

- (5) biababe láyu
 bi-aba=*be*^{LH} layu
 COMPL-fall=3SG ground
 ‘S/he fell on the ground.’
- (6) biaba layu
 bi-aba=∅ layú
 COMPL-fall=3SG ground
 ‘S/he fell on the ground.’
- (7) biiyabe bá’du quě
 bi-iya=*be*^{LH} ba’du’ que^{LH}
 COMPL-see=3SG child DIST
 ‘S/he saw the child.’
- (8) biiya ba’du quě
 bi-iya=∅ ba’du’ que^{LH}
 COMPL-see=3SG child DIST
 ‘S/he saw the child.’

The intransitive clauses in (5) and (6) convey the same propositional content. However, whereas in (5) the S role is occupied by the overt third person pronoun *=be*, in (6) the S role is occupied by the zero form. This alternation is possible in transitive clauses as well, as is shown in (7), which contains the overt form, and (8), which contains the zero form.

If the choice between the two forms is indeed free at the level of the main clause, it is important to consider the discourse conditions are under which each of the two forms is used. One possibility is that the distribution of the forms is conditioned by grammatical role. This possibility is explored in Table 3.19. What

Table 3.19: Frequency of third-person singular overt vs. zero DPR per grammatical role

	A	S
=BE	78% (73/93)	73 % (58/79)
=∅	22% (20/93)	27% (21/79)
Total	100% (93/93)	100% (79/79)

emerges from this table is a strong preference for overt marking. However, although there may be a slight preference for the overt form to appear in the A role, there does not seem to be a significant difference between the two forms in the grammatical role with which they are associated.

A second possibility is that the distribution of the overt versus the zero form correlates with one or more levels of salience. This is represented in Table 3.20. These data show that zero forms are much more restricted in terms of the degree

Table 3.20: Frequency of third-person singular overt vs. zero for each level of salience

	New	Old	Act	PrS
=BE	0	73% (8/11)	90% (18/20)	74% (106/144)
=∅	0	27% (3/11)	10% (2/20)	26% (38/144)
Total	0	100% (11/11)	100% (20/20)	100% (144/144)

of salience compared to the overt forms. That is, while overt pronouns may occur somewhat freely at each level of salience (except, of course, for “New”), zero pronouns appear to be much more restricted to “PrS” contexts– there are only five total uses of the zero form outside of “PrS” contexts.

Here for “Old” and “PrS” the distribution is very similar to Table 3.19 (in PrS it is basically identical). Only the numbers reported for “Act” stand out. This pattern would appear to imply that topicality and not recency of mention is the crucial factor in whether the zero form is employed. That is, the use of the zero form, higher on the accessibility scale than the overt form, is restricted to highly topical referents whereas the overt form may be used for either highly topical or recently mentioned referents. For the purposes of this section, I leave this question unresolved for now and return to it in §4, where I argue that the distribution of the two forms is related to a distinction between primary and secondary topic.

3.1.7.2 Independent pronouns in the A or S role

While the most common way to refer to subjects in the A or S role is through the use of dependent pronouns, it is also possible in ZAI to use an independent pronoun in pre-verbal position. These are cases that Du Bois terms “marked topics”: “NPs which are topicalized and set off in a separate intonation unit without a verb, and usually precede a predication about the same referent in the imme-

diately following clausal intonation unit” (Du Bois 1987: 814, note 11)).¹² In the ZAI data, there are 25 instances in which an independent pronoun is used in this way. Consider the following example:

- (9) 01 biabandabě
bi-abanda=be^{LH}
COMPL-fall.hard=3SG
'He fell.'
- 02 birěeche dxúmí pěrá stí=bě
bi-ree^{LH}che^{LH} dxumi^{LH} pe^{LH}ra sti^{LH}=be^{LH}
COMPL-spill basket pear POSS-3SG
'His basket of pears spilled.'
- 03 laabe lá,
laa=be^{LH} la^H
BASE=3SG LA
'As for him,'
- 04 biiyadxisibé bádudxaapahuiini quě
bi-iyadxisi^{LH}=be^{LH} badudxaapa-huiini que^{LH}
COMPL-see.fixedly=3SG girl-DIM DIST
'he looked fixedly at the little girl.'

Here, the subject of the intransitive verb in the first intonation unit is the bike boy and the subject of the intransitive verb in the second intonation unit is the basket of pears. In the immediately following intonation unit, line 3, the third person singular independent pronoun is used to refer to the bike boy, followed by the particle LA.¹³ The marked topic in the third line therefore helps to signal the change in subject from the basket back to the bike boy. The last intonation unit, line 4, consists of a transitive clause in which the A role is filled by the third person singular pronoun =be and the O role by a LNP that refers to the girl. Of the 25 instances in which this strategy is used in the corpus, 20 (or 80%) signal a change in subject from the previous sentence.

¹²Importantly, for the purposes of coding the data, marked topics were treated as one mention (of an independent pronoun), not two mentions (one independent pronoun plus one dependent pronoun).

¹³The LA particle always appears at the end of an intonation unit. It appears in 23 of the 25 tokens in which the marked topic strategy is used. It also appears consistently at the end when-clauses and if-clauses. One possibility, then, is that it is used as a topic or contrastive topic marker. This issue will be taken up again in §6.2.

