

Pre-print version, November 2008.

Accepted for publication in the *International Journal of American Linguistics*.

Word order in Paraguayan Guaraní

Judith Tonhauser and Erika Colijn

The Ohio State University

Abstract

This paper explores constraints on clausal word order in Paraguayan Guaraní on the basis of a corpus of naturally occurring data. We find that grammatical function is a strong predictor of word order for direct objects since they are almost exclusively realized post-verbally, as well as for indirect objects, which all occur post-verbally. The placement of subjects, on the other hand, depends on a variety of factors, in particular discourse status and the transitivity of the verb. We find only partial support for the claim that Paraguayan Guaraní has a basic subject-verb-object word order (e.g. Gregores and Suárez 1967) since there does not seem to be a basic position for subjects. We conclude by comparing constraints on word order in Paraguayan Guaraní to constraints reported for other Tupí-Guaraní languages.

Keywords: Paraguayan Guaraní (Tupí-Guaraní), constraints on word order (grammatical function, animacy, discourse status, transitivity), basic word order.

Contents

1	Introduction	1
2	Annotation of the Guaraní corpus	3
2.1	Transitivity of the verb	5
2.2	Grammatical function	5
2.3	Animacy	6
2.4	Discourse status	8
2.5	Word order	12
2.6	Coding decisions	13
3	Constraints on word order: results of the quantitative study	15
3.1	Placement of subject noun phrases	18
3.2	Placement of object noun phrases	23
3.3	Summary	24
4	Does Guaraní have a basic word order?	25
4.1	Frequency	26
4.2	Markedness	26
4.3	Disambiguation	27
5	Word order in other Tupí-Guaraní languages	30

Word order in Paraguayan Guaraní

1 Introduction

This paper explores constraints on word order in Paraguayan Guaraní¹ (henceforth Guaraní) on the basis of a corpus of naturally occurring data. Whereas the order of affixes, clitics, and stems in Guaraní is rather well-described (cf. Gregores and Suárez 1967, Velázquez-Castillo 2002a,b), only few claims have been made about word order at the clausal level. Gregores and Suárez (1967), a reference grammar of the language, suggests that subject–verb–object is the most frequent and the basic word order, but cautions that these claims are “very rough approximations, based on impressionistic evaluations of what is more frequent” (p.182). Velázquez-Castillo (1996:11) states that Guaraní “word order is often determined by discourse” (cf. also Tonhauser 2006:144f.) and, in contrast to Gregores and Suárez (1967), maintains that “the default order of elements seems to be VO [verb–object], with the subject appearing either before or after the verb”.² While Velázquez-Castillo (1996) does not indicate whether these results are based on a corpus study, Velázquez-Castillo (1995) examines the placement of objects in nine folktales using Givón’s (1983) measures for topicality, and finds that the placement of object noun phrases is affected by topicality: post-verbally realized objects tend to be more topical than pre-verbal objects. In this paper, we systematically explore constraints on clausal word order on the basis of a corpus of naturally occurring Guaraní discourse. The central advantage for using a corpus rather than relying on elicited examples is that corpus examples come with a discourse context, which allows one to examine various factors that have been reported in the literature to affect word order (such as discourse status, animacy and grammatical function), as well as their

¹Paraguayan Guaraní is a Tupí-Guaraní language spoken by about four million people in Paraguay and surrounding countries. The data presented in this paper was collected by the first author during several fieldwork trips between 2004 and 2007. Examples are marked to identify their origin: elicited examples with [E], examples from a corpus of naturally occurring texts with [C], examples from a theater play not part of the corpus with [T], and examples from published texts with the name and page of the publication. For providing and discussing the Guaraní texts we thank Maria de la Cruz Bogado, Nicolas Cantero, Ancia Sabina Maciel and Marité Maldonado. We also thank Maura Velázquez-Castillo and two anonymous reviewers for IJAL for their detailed comments, which led to significant improvements in the paper.

²Our paper is concerned with constraints on word order in modern (Paraguayan) Guaraní. Velázquez-Castillo (1995) provides evidence that 17th-century Guaraní had an OV basic word order (cf. also Ruiz de Montoya 1993:34-35). We thank an anonymous reviewer for pointing out that present-day tribal Tupí-Guaraní languages tend to have an OV order (as reported e.g. for Chiriguano in Dietrich (1986) and for Tapiete in González (2005)) and the suggestion that the VO word order of modern Paraguayan Guaraní is a result of its extensive contact with Spanish (cf. also Martins 2003:163).

interaction. We believe that elicitation is a useful, even necessary, fieldwork methodology for investigating other areas of grammar (cf. Matthewson 2004), but it would be difficult to identify constraints on Guaraní word order via elicitation: consultants will, in (1) for example, accept any of the six possible orders of the subject (S) *Juan*, the object (O) *Maria-pe* (*-pe* functions as an object marker here) and the (third person inflected) verb (V) *o-hecha* ‘s/he sees/saw’ as an (out of context) translation of the Spanish sentence *Juan vió a Maria* (‘Juan sees/saw Maria’):³

(1) *Juan o-hecha Maria-pe.*

Juan A3-see Maria-PE

‘Juan sees/saw Maria.’

[E]

To identify constraints on word order, several consultants would need to be presented with a great number of contextualized examples to identify which (interplay of) factors leads them to prefer one (or more) of the six possible word orders over the others. By using a corpus, we furthermore ensure that our study is representative of the actual language, not of data obtained via elicitation. We find that grammatical function, animacy, discourse status and transitivity (of the verb) play a role in the placement of noun phrases in Guaraní, although to different degrees. Grammatical function emerges as a strong predictor of noun phrase placement since almost all (94%) direct object and all indirect object noun phrases occur post-verbally. Several criteria for determining basic word order support the claim (cf. Gregores and Suárez 1967, Velázquez-Castillo 1995, 1996) that VO is a basic word order of Guaraní. This finding is typologically surprising, as already pointed out by Velázquez-Castillo (1995:568), since languages with postpositions tend to have an OV basic order (Greenberg 1963, Dryer 1992, Haspelmath et al. 2005). The placement of subjects is more complex and depends on a variety of factors, in particular the discourse status of the denotation of the noun phrase and the transitivity of the verb. Given the flexibility of the placement of subjects, we conclude that SVO is not the basic word order of Guaraní.

³The examples in this paper are given in the standardized orthography of Guaraní used in Paraguay (Ministerio de Educación y Cultura 2004, cf. also Velázquez-Castillo 2004b:1421f.), except that all postpositions are attached to their host. Following the official orthography, accents are not written for normally accented words (stress on the final syllable); stressed nasal syllables are marked with a tilde. The following glosses are used (cf. footnote 5 for glosses of cross-reference markers): 3.pron = 3rd person pronoun, -ABL = ablative case, -AT = cotemporaneity, CAUS- = causative, -COMPLETE = completive aspect, -DIM = diminutive, -EMPH = emphatic, excl = exclusive, -FUT = future, incl = inclusive, JE- = reflexive/passive, NEG = negation, -NOM = nominalizer, -NOM.FUT = future-oriented nominal aspect/mood, -NOM.PAST = past-oriented nominal aspect (cf. the analysis of the nominal markers in Tonhauser (2006, 2007b)), -PE = case marker, -PL = plural, PROG = progressive, -RC = relative clause, -REHE = case marker.

2 Annotation of the Guaraní corpus

Paraguayan Guaraní is a comparatively well-described language among the South American languages, and we refer to Gregores and Suárez (1967) and Maura Velázquez-Castillo's (1996, 2002a, 2002b, 2004a, 2004b) work for descriptions and analyses of the morphosyntax of the language. In this section, we describe only those features of Guaraní that are relevant for the study of word order and the annotation of the corpus, in particular the three main factors that have been identified in the literature as affecting word order across languages, namely grammatical function, animacy and discourse status (e.g. Brody 1984, Mithun 1992, Rude 1992, Choi 1999 and references therein).⁴

Cross-reference prefixes mark arguments of verbs as well as possessors. The example in (2) illustrates the third person (set A) cross-reference prefix *o-*, which identifies the subject of the transitive verb *mby-aku* (CAUS-hot) 'heat' as a third person entity.⁵

(2) Context: Dominga entered.

O-*mby-akú-ta hína y ka'ay-rã.*
A3-CAUS-hot-FUT PROG water ka'ay-NOM.FUT

'She was getting ready to heat water for ka'ay (= herbal hot drink).' (adapted from Velázquez-Castillo 2002b:521)

In Guaraní, transitive verbs are realized with only one cross-reference prefix, which cross-references the argument that is higher on the person hierarchy ($1 < 2 < 3$, cf. Velázquez-Castillo 1996:16). Subjects of transitive verbs are cross-referenced with set A prefixes, objects with set B prefixes (cf. footnote 5); the single argument of an intransitive verb is cross-referenced with a set A or a set B prefix (see Velázquez-Castillo (2002a) for details). Arguments need not be realized with a full noun phrase (or an independent pronoun): in (2), for example, the agent of the heating event is referred to only with a cross-reference marker. In (3), the agent of the transitive verbs *trata* 'treat' (a Spanish borrowing) and *nupã* 'hit' is not overtly realized at all: since the patient of both verbs is the (1st person) speaker (and hence higher on the person hierarchy than the

⁴Other factors that have been identified as playing a role in word order are the length of the expression (Arnold et al. 2000) and performance factors (Hawkins 1994). These are not addressed here and remain a topic for future research.

⁵The set A prefixes are *a(i)-* 'A1sg', *ja(i)-* 'A1pl.incl', *ro(i)-* 'A1pl.excl', *re(i)-* 'A2sg', *pe(i)-* 'A2pl', and *o(i)-* 'A3'; the set B prefixes are *che(r)-* 'B1sg', *ñane(r)-* 'B1pl.incl', *ore(r)-* 'B1pl.excl', *nde(r)-* 'B2sg', *pende(r)-* 'B2pl', and *i-/h-* 'B3'. The two portmanteaux prefixes *ro(i)-* '12sg' and *po(i)-* '12pl' refer to a first person subject and a second person (singular/plural) object.

contextually salient third person agent), the speaker is cross-referenced on the verbs with the first person set B marker *che-*. The agent is neither realized with a cross-reference marker nor with an overt noun phrase.