Contrastive forms such as these are generally used in contexts where there is a switch in subject from the previous sentence because they signal referents that are not predicted to occur in particular roles. The account sketched here based on accessibility in fact predicts this to be the case. Ariel (2001: 37) states that, “when an entity is not predicted to appear in a certain role, its degree of accessibility is (relatively) low.” In other words, despite having the exact same form, marked topics with topicalized IPRs indicate a lower degree of accessibility (i.e. they signal a change in subject) than do IPRs in their more common O role position.

To this point, I have tried to show that there exists a strong correlation in the ZAI data between nominal expressions such as LNPs, overt and zero dependent pronouns, and independent pronouns on the one hand and certain grammatical roles (S, A, or O) on the other. Further, I have argued that the reasons for the strong correlation can be traced to different degrees of salience that are associated with the grammatical roles in which the nominal expressions are used. Overt and zero dependent pronouns are preferred over LNPs in the S and A roles because those roles tend to house more salient referents. In contrast, independent pronouns and LNPs are preferred in the O role because of the tendency for the O role to house less salient referents. In the next section, I conclude this analysis by looking closely at one additional factor involved in the distribution of these nominal expressions across grammatical roles: episode boundaries.

3.1.8 Episode boundaries

Do speakers use different nominal forms according to different episode boundaries? We can distinguish five main episode boundaries that each of the speakers marked in their narratives about the Pear film. These are listed in (10):

- (10) Five episode boundaries
1. The Pear man is picking pears.
 2. The Bike boy passes by on his bike and steals a basket of pears.
 3. The Bike boy passes the Bike girl, hits a rock and falls.
 4. Three boys appear and help the boy get up and pick up the pears that spilled.
 5. The Three boys walk away, passing the Pear man by the pear tree

Out of the 35 episode boundaries in the seven narratives, 16 were marked with an intransitive clause and 19 with a transitive clause.

Since low accessibility markers regularly occur in paragraph-initial positions such as episode boundaries (Ariel 2001: 52), we would expect the clauses at episode boundaries to contain higher proportions of LNPs in the A and S roles than throughout the rest of the narratives. This is in fact the case. In Table 3.21, we see that the majority of the arguments (75%) that appear in the S role at episode boundaries are coded with a LNP. More significantly, all of the LNPs that occur

Table 3.21: Lexical arguments at episode boundaries, intransitive clauses

new lexical S	75% (12/16)
non-new lexical S	0
non lexical S	25% (4/16)

in the S role at episode boundaries introduce new referents. Moreover, of the 18 total new LNPs introduced in S position in the entire corpus, 12 (or 67%) occur at episode boundaries. This conforms to the observation by Du Bois (1987: 831) that the S position acts as a cognitive “staging area” for the introduction of referents that are later tracked through combinations of transitive and intransitive clauses.

We also see a higher percentage of LNPs in transitive clauses at episode boundaries. This is shown in Table 3.22. LNPs occur at a much lower rate in the A role

Table 3.22: Lexical arguments at episode boundaries, transitive clauses

	new lexical A	non-new lexical A	non-lexical A
new lexical O	0	1	11
non-new lexical O	1	4	1
non lexical O	1	0	0

than in the S role, even at episode boundaries. However, 7 of the 19 total As at episode boundaries are LNPs. This percentage (37%) is much higher than the percentage of lexical As found overall. In addition, it is interesting to note that of the two new lexical As that appear in the entire corpus, both occur at episode boundaries.

In summary, LNPs in the A and S roles occur at a much higher rate at episode boundaries than they do at other parts of the narratives. I propose that the reason

for this pattern can be also explained in terms of accessibility: episode boundaries are cross-linguistically very common sites for the occurrence of low accessibility markers (Ariel 2001: 52; see also Downing 1980).

3.1.9 Summary

The ZAI data patterns as predicted by PAS: lexical and new arguments are avoided in A position and the number of clauses with more than one lexical or new argument are extremely rare. The question this chapter has been concerned with is: *Why?* Why should the four PAS constraints hold in ZAI as well as cross-linguistically? How are they to be explained? Are the constraints discursively motivated? If so, what are these motivations?

Other researchers (e.g. Haspelmath 2006; Everett 2009; Haig & Schnell 2016; Brickell & Schnell 2017), however, have pointed out that the cross-linguistic tendency to observe these constraints may in fact be due to more fundamental generalizations about the nature of discourse. Three main observations stand out. First, there is a well-established correlation that exists between human, topical referents and the A role in transitive clauses. Second, cross-linguistically what lexical arguments have in common with new arguments is that it is precisely full lexical forms that are used to introduce and track less-accessible (Ariel 1990) referents, i.e. new information. This conforms to the more general observation in the literature that the use of more coding material, i.e. fuller nominal forms, correlates strongly with referents that are lower on the accessibility scale (Givon 1983).