- (3) *Jaguá-icha che-trata. Por poco na-che-nupã-i.*
 dog-like B1sg-treat barely NEG-B1sg-hit-NEG
 ‘He treated me like a dog. He almost hit me.’ [T]

The Guaraní corpus used for this study consists of eight texts, amounting to about 2800 Guaraní words (which correspond to about 6,000 English words since Guaraní is mildly polysynthetic). Table 1 gives an overview of the texts in the corpus.⁶

Text	description	word count
<i>Michi</i> ‘small’	personal narrative about childhood told by MM	283
<i>Erída</i> ‘wound’	personal narrative about an accident told by MB	1061
<i>BDF-1</i>	<i>A boy, a dog and a frog</i> (Mayer 1967) narrated by SC	412
<i>BDF-2</i>	<i>A boy, a dog and a frog</i> (Mayer 1967) narrated by NC	247
<i>Jakare</i> ‘crocodile’	story about the life of crocodiles (author unknown)	143
<i>Kirikiri</i> ‘cricket’	fable about a cricket’s adventure (author unknown)	196
<i>Ka’i</i> ‘monkey’	fable about a monkey’s adventure (author unknown)	375
<i>Ypei</i> ‘duck’	fable about a frog and a duck (author unknown)	82
		2799

Table 1: Texts in the Guaraní corpus

We annotated the main, declarative clauses⁷ of the corpus to reflect the following five features: (i) the valency of the verb, (ii) the grammatical function of overt noun phrases, (iii) the animacy of the denotation of overt noun phrases, (iv) the discourse status of the denotation of overt noun phrases, and (v) word order. The following five sections illustrate the annotated features.

⁶The four texts with unknown authors were collected from Paraguayan school books and were checked with Guaraní speakers to verify their naturalness.

⁷We follow other word order studies in restricting our attention to main, declarative clauses since subordinate clauses have a tendency to exhibit a more rigid word order and non-declarative clauses have word order requirements not found in declarative clauses (cf. Steele 1978:592).

2.1 Transitivity of the verb

We distinguish intransitive verbs from verbs of a higher transitivity (i.e. transitive and ditransitive verbs). The verb *o-po* ‘A3-jump’ in (4a) is annotated as an intransitive verb (since it only requires one argument), in contrast to the transitive verb *ho’u* ‘A3.eat’ in (4b) which takes two arguments.⁸

- (4) a. *Ha upé-icha, ko’ã mokõi mymba-mi o-po vy’á-pe y-no’õ-me.*
 and this-like those two animal-DIM A3-jump happy-PE water-united-PE
 ‘And this is why these two animals jump happily together in the pond.’ [C]
- b. *Kyju kiri-kiri ho’u-pa avei upe hogue, ha oi-ke i-kuára-pe,*
 cricket (cricket sound) A3.drink-COMLETE also that B3.leaf and A3-enter B3-hole-PE
o-vy’a-há-pe.
 A3-happy-NOM-PE
 ‘The cricket also finished its leaf and happily went to its hole.’ [C]

2.2 Grammatical function

Grammatical function is one of the factors that has been identified in the literature as playing a role in word order. In English, grammatical function has a strong effect on word order since, typically, subjects precede the verb and objects follow it (e.g. *I know Jim*). Our annotation of the Guaraní corpus distinguishes subjects, direct objects and indirect objects, and also marks subjects for whether they are subjects of transitive or intransitive verbs. We include in our study only those verbal dependents that (i) are arguments, i.e. subcategorized for by the verb of the clause, to the exclusion of adjuncts, and (ii) denote (human, animate or inanimate) entities, to the exclusion of proposition-denoting, spatial or temporal arguments. The example in (5a) illustrates the subject noun phrase *umi karia’y* ‘those young men’ and the object noun phrase *h-embí’u ha i-sapatu-kuéra* ‘their food and shoes’ of the transitive verb *nohẽ* ‘remove’. The noun phrase *hymba jaguá-pe* ‘his dog’ in (5b) is an example of an indirect object.⁹

⁸Transitive and ditransitive verbs are not distinguished by the annotation of the verbs but by the kind of objects they take: only ditransitive verbs occur with indirect objects, which were annotated differently than direct objects.

⁹Expressions that fulfill conditions (i) and (ii), and are realized with markers such as *-pe* and *-re(he)* were also included in the study. These markers indicate the grammatical or semantic relation of the dependent to the verb: In (5b), *-pe* identifies the noun phrase *hymba jagua* ‘his dog’ as the indirect object argument of the verb *he’i* ‘A3.say’, and in (1) and (25b,c) *-pe* functions as a direct object marker. In (i) below, the verb *ñangareko* ‘protect’ subcategorizes for an object marked with *-re(he)* (here, *hupi’a-kuéra* ‘her eggs’); cf. also (28b). In addition to these more abstract, grammatical uses, such markers also express causal, spatial and temporal meanings (e.g. Gregores and Suárez 1967:163f.): for example, in (iia) *-pe* translates as *on*, and *-re* in (iib) as *against*.