This chapter has presented discourse data from ZAI and has argued, in line with Haspelmath (2006), Everett (2009), Haig & Schnell (2016), and Brickell & Schnell (2017) that the constraints on new arguments and new As can be viewed as a subset of the constraints on lexical arguments and lexical As. I have proposed that the fundamental mechanism driving the tendencies captured by PAS can be traced to the notion of accessibility (Ariel 1990; 2001). This mechanism may be summarized as a reduction of the four PAS constraints to a single constraint that refers directly to the accessibility of referents in the A role: *Avoid low-accessible As*. In other words, the avoidance of new referents and LNPs in the A role can be understood as an avoidance of referents with a low degree of accessibility in that role. That this should be the case is natural given the factors involved in determining a referent's accessibility (as listed above in (2)): newly mentioned vs. already mentioned, non-subject vs. subject, animacy, topicality, recency of mention, and episode boundaries.

Highly accessible referents are referents that have already been mentioned,

subjects, animate, topical, recently mentioned, and/or that do not tend to appear at episode boundaries. These are represented with relatively little *coding material* (Givon 1983). In contrast, low accessible referents are referents that are new mentions, non-subjects, inanimate, non-topical, not recently mentioned, and/or that tend to appear at episode boundaries. These are represented with relatively more coding material. Most significantly, this correlates with grammatical role: while highly accessible referents are very likely to occur in the A role, low accessible referents are very *unlikely* to occur in the A role. The correlations between accessibility factors, nominal expressions and grammatical role are summarized in (3.23):

Table 3.23: Accessibility scale for ZAI nominal expressions

	<u>low accessibility</u>	<u>high accessibility</u>
accessibility factors	newly mentioned non-subject inanimate non-topical not mentioned recently at episode boundary	already mentioned subject animate topical recently mentioned not at episode boundary
type of referring expression	INDEF + LNP > LNP + DEM > IPR > overt DPR > zero	
grammatical role	O	S A

These patterns are corroborated in the ZAI data presented above. On one hand, new and/or lexical arguments are low on the accessibility scale and tend to be referred to using the forms ‘INDEF + LNP’ and ‘LNP + DEM’. These occur most commonly in the O role. On the other hand, already introduced referents are high on the accessibility scale and tend to be referred to using more attenuated pronominal forms. These occur most commonly in the S or A role.

Interestingly, independent pronouns occupy a kind of middle ground since they are usually used to refer to objects which tend to be less accessible than subjects, but, as in the case of “marked topics”, they can also be used to refer to subjects that are relatively less accessible, i.e. subjects that are not particularly salient at a certain moment in the discourse and/or subjects that occur at episode boundaries. The function of this construction in these cases is one of topic promotion (this construction will be an important part of the discussion of topic relations in §6).

Similarly, the S role also has an intermediate function between the O and A

role. The S role will often house previously mentioned, animate, salient, topical, and recent referents but, as we saw, it also frequently functions as a “cognitive staging area” for the introduction of new referents at episode boundaries.

In the next section, I move away from the analysis of Preferred Argument Structure and accessibility to examine the relationship between nominal forms and the pragmatic status of referents.

3.2 Nominal forms and the pragmatic status of referents

As we have seen throughout the course of this chapter, the forms of nominal expressions that speakers use depend on the assumed cognitive status of the referents, that is, on assumptions that a speaker can reasonably make regarding the addressee’s knowledge and attention state in the specific context in which nominal expressions are used (cf. Chafe 1976; Prince 1981; Ariel 1988; *inter alia*). Certain correlations therefore hold in ZAI between the formal category and the pragmatic status of the referents such that the lexical form of an NP may convey either 1) a request to the hearer to act as if the NP were already pragmatically available or “given”, albeit to varying degrees, or 2) a request to the hearer to act as if the NP constitutes unavailable or “new” information. The various nominal forms in ZAI, namely independent and dependent pronouns, demonstratives and indefinite articles, indicate the status of their denotations as more or less activated in the speaker/hearer’s mind, the discourse, or some real or possible world.¹⁴

Gundel et al. (1993) propose six cognitive (memory and attention) statuses relevant to the form of nominal expressions in discourse which are implicationally related such that each status entails (and is therefore included by) all lower statuses, but not vice versa. The statuses that an entity mentioned in a sentence may have in the mind of the addressee and their relation to each other is represented in the Givenness Hierarchy in Table 3.24:

Table 3.24: Givenness Hierarchy (Gundel et. al.1993)

in focus > activated > familiar > uniquely identifiable > referential > type identifiable

Each status on the hierarchy is a necessary and sufficient condition for the appropriate use of a different form or forms. In using a particular form, a speaker

¹⁴“Depending on where the referents or corresponding meanings of these linguistic expressions are assumed to reside” (Gundel & Fretheim 2001: 177).

signals that s/he assumes the associated cognitive status is met and, since each status entails all lower statuses, s/he also signals that all lower statuses (the statuses to the right) have been met (Gundel et al. 1993: 275). For example, anything in focus is also activated, anything activated is also familiar, and so on, but something that is familiar is not necessarily activated or in focus. The statuses are therefore ordered from most restrictive (in focus) to least restrictive (type identifiable), with respect to the set of possible referents they include. These are summarized in (11):

(11) Six cognitive statuses proposed by Gundel et al. (1993)

- *Type identifiable*. The addressee is able to access a representation of the type of object described by the expression.
- *Referential*. The addressee not only needs to access an appropriate type-representation, he must either retrieve an existing representation of the speaker's intended referent or construct a new representation by the time the sentence has been processed.
- *Uniquely identifiable*. In contrast to expressions which are referential but not uniquely identifiable, expressions which are both referential and uniquely identifiable require the addressee to construct or retrieve a representation on the basis of the nominal expression alone. Identifiability may be based on an already existing representation in the addressee's memory.
- *Familiar*. The addressee is able to uniquely identify the intended referent because he already has a representation of it in memory (in long-term memory if it has not been recently mentioned or perceived, or in short-term memory if it has).
- *Activated*. The referent is represented in current short-term memory. Activated representations may have been retrieved from long-term memory, or they may arise from the immediate linguistic or extralinguistic context. They therefore always include the speech participants themselves.
- *In focus*. The referent is not only in short-term memory, but is also at the current center of attention. Entities in focus generally include at least the topic of the preceding utterance, as well as any still-relevant higher-order topics.

The forms that encode statuses on the Givenness Hierarchy thus provide procedural information about the manner of cognitive accessibility (or accessibility of representations of the intended referent) and thereby guide the addressee in

restricting possible interpretations to ones whose status is explicitly indicated by particular forms. Furthermore, these hierarchical relations predict that a particular form will be inappropriate if the required cognitive status is not met.

Table ?? shows the correlations between pragmatic status and nominal forms in ZAI.¹⁵

Table 3.25: Correlations between linguistic form and pragmatic status in ZAI

In focus	Activated	Familiar	Uniquely identifiable	Referential	Type identifiable
= <i>bě</i>	independent		NP + DIST		<i>ti</i> NP ‘a NP’
= \emptyset	pronoun NP + DEM				\emptyset N

Zero pronouns require that the referents be “in focus” while both dependent and independent pronouns require that referents be at least familiar. Indefinite NPs, in contrast, may require only that referents be type identifiable.

The four-way distinction in demonstratives (proximal, mesioproximal, mesiodistal and distal) summarized above in (3.10) is relevant here as well. As we saw, important differences occur in the Pear Story corpus with respect to how each is used anaphorically to refer to already introduced referents. Of the 147 lexical NPs + DEM used this way, the proximal form *ri*’ is used only two times, the mesioproximal form *ca* only once and the mesiodistal form *rica*’ zero times. The distal demonstrative *que* is by far the most frequent having been employed in the remaining 144 cases. What is interesting is that the few uses of the proximate and the mesioproximal demonstratives are limited to cases in which the lexical NP refers anaphorically to a referent mentioned within the previous three clauses, i.e more familiar or more activated referents.

The above cognitive statuses generally correlate formally with type of nominal expression. As was shown, these statuses also have correlates in syntax, in particular, with the grammatical roles of core arguments. In short, the O role tends to house less activated or ‘new’ referents that are coded with more linguistic material such as Lexical NPs. The A role tends to house referents that are in focus (in

¹⁵Note that, based on further cross-linguistic investigation, Gundel et al. (2010) claim that 1) if a language encodes the distinction between two adjacent statuses on the Givenness Hierarchy, it will also encode distinctions between higher statuses, and 2) all languages encode distinctions between the two highest statuses, ‘in focus’ and ‘activated’.

the sense of Gundel et al. (1993)) and that are coded with less linguistic material such as zeros. The tendencies for the S role are found somewhere between these two poles, tending more towards the O role in the marking of new information, but more towards the A role in contexts of topic continuity, i.e. the marking of topical or human elements.

Finally, the cognitive status “in focus” has also been claimed to have prosodic correlates, i.e. phonological attenuation (Gundel et al. 1993: 285; but see also the cognitive category “activeness” in Lambrecht 1994; Ariel 1990; 2001). As mentioned in §2.2.4 and discussed in more detail in §5, such correlates do not exist in ZAI, at least in the form of pitch accent. In this, it may be important to consider that, in Lambrecht’s words:

“While it is true that the referent of a pronominal expression or of a nominal expression spoken with attenuated pronunciation is always taken to be active..., it is NOT the case that an expression coding a referent which is assumed to be active is necessarily also spoken with attenuated pronunciation. In other words, weak prosodic manifestation is only a sufficient, not a necessary condition for assumed activeness of a discourse referent” (Lambrecht (1994: 97); EMPHASIS in original).

For Lambrecht, then, the link between attenuated pronunciation and/or pronominal marking and highly activated referents represents the unmarked or default case whereas, in more “marked” environments, these same referents may receive emphatic pronunciation and be coded using fuller nominal forms.

Similarly, Ariel (2001: 50) emphasizes the role that phonetic and intonational cues might play in marking the degree of accessibility of a referent. She cites Mithun (1995) who shows how the same accessibility marker, a definite NP, can encode different degrees of accessibility through prosodic cues: low degrees of accessibility are encoded by definite NPs which occur in separate intonation units, slightly higher degrees of accessibility are encoded by definite NPs which are not separated by any intonational cues, and high degrees of accessibility are encoded by definite NPs that occur in the more given syntactic position (in Central Pomo) with a specific, unmarked intonation.

In the next chapter, I leave behind the relation between grammatical role, accessibility and pragmatic status, which I will come back to in §6, and I continue with the analysis of ZAI nominal forms by focusing on the alternation and distribution of overt and zero third-person clitics that was mentioned in §3.1.7.