- (5) a. *O-nohẽ sapy'a umi karia'y h-embí'u ha i-sapatu-kuéra.*
 A3-remove suddenly those young.man B3-food and B3-shoe-PL
 'Finally, the young men gathered their food and shoes. [C]
- b. *Peteĩ jey he'i hymba jaguá-pe: "Ja-ha ja-heka*
 one time A3.say B3.animal dog-PE A1pl.incl-go A1pl.incl-search
ñane-iru-rã".
 B1pl.incl-friend-NOM.FUT
 'One time, he said to his dog: Let's go and look for a friend.' [C]

2.3 Animacy

The grammatical category *animacy* has been shown to play a role in word order e.g. in the Mayan language Tojolabal (Brody 1984) and in Haida (Na Dene, Enrico 1986). With expressions that denote individuals, animacy refers to the level of sentience of the individual. The animacy hierarchy, in the form given in (6), ranks humans higher than non-human animates, which in turn are higher than inanimate entities.

- (6) human > (non-human) animate > inanimate (adapted from Silverstein 1976:122)

Brody (1984) reports that in Tojolabal all of the six logically possible word orders are attested, but SOV, VSO, OSV, and OVS are more marked than VOS and SVO, and are acceptable only if the subject is higher than the object on the modified animacy hierarchy in (7), which additionally distinguishes humans referred

-
- (i) *Jakare kuña o-ñangareko hupi'a-kuéra-rehe mombyry-mí-guive.*
 crocodile female A3-protect B3.egg-PL-REHE far-DIM-from
 'The female crocodile takes care of her eggs from a little way away.' [C]
- (ii) a. *A-ha a-ha'arõ chu-pe la tapé-pe.*
 A1sg-go A1sg-wait 3-PE the path-PE
 'I went to wait for her on the path.' [C]
- b. *O-mbo-jeko iñ-akã yvyrá-re.*
 A3-CAUS-lean B3-head tree-REHE
 'He leaned his head against a tree.' [C]

Since, cross-linguistically, cases and adpositions share distributional, functional and semantic similarities (e.g. Zwicky 1992, Dryer 2005), the morphological status of a marker is often the only recourse to deciding whether it is a case (affix) or an adposition (word) (Haspelmath 2006). For the Guaraní markers, this question is not yet settled: Gregores and Suárez (1967) and Velázquez-Castillo (2004a) consider *-pe* and *-re(he)* postpositions while Adelaar (1994) provides evidence that they are suffixes (cf. also Nordhoff 2004). In this paper, we follow these latter authors in assuming they are suffixes and hence include noun phrases marked with such (case) suffixes in our study.

to by proper names and independent pronouns from other humans.

(7) Animacy hierarchy used in Brody (1984:720)

proper names and independent pronouns > other humans > animate > inanimate

For instance, (8a), which realizes an OVS word order, is acceptable because the object, *chenek* ‘bean’, is lower on the animacy hierarchy than the subject *ja migel-i* ‘Miguel’. The SVO example in (8b), where the subject and object are on an equal level on the animacy hierarchy, is acceptable since the SVO word order is not constrained by the animacy hierarchy.¹⁰

(8) Tojolabal (Brody 1984:722)

- a. *miyuk chenek' wa s-chon-o-∅ ja migel-i*
NO BEAN pro 3e-SELL-tvm-3a det MIGUEL-term
‘No, Miguel sells BEANS.’
- b. *ja Roberto-i ∅-s-mak'-a-∅ ja julio-i*
det ROBERTO-term com-3e-HIT-tvm-3a det JULIO-term
‘Roberto, he hit Julio.’

Besides animacy, discourse status has an effect on Tojolabal word order (Brody 1984, cf. also for other Mayan languages, e.g. Dayley 1981, Aissen 1992, Tonhauser 2007a). In (8a), for instance, the pre-verbal object is licensed in a context where the object is contrastively focused.

To examine whether animacy plays a role in noun phrase placement in Guaraní, each overt noun phrase argument of the verb of a main, declarative clause was annotated for the animacy of its referent. The annotation distinguishes the three levels of animacy illustrated in the examples in (9a-c), respectively: human referents, (non-human) animate referents, and inanimate referents.¹¹

- (9) a. **I-memby** *o-heja*.
B3-child A3-leave
‘She left her child.’

[C]

¹⁰Brody (1984) uses the following glosses in these examples: 3 = third person, a = absolutive, com = completive aspect, det = determiner, e = ergative, pro = progressive aspect, term = terminal, tvn = transitive verb marker.

¹¹Since our corpus consists of many fables, where animals act as if they had human properties, we decided (after the corpus was coded and evaluated) to lump together human and animate referents in the discussion of the results (section 3).