3.3 Summary and conclusions

This chapter explored the relationship in ZAI between form and distribution of nominals by function, focusing on the ways that the different forms are used to introduce and track referents and to mark referents as more or less accessible. Through the lens of Preferred Argument Structure (Du Bois 2003a) and the theory of Accessibility (Ariel 2001), the chapter argued that the fundamental mechanism driving the PAS tendencies captured by PAS can be traced to the notion of accessibility.

More specifically, one of the tendencies identified by PAS, the avoidance of new referents and lexical NPs in the A role, was understood as an avoidance of referents in the A role with a low degree of accessibility. More directly, the tendency is: *Avoid low accessible As*. This is because, as we saw, highly accessible referents with less coding material are likely to occur in the A role. In contrast, low accessible referents with characteristically more coding material are unlikely to occur in that role and more consistently occur in the O role instead. The S role, for its part, exhibits a tendency in between the A and O roles. On one hand, it can house previously mentioned, animate, salient, topical, and recent referents. On the other hand, it can house new referents at episode boundaries, thereby functioning as a “cognitive staging area” (cf. §3.1.8).

In summary, the A role tends to house referents that are ‘in focus’ (Gundel et al. 1993) and coded with less linguistic material and the O role houses referents that are less activated or “new” and coded with more linguistic material. The S role tends more towards the O role in contexts of marking new information and more towards A role in contexts of topic continuity.

Furthermore, we saw that there is a relation between grammatical role of core arguments, accessibility, and cognitive or pragmatic status. In other words, cognitive status correlates with type of nominal expression as well as with the grammatical roles of core arguments. These correlations were summarized in Table ?? . That this is the case is because nominal forms indicate the status of their denotations as more or less activated in the speaker or hearer’s mind, as pragmatically more or less available, such that the forms of nominals that speakers use depend on the assumed cognitive status of the referents involved. That is, nominal forms depend on assumptions that a speaker can reasonably make regarding the addressee’s knowledge and attention state in the specific context in which the form is used.

References

- Alfabeto popular para la escritura del zapoteco del Istmo*. 1956. Mexico City, Mexico.
- Arellanes, F. 2009. *El sistema fonológico y las propiedades fonéticas del zapoteco de san pablo güilá: descripción y análisis formal*. El Colegio de México dissertation.
- Ariel, M. 1988. Referring and accessibility. *Journal of linguistics* 24(01). 65–87.
- Ariel, M. 1990. *Accessing noun-phrase antecedents*. New York: Routledge.
- Ariel, M. 2001. Accessibility theory: an overview. In T. Sanders, J. Schilperoord & W. Spooren (eds.), *Text representation: linguistic and psycholinguistic aspects*, 29–87. Philadelphia: John Benjamins.
- Arnold, J.E. 2003. Multiple constraints on reference form. In J.W. Du Bois, L.E. Kumpf & W.J. Ashby (eds.), *Preferred argument structure: grammar as architecture for function*, 225–246. Philadelphia: John Benjamins.
- Auer, P. 2005. Projection in interaction and projection in grammar. *Text* 25(1). 7–36.
- Augsburger, D. 2004. *Language socialization and shift in an Isthmus Zapotec community of Mexico*. University of Pennsylvania dissertation.
- Avelino, H. 2004. *Topics in Yalálag Zapotec, with particular reference to its phonetic structures*. University of California, Los Angeles dissertation.
- Beam de Azcona, R. 2004. *A Coatlán-Loxicha Zapotec Grammar*. University of California, Berkeley dissertation.
- Benton, J.P. 1987. Clause and sentence-level word order and discourse strategy in Chichicapan Zapotec oral narrative discourse. *SIL Mexico Workpapers* 9. 72–84.
- Benton, J.P. 1997. Aspect shift in Chichicapan Zapotec narrative discourse. *SIL Mexico Workpapers* 12. 34–46.
- Bernini, G. & M.L. Schwartz. 2006. *Pragmatic organization of discourse in the languages of Europe*. Berlin: Walter de Gruyter.
- Black, C. 2000. *Quiébolani zapotec syntax: a principles and parameters account*. Dallas: SIL International & The University of Texas at Arlington Publications in Linguistics.

- Brickell, T. & S. Schnell. 2017. Do grammatical relations reflect information status? reassessing preferred argument structure theory against discourse data from tondano. *Linguistic Typology* 21(1). 177–208.
- Broadwell, G.A. 1999. The interaction of focus and constituent order in San Dionicio Ocotepéc Zapotec. In *Proceedings of the LFG 99 conference*.
- Broadwell, G.A. 2002. Preverbal positions and phrase boundaries in Zapotec. In *Annual meeting of the linguistic society of america*. San Francisco, CA.
- Camacho, J., R. Gutiérrez-Bravo & L. Sánchez. 2010. *Information structure in indigenous languages of the americas: syntactic approaches*. Vol. 225. Walter de Gruyter.
- Campbell, E. 2011. Zenzontepec Chatino aspect morphology and Zapotecan verb classes. *International Journal of American Linguistics* 77(2). 219–246.
- Campbell, E. 2017a. Otomanguean historical linguistics: exploring the subgroups. *Language and Linguistics Compass* 11(7). e12244.
- Campbell, E. 2017b. Otomanguean historical linguistics: past, present, and prospects for the future. *Language and Linguistics Compass* 11(4). e12240.
- Castillo Hernández, C. 2014. Observaciones sobre la interfaz sintaxis-pragmática en narrativas de tres lenguas indígenas mexicanas. *Signo y Señal Revista del Instituto de Lingüística* (25). 35–58.
- Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C.N. Li (ed.), *Subject and Topic*, 27–55. New York: Academic Press.
- Chafe, W. (ed.). 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood: Ablex.
- Chafe, W. 1994. *Discourse, consciousness and time*. Chicago: University of Chicago Press.
- Chávez Peón, M. 2010. *The interaction of metrical structure, tone, and phonation types in quiaviní zapotec*. University of British Columbia dissertation.
- Crocco, C. 2009. Topic accent and prosodic structure. In Lunella Mereu (ed.), *Information structure and its interfaces*, 15–49. Berlin: Mouton de Gruyter.
- Dahlstrom, A. 1991. *Plains Cree Morphosyntax*. New York: Garland.
- Dahlstrom, A. 2003. Focus constructions in Meskwaki (Fox). In M. Butt & T. Holloway King (eds.), *The proceedings of the lfg'03 conference*, 144–163. Stanford: CSLI Publications.
- Dahlstrom, A. 2014. Obviation and Information Structure in Meskwaki. In M. Macaulay & M. Noodin (eds.), *Forty-sixth Algonquian Conference 2014*. Michigan State University Press.
- Dixon, R. 1979. Ergativity. *Language* 55(1). 59–138.

- Downing, P. 1980. Factors influencing lexical choice in narrative. In W. Chafe (ed.), *The pear stories. cognitive, cultural, and linguistic aspects of narrative production*, 89–126. Norwood: Ablex.
- Dryer, M.S. 2007. Word order. *Language typology and syntactic description* 1. 61–131.
- Du Bois, J. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.
- Du Bois, J. 2003a. Argument structure: grammar in use. In J. Du Bois, L.E. Kumpf & W.J. Ashby (eds.), *Preferred argument structure: grammar as architecture for function*, 11–60. Philadelphia: John Benjamins.
- Du Bois, J. 2003b. Discourse and grammar. In M. Tomasello (ed.), *The new psychology of language: cognitive and functional approaches to language structure*, vol. 2, 47–87. Mahwah, NJ: Lawrence Erlbaum.
- Du Bois, J. 2006. The pear story in sakapultek maya: a case study of information flow and preferred argument structure. *Haciendo Lingüística: Homenaje a Paola Bentivoglio*.
- Du Bois, J., L.E. Kumpf & W.J. Ashby (eds.). 2003. *Preferred Argument Structure: Grammar as architecture for function*. Philadelphia: John Benjamins.
- DuBois, J., S. Schuetze-Coburn, S. Cummings & D. Paolino. 1993. Outline of discourse transcription. In J.A. Edwards & M.D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*, 45–89. Hillsdale, NJ: Erlbaum.
- Enríquez Licón, M. 2008. Clases verbales en zapoteco del istmo. In A. López Cruz & M. Swanton (eds.), *Memorias del Coloquio María Teresa Fernández de Miranda*. Oaxaca, Mexico: Biblioteca Francisco de Burgoa.
- Erteschik-Shir, N. 2007. *Information structure: the syntax-discourse interface*. Oxford: Oxford University Press.
- Everett, C. 2009. A reconsideration of the motivations for preferred argument structure. *Studies in Language* 33(1). 1–24.
- Fox, J. 1977. Roman Jakobson and the comparative study of parallelism. *Roman Jakobson: Echoes of his scholarship*. 59–90.
- Givon, T. (ed.). 1983. *Topic continuity in discourse: A quantitative cross-language study*. Philadelphia: John Benjamins.
- Gundel, J.K., M. Bassene, B. Gordon, L. Humnick & A. Khalfaoui. 2010. Testing predictions of the Givenness Hierarchy framework: A crosslinguistic investigation. *Journal of Pragmatics* 42(7). 1770–1785.
- Gundel, J.K. & T. Fretheim. 2001. Topic and focus. In L. Horn & G. Ward (eds.), *Handbook of pragmatics*, 175–196. Oxford: Blackwell.
- Gundel, J.K., N. Hedberg & R. Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