- b. **Kururu** *nd-o-vy'á-i-ete*.
 toad NEG-A3-happy-NEG-very
 'The toad was very unhappy.' [C]
- c. **Kuarahy** *o-mimbi*.
 sun A3-shine
 'The sun was shining.' [C]

2.4 Discourse status

Word order is also affected by the discourse status of the denotation of a noun phrase. English has a topicalization construction that can realize e.g. topical objects in a pre-clausal position. In so-called scrambling languages, such as German or Korean (Choi 1999), deviations from the basic word order (which is determined by grammatical functions) are motivated in particular discourse contexts. Discourse configurational languages, such as Russian (King 1995), Hungarian (Kiss 2002), and the Mayan languages (e.g. Brody 1984, Aissen 1992) have special positions in the sentence for noun phrases whose denotation has a particular discourse status.

We examine the effect of discourse status on noun phrase placement in Guaraní in two ways. The first is Givón's (1983) topicality measure, which allows one to compare the relative topicality of particular noun phrase types and was already employed for Guaraní objects in Velázquez-Castillo (1995). We did not annotate all noun phrases in the corpus for this measure but only those subjects that emerged as of interest in section 3 (where we explain Givón's measure in more detail). The second way relies on the finding that, cross-linguistically, the form of a noun phrase correlates to some extent with the discourse status of its denotation (cf. e.g. Chafe 1976, Prince 1992, Gundel et al. 1993). For example, English noun phrases with the article *a* are used to denote entities that have not yet been introduced to the discourse, while noun phrases with the article *the* can be felicitously used only in contexts where a unique entity that bears the property of the noun is salient in the context (Heim 1982, Roberts 2003). Our annotation distinguishes eight kinds of noun phrase forms, which are described in the following sections.

(i) *peteĩ*-marked noun phrases The numeral *peteĩ* 'one' is used in Guaraní as an indefinite determiner. Noun phrases marked with *peteĩ* 'one' denote entities that are *hearer-new* (Prince 1992), i.e. entities that are novel to the hearer. We assume, with other authors, that hearer-new entities are also *discourse-new*, i.e. have

not yet been introduced in the discourse. An example of a hearer-new (and hence discourse-new) entity is the stone in the example in (10):

(10) Context: The frog was scared of the boy and looked for a place to hide.

O-topa **peteĩ ita.**
A3-find one stone

‘He found a stone.’

[C]

Not all hearer-new entities are introduced with noun phrases marked with *peteĩ* ‘one’: bare noun phrases, introduced next, can also be used to this effect.

(ii) Hearer-new bare noun phrases We call those Guaraní noun phrases *bare noun phrases* that consist of a noun (or a compound noun, and possibly adjectival or relative clause modifiers) but not a pre-nominal numeral, a demonstrative or a possessive prefix. Since Guaraní bare noun phrases can denote entities that are hearer-new as well as ones that are not (these are discussed in (iii) below), our annotation distinguishes two kinds of bare noun phrases. The bold-faced expressions in (11) are examples of *hearer-new bare noun phrases*:

(11) a. *Ro-moĩ* **ñuhã** *ho’á-pype* *la guyra.*
A1pl.excl-put trap A3.fall-inside the bird

‘We laid traps for the birds (to fall into).’

[C]

b. *A-japó-ta* **ta’anga araity kakuaa porã-va.**
A1sg-make-FUT figure wax big pretty-RC

‘I will make a pretty big wax figure.’

[C]

The denotation of the bare noun phrases *ñuhã* ‘trap’ in (11a) and of *ta’anga araity kakuaa porã-va* ‘pretty big wax figure’ in (11b) cannot be assumed to be familiar to the hearer (i.e. are hearer-new) and were also not mentioned in previous discourse (i.e. are discourse-new).

Guaraní has a variety of noun phrase forms that denote *discourse-old* entities, i.e. entities that have already been explicitly introduced in previous discourse, or *hearer-old* entities, which are entities that can be assumed to be familiar to the hearer either because they are inferable from the situation in which the discourse takes place, such as the sun, or because they are visible to the hearer (but have not yet been

mentioned), such as a dog that comes running towards the hearer. Discourse-old entities are also hearer-old, but not vice versa. Our annotation distinguishes six different noun phrase forms denoting discourse-/hearer-old entities; we use Gundel et al.'s (1993) *Givenness Hierarchy* to further distinguish their respective discourse statuses.

(iii) Discourse-/hearer-old bare noun phrases The vast majority of bare noun phrases in Guaraní refer to discourse- or hearer-old entities. The denotation of the bare noun phrase *kururu* ‘toad’ in (12a) is discourse-old since the toad has been previously introduced in the story and has been mentioned several times (as one of the main characters). The denotation of the bare noun *kuarahy* ‘sun’ in (12b), on the other hand, is hearer-old since the existence of the sun can be assumed to be known by the hearer (but the sun had not been explicitly introduced in the story).

- (12) a. Context: The frog has been commiserating with the duck about his life. The duck convinces him that he does many good things.