- Haig, G. & S. Schnell. 2016. The discourse basis of ergativity revisited. *Language* 92(3). 591–618.
- Haiman, J. 1978. Conditionals are topics. *Language* 54(3). 564–589.
- Haspelmath, M. 2006. Preferred Argument Structure: Grammar as architecture for function (review). *Language* 83(4). 908–912.
- Heath, S.B. 1972. *Telling tongues: Language policy in Mexico, colony to nation*. New York: Teachers College Press.
- Heise, J.L. 2003. *Participant reference and tracking in San Francisco Ozolotepec Zapotec*. Graduate Institute of Applied Linguistics. MA thesis.
- Hopper, P.J. & S.A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56(2). 251–299.
- Jakobson, R. 1966. Grammatical parallelism and its Russian facet. *Language* 42(2). 399–429.
- Kaufman, T., M. Feke & G. Pérez Báez. n.d. Lexical Database of Sapoteko of Juchitán. Project for the Documentation of the Languages of Meso-America.
- Kerslake, C. 1996. Future time reference in subordinate clauses in Turkish. In *Proceedings of the VIIIth International Conference on Turkish Linguistics*, 49–59.
- König, E. & P. Siemund. 2007. Speech act distinctions in grammar. In E. Shopen (ed.), *Language typology and syntactic description*, vol. 1, 276–324. Cambridge: Cambridge University Press.
- Konnerth, L. 2013. Additive particle and discourse contrast marker: Evidence from Karbi (Tibeto-Burman) =*ta*. In *Information Structure in Spoken Language Corpora*. University of Bielefeld, Germany.
- Kreikebaum, W. 1987. Fronting and related features in Santo Domingo Albarradas Zapotec. *SIL Mexico Workpapers* 9. 33–71.
- Lambrecht, K. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Lee, F. 2000. VP remnant movement and VSO in Quiavini Zapotec. In A. Carnie & E. Guilfoyle (eds.), *The syntax of verb-initial languages*, 143–162. New York: Oxford University Press.
- Lewis, M.P., G.F. Simons & C.D. Fennig (eds.). 2016. *Ethnologue: languages of the world*. 19th. Dallas: SIL International.
- Li, C.N. & S.A. Thompson. 1976. Subject and Topic: a new typology of language. In C.N. Li (ed.), *Subject and Topic*, 459–489. New York: Academic Press.
- Lillehaugen, B.D. 2006. *Expressing Location in Tlacolula Valley Zapotec*. University of California, Los Angeles dissertation.

- Lillehaugen, B.D. 2008. El morfema làa' y su uso en el zapoteco de tlacolula de matamoros. In *X encuentro internacional de lingüística en el noroeste*. Hermosillo, Sonora, Mexico.
- Lillehaugen, B.D. 2016. The syntax of preverbal subjects in colonial valley zapotec. In *Vii syntax of the world's languages international conference and workshop*. Mexico City, Mexico.
- Long, R. 1985. Topicalization in Zoogocho Zapotec expository discourse. *SIL Mexico Workpapers* 7. 61–100.
- MacLaury, R.E. 1989. Zapotec body-part locatives: Prototypes and metaphoric extensions. *International Journal of American Linguistics* 55(2). 119–154.
- Marchese, L. 1977. Subordinate clauses as topics in godié. *Studies in African Linguistics, Supplement* 7. 157–164.
- Marchese, L. 1987. On the role of conditionals in godié procedural discourse. In *Coherence and grounding in discourse: outcome of a symposium, eugene, oregon, june 1984*, vol. 11.
- Marlett, S. & V. Pickett. 1987. The syllable structure and aspect morphology of Isthmus Zapotec. *International Journal of American Linguistics* 53(4). 398–422.
- Marlett, S. & V. Pickett. 1996. El pronombre inaudible en el zapoteco del Istmo. *III Encuentro de Lingüística en el Noroeste*. 119–150.
- Matić, D. & D. Wedgwood. 2013. The meanings of focus: the significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49(1). 127–163.
- McComsey, M. 2015. *Bilingual spaces: approaches to linguistic relativity in bilingual mexico*. University of California, San Diego dissertation.
- Mereu, L. (ed.). 2009. *Information structure and its interfaces*. Berlin: Mouton de Gruyter.
- Mithun, M. 1992. Is basic word order universal. In D.L. Payne (ed.), *Pragmatics of word order flexibility*, 15–61. Philadelphia: John Benjamins.
- Mithun, M. 1995. Morphological and prosodic forces shaping word order. In P. Downing & M. Noonan (eds.), *Word order in discourse*, 387–423. Philadelphia: John Benjamins.
- Mock, C. 1983. Tone Sandhi in Isthmus Zapotec: an Autosegmental Account. *Linguistic Analysis* 12(1). 91–139.
- Mock, C. 1985a. A systemic phonology of Isthmus Zapotec prosodies. In J.D. Benson & W.S. Greaves (eds.), *Systemic perspectives on discourse: selected theoretical papers from the ninth international systemic workshop*, 349–372. Norwood: Ablex.