Kururu *o-po o-po vy'á-gui.*
toad A3-jump A3-jump happy-ABL

‘The toad jumped and jumped with joy.’ [C]

- b. Context: From the beginning of a story; a cricket is sitting in a meadow.

Ha kuarahy *o-ĩ-ha-gué-re-ma yvate, o-mo-ñepyrũ i-purahéi.*
and sun A3-be-NOM-NOM.PAST-REHE-already high A3-CAUS-begin B3-song

‘And since the sun was already high, it made him start his song.’ [C]

(iv) *la*-marked noun phrases Guaraní has borrowed from Spanish the definite article *la*, cf. *la ju'i* ‘the frog’ in (13).¹² Just like discourse-/hearer-old bare noun phrases, noun phrases marked with *la* denote discourse-/hearer-old entities and are *uniquely identifiable*, i.e. the nominal alone suffices for the hearer to identify the speaker’s intended referent (Gundel et al. 1993:277).

- (13) Context: The boy tried to catch the frog and the frog got scared.

O-kañy chu-gui la ju'i.
A3-hide 3-ABL the frog

‘The frog hid from him.’ [C]

¹²In our corpus, *la* is used for both singular and plural reference (e.g. (13) and (11a)), while Velázquez-Castillo (2004a) reports that *la* is used for singular reference and *lo* for plural reference.

(v) **Independent pronouns** Independent pronouns also denote entities that are discourse-/hearer-old but, in contrast to discourse-/hearer-old bare noun phrases, the referent of a (third person) pronoun has to be *in focus* (Gundel et al. 1993), i.e. the current center of attention, as well as represented in short term memory (*activated*). Independent pronouns have a variety of uses. One is to mark contrastive focus, as in (14a) where the speaker (the little boy) is contrasting himself with the dog. Pronouns are also used to (re-)instate an already introduced discourse participant as the topic. An example of this is (14b), where the third person pronoun *ha'e* makes the grandmother the topic of the current sentence (and subsequent discourse).

- (14) a. Context: The little boy says to his dog: “YOU will go to him from over there,...

ha che a-há-ta chu-pe ko ótro ládo.
and B1sg A1sg-go-FUT 3-PE this other side

‘and, me, I’ll go to him by the other side.’ [C]

- b. Context: When we were small, my mother, my brother and my sisters, we all lived in my grandmother’s house.

Ha’e *o-pu’ã voi-eterei.*
3.pron A3-get.up early-very

‘She got up very early.’ [C]

(vi) **Demonstrative noun phrases** Guaraní has a series of demonstrative markers that precede the noun. Two are illustrated in (15): The demonstrative *pe* in (15a) indicates proximity to the addressee, and the demonstrative *ko’ã* in (15b) indicates plurality, but not relative distance to the speaker (Gregores and Suárez 1967:141). In contrast to pronouns, the referent of a demonstrative noun phrase has to be activated but need not be in focus.

- (15) a. Context: A gentleman had a daughter who abandoned her child.

O-mbo-kakuaa karai pe mitã’í
A3-CAUS-grow gentleman that child

‘The gentleman raised that child.’ [C]

- b. Context: The duck and the frog are friends living in the pond.

Ha upé-icha, ko’ã mokõi mymba-mi o-po vy’á-pe y-no’õ-me.
and this-like these two animal-DIM A3-jump happy-PE water-united-PE

‘And this is why these two animals jump happily together in the pond.’ [C]

(vii) **Proper names** Like pronouns, proper names do not have descriptive (nominal) material. They refer to entities that are discourse-old, but need not be activated or in focus.

- (16) *O-jevy jey Juán-chi.*
A3-return again Juan-little
'Juanito went back (home).'
- [C]

(viii) **Possessive noun phrases** The possessor in possessive noun phrases is realized by a set B prefix, as in *i-memby* 'her child' in (17a); it can be additionally specified by a pre-nominal expression as in *Maria i-memby* 'Maria's child'. Only the first kind occurs in our corpus.

- (17) a. Context: A woman has a child out of wedlock.
I-memby *o-heja ha o-ho mombyry.*
B3-child A3-leave and A3-go far
'She left her child and went far away.'
- [C]
- b. *O-nohẽ sapy'a umi karia'y h-embí'u ha i-sapatu-kuéra.*
A3-remove suddenly those young.man B3-food and B3-shoe-PL
'Finally, the young people gathered their food and shoes.'
- [C]

Possessive noun phrases can denote discourse-old entities, such as the child in (17a), entities whose existence the hearer can be assumed to know (hearer-old) or entities whose existence the (cooperating) hearer is willing to accommodate (i.e. the hearer acts as if the information was part of his/her background assumptions, Stalnaker 1974). (17b) is an example of the last kind: the referents of the possessive noun phrases *h-embí'u* 'their food' and *i-sapatu-kuéra* 'their shoes' were not previously mentioned (not discourse-old) and the speaker cannot assume that the hearer knows they exist (not hearer-old) but the hearer accommodates their existence.