- Mock, C. 1985b. Relations between pitch accent and stress. In *Papers from the general session at the twenty-first regional meeting of the Chicago Linguistic Society*, 256–274.
- Mock, C. 1988. Pitch accent and stress in Isthmus Zapotec. In H. Van der Hulst (ed.), *Autosegmental studies on pitch accent*, 197–223. Dordrecht: Foris.
- Mock, C. 1990. Temporal orientation without tenses: the deixis of time in Isthmus Zapotec. *Homenaje a Jorge A. Suárez: lingüística indoeuropea e hispánica*. 367–379.
- Opengin, Ergin. 2013. Topicalisation in central Kurdish: additive particle and other means. In *Information Structure in Spoken Language Corpora*. University of Bielefeld, Germany.
- Ozerov, Pavel. 2015. Information structure without topic and focus: differential object marking in Burmese. *Studies in Language* 39(2). 386–423.
- Payne, D.L. 1990. *The pragmatics of word order: typological dimensions of verb initial languages*. Berlin: Walter de Gruyter.
- Payne, D.L. 1995. Verb initial languages and information order. In P. Downing & M. Noonan (eds.), *Word order in discourse*, 449–85. Philadelphia: John Benjamins.
- Pekarek Doehler, Simona. 2011. Emergent grammar for all practical purposes: the on-line formatting of left and right dislocations in French conversation. In *Constructions: emerging and emergent*, 45–87. Berlin: Walter de Gruyter.
- Pérez Báez, G. 2011. Spatial frames of reference preferences in Juchitán Zapotec. *Language Sciences* 33(6). 943–960.
- Pérez Báez, G. 2015. Morphological valence-changing processes in Juchitán Zapotec. *Valence Changes in Zapotec: Synchrony, diachrony, typology* 110. 93.
- Persons, D. 1979. Plot structure in Lachixio Zapotec discourse. In L.K. Jones & R.E. Longacre (eds.), *Discourse studies in Mesoamerican languages*, vol. 1, 123–40. Dallas: Summer Institute of Linguistics.
- Pickett, V. 1960. The grammatical hierarchy of Isthmus Zapotec. *Language* 36(1). 3–101.
- Pickett, V. 1979. *Vocabulario zapoteco del Istmo*. Mexico, DF: Summer Institute of Linguistics.
- Pickett, V., C. Black & V.M. Cerqueda. 1998. *Gramática popular del zapoteco del Istmo*. Oaxaca, Mexico: Centro de Investigación y Desarrollo Binnizá; Instituto Lingüístico de Verano.
- Piper, M.J. 1995. The functions of 'lëë' in Xanica Zapotec narrative discourse with some implications for comparative Zapotec. *SIL Mexico Workpapers* 11. 67–78.

- Prince, E. 1981. Toward a taxonomy of given-new information. In P. Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Rendón, J.J. 1995. *Diversificación de las lenguas zapotecas*. Oaxaca, Mexico: Instituto Oaxaqueño de las Culturas, Centro de Investigaciones y Estudios Superiores de Antropología Social.
- Riggs, D.B. 1987. Paragraph analysis for Amatlán Zapotec. *SIL Mexico Workpapers* 9. 1–11.
- Riggs, D.B. & S. Marlett. 2010. The le'e focus phrase: structural aspects. In Black H.A. Black C. & S. Marlett (eds.), *The zapotec grammar files*. Summer Institute of Linguistics.
- Sacks, H., E.A. Schegloff & G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4). 696–735.
- Saynes-Vásquez, F.E. 2002. *Zapotec language shift and reversal in Juchitán, Mexico*. University of Arizona dissertation.
- Scarano, Antonietta. 2009. A the prosodic annotation of c-oral-rom and the structure of information in spoken language. *Information structures and its interfaces*. Berlin and New York: Mouton de Gruyter. 51–74.
- Sherzer, J. 1987. A discourse-centered approach to language and culture. *American Anthropologist* 89(2). 295–309.
- Sicoli, M.A. 2007. *Tono: A linguistic ethnography of tone and voice in a Zapotec region (Mexico)*. University of Michigan dissertation.
- Sicoli, M.A. 2010. Shifting voices with participant roles: voice qualities and speech registers in Mesoamerica. *Language in Society* 39(04). 521–553.
- Silverstein, M. 1976. Hierarchy of features and ergativity. In R.M.W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Silverstein, M. 1984. On the pragmatic 'poetry' of prose: parallelism, repetition, and cohesive structure in the time course of dyadic conversation. *Meaning, form, and use in context: Linguistic applications*. 181–199.
- Smith-Stark, T.C. 2002. Las clases verbales del zapoteco de Chichicapan. In Z. Estrada Fernández, M. Figueroa Esteva & G. López Cruz (eds.), *Proceedings of the VI Encuentro Internacional de Lingüística en el Noroeste*. Sonora, Mexico: Editorial Universidad de Sonora.
- Smith-Stark, T.C. 2003. Algunas isoglosas zapotecas. In *Proceedings of the III Coloquio de Mauricio Swadesh*. Mexico City, Mexico.
- Sonnenschein, A.H. 2005. *A descriptive grammar of San Bartolomé Zoogocho Zapotec*. Munich: Lincom Europa.

- Suárez, J.A. 1983. *The Mesoamerican Indian Languages*. Cambridge: Cambridge University Press.
- Swadesh, M. 1947. The phonemic structure of proto-zapotec. *International Journal of American Linguistics* 13(4). 220–230.
- Thompson, S.A., R.E. Longacre & S.J. Hwang. 2007. Adverbial clauses. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 2, 237–300. Cambridge: Cambridge University Press.
- Toledo Bustamante, N. 2018. *Socialization patterns, emerging attention, and language choice in juchitán, oaxaca, mexico*. University of Chicago dissertation.
- Van Valin, R.D. 1999. A typology of the interaction of focus structure and syntax. In E. Raxilina & J. Testelec (eds.), *Typology and the theory of language: from description to explanation*, 511–524. Moscow: Languages of Russian Culture.
- Ward, M. 1987. A focus particle in quioquitani zapotec. *SIL Mexico Workpapers* 9. 26–32.
- Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.
- Zhang, J. 2002. *The effects of duration and sonority on contour tone distribution: a typological survey and formal analysis*. New York: Routledge.
- Zoll, C. 2003. Optimal tone mapping. *Linguistic Inquiry* 34(2). 225–268.