2.5 Word order

For word order, we annotated which of the arguments of a verb were overtly realized, and the position in which each overt argument occurred with respect to the verb and the other overt arguments. (18), for example, was annotated as VSO, since both the subject and the object are realized as overt noun phrases

(*karai* ‘gentleman’ and *pe mitã’í* ‘that child’, respectively), the subject follows the verb, and the object follows the subject.

- (18) *O-mbo-kakuaa karai pe mitã’í.*
 A3-CAUS-grow gentleman that child
 ‘The gentleman raised the child.’ [C]

Clauses where one or more arguments are not realized with a pronoun or a noun phrase are annotated accordingly. In (19), for example, the transitive verb *o-heja* ‘A3-leave’ is realized only with the (pre-verbal) object *i-memby* ‘her child’ and the subject of the intransitive verb *o-ho* ‘A3-go’ in the same example is referred to only with a cross-reference prefix on the verb. These clauses are annotated as OV and V, respectively.

- (19) **I-memby** *o-heja ha o-ho mombyry.*
 B3-child A3-leave and A3-go far
 ‘She left her child and went far away.’ [C]

2.6 Coding decisions

In the process of annotating the corpus, we made a number of coding decisions that we report on here in the hope of making our quantitative study as transparent and replicable as possible. First, identificational/specificational constructions, which in Guaraní can consist of two juxtaposed noun phrases, such as *jakare* ‘crocodile’ and *peteĩ mymba* ‘an animal’ in (20), were excluded from the study.

- (20) *Jakare peteĩ mymba oi-kó-va y ha yvy-pe.*
 crocodile one animal A3-live-RC water and earth-PE
 ‘The crocodile is an animal that lives in the water and on the ground.’ [C]

Neither of the two noun phrases is realized with a cross-reference marker, i.e. functions as a verb. We exclude such constructions from our study since it is not clear how their word order properties should be compared to examples that consist of verbs and noun phrase arguments.

Second, in examples where a subject argument is realized with two coordinated verbs, the subject was only counted as overtly realized for the first verb. For example, in (21), the noun phrase *i-sy* ‘their mother’ was annotated as the overt subject argument of the first verb (*o-gueraha* ‘A3-take’) only, not of the the second one (*o-mbo-je-poi* (A3-CAUS-JE-release) ‘throw’).

- (21) *Umi jakare-ra'y i-sy o-gueraha ha o-mbo-je-poi y-pe.*
 those crocodile-young B3-mother A3-take and A3-CAUS-JE-release water-PE
 ‘Their mother takes the crocodile young and throws them into the water.’ [C]

Likewise, if a single argument was realized with verbs in a serial verb construction (see Velázquez-Castillo (2004b) for discussion), only the first verb was annotated as if it occurred with an overt argument. In (22), for example, where the subject *ju'i* ‘frog’ is the understood subject of both *o-po* ‘A3-jump’ and of *o-je-poi* (A3-JE-release) ‘fall’, only the first verb was annotated as having an overt argument realized.

- (22) **Ju'i** *o-po o-je-poi chu-gui y-pe.*
 frog A3-jump A3-JE-release 3-ABL water-PE
 ‘The frog jumped and fell into the water.’ [C]

Third, constructions with the motion verb root *ho* ‘go’ were treated as serial verb constructions (as in (23a)),¹³ unless the motion verb was marked with the cotemporaneity marker *-vo* ‘-AT’, as in (23b), in which case we took it to be a (progressive) aspectual marker and did not annotate it.

- (23) a. **A-ha** *a-je-po-reko algun mymba*
 A1sg-go A1sg-JE-hand-have some animal
 ‘I’ll go catch some animal.’ [C]
- b. *O-segi mombyry o-hó-vo chu-pe-kuéra ju'i*
 A3-follow far A3-go-AT 3-PE-PL frog
 ‘The frog was following them far.’ [C]

Fourth, a possessive noun phrase that occurred with a demonstrative marker, such as *upe hogue* (that B3.leaf) ‘his leaf’ in (24), was annotated as a possessive noun phrase.

- (24) *Kyju kiri-kiri ho'u-pa avei upe hogue, ha oi-ke i-kuára-pe,*
 cricket (cricket sound) A3.eat-COMplete also that B3.leaf and A3-enter B3-hole-PE
o-vy'a-há-pe
 A3-happy-NOM-PE
 ‘The cricket also finished its leaf and happily went to its hole.’ [C]

Possessive noun phrases that were marked with *la* were also coded as possessive noun phrases (cf. examples (25e,f)).

¹³The first person form of the verb root *ho* ‘go’ is *a-ha* ‘A1sg-go’.

3 Constraints on word order: results of the quantitative study

There are 573 main, declarative clauses in our corpus, which we evaluated with respect to constraints on word order. These 573 clauses consist of 305 intransitive clauses and 268 (di)transitive clauses. 165 of the (di)transitive clauses contain verbs that can take entity-denoting objects; the other 103 have verbs that take proposition-denoting or spatio/temporal objects. We therefore consider the full set of 573 clauses when examining subjects, but, when considering direct objects, only the 165 (di)transitive clauses that contain verbs that can take an entity-denoting direct object. A first result of our study is a difference in how often subjects and direct objects are overtly realized. Of the 305 intransitive verbs, 111 (36%) are realized with their subject argument, i.e. the majority of intransitive verbs are realized without the subject. The same is true for (di)transitive verbs, where 84 of 268 (31%) occur with a subject. Direct objects, on the other hand, are much more often realized than not: 130 of the 165 (79%) (di)transitive verbs are realized with a direct object noun phrase.

Grammatical function is one of the strongest predictors of noun phrase placement in our study. The data in Table 2 reveal that there is a very strong preference for direct object noun phrases to be realized post-verbally: 124 of the 130 (95%) objects occur post-verbally.¹⁴ Subjects overall show a very slight preference for pre-verbal realization (108 of 195, 55%), and all 47 indirect objects are realized in post-verbal position.

	Subjects	Direct objects	Indirect objects
Pre-verbal	108	6	0
Post-verbal	87	124	47
Total	195	130	47

Table 2: Grammatical function and noun phrase placement

Since there is a categorical restriction against pre-verbal indirect objects, the remainder of our study is concerned only with constraints on the placement of the 325 subject and direct object noun phrases.

¹⁴The preference for objects to be realized post-verbally was weaker in the study on object placement reported in Velázquez-Castillo (1995), where only 55 of 69 (80%) objects occur post-verbally (p.570, Table 1). A possible explanation for the difference is that the two studies include different sets of objects. If Velázquez-Castillo's study included objects in embedded clauses (no indication is given that only main clause objects were considered), we can hypothesize, given that embedded clauses are more conservative with respect to change than main clauses (Givón 1979, Hock 1986) and that OV was the preferred order at an earlier stage of the language (cf. footnote 2), that the difference in object placement between the two studies is a result of objects in embedded clauses being more likely to be realized in pre-verbal position than objects of main clauses.

Noun phrases across all animacy categories may occur in pre- and post-verbal position and all exhibit a preference for post-verbal realization, but to varying degrees. As illustrated in Table 3, 70 of 117 (60%) noun phrases with human referents occur post-verbally, 56% of noun phrases with animate referents (58 of 104), and 80% of noun phrases with inanimate referents (83 of 104). Since five of the eight stories in our corpus are fables where animals (annotated as ‘animate’) act as humans, we group together the categories ‘human’ and ‘animate’ in the remainder of this study. What emerges for the effect of animacy on noun phrase placement is that human/animate noun phrases¹⁵ exhibit a much smaller preference for post-verbal realization (128 of 221, 58%) than inanimate noun phrases (80%).

	Human	Animate	Inanimate	Human/Animate
Pre-Verbal	47	46	21	93
Post-Verbal	70	58	83	128
	117	104	104	221

Table 3: Animacy and noun phrase placement

If we cross-pair the features grammatical function and animacy (Table 4), we find that most inanimate noun phrases are objects (75 of 104, 72%) whereas most human/animate noun phrases are subjects (166 of the 221 (75%)). This is expected since human/animate referents are more likely to be agents (and hence subjects) than inanimate entities. Likewise, subjects are much more likely to be human/animate (85%) than objects (42%).

	Human/Animate	Inanimate	
Subjects	166	29	195
Objects	55	75	130
	221	104	

Table 4: Animacy and grammatical function

The data presented in Table 5 identifies that noun phrase form, and hence the discourse status of the denotation of the noun phrase, has an effect on noun phrase placement in Guaraní. In particular, noun phrases denoting hearer-new entities (i.e. *peteĩ*-marked and hearer-new bare noun phrases) exhibit a strong prefer-

¹⁵For convenience we sometimes say e.g. ‘a human noun phrase’ instead of ‘a noun phrase whose denotation is a (set of) human(s)’.

ence for post-verbal realization (only 2 of 34 (6%) are pre-verbal). The remaining 291 noun phrases have denotations that are discourse-/hearer-old (or accommodatable), and 112 of these (38%) noun phrases are pre-verbal. Discourse-/hearer-old bare noun phrases exhibit the strongest preference for pre-verbal realization (36 of 71 (51%) are pre-verbal).

	<i>peteĩ</i> -marked NPs	Hearer-new bare NPs	Discourse-/hearer-old bare NPs	Pronouns	Demonstrative NPs	Names	Possessive NPs	<i>la</i> -marked NPs
Pre-Verbal	0	2	36	38	7	9	21	1
Post-Verbal	19	13	35	63	11	12	47	11

Table 5: Noun phrase form and noun phrase placement

When we cross-pair noun phrase form with grammatical function (Table 6), we find that noun phrases denoting hearer-new entities (*peteĩ*-marked noun phrases and hearer-new bare noun phrases) tend to be objects (only 6 of 34 (18%) are subjects), while noun phrases denoting discourse-/hearer-old entities tend to be subjects (189 of 291 (65%) are subjects). This is not surprising since, as discussed above, noun phrase form correlates with discourse status and subjects tend to be more topical than objects (Foley and van Valin 1984, Givón 1990, Lambrecht 1994). The figure for discourse-/hearer-old noun phrases is even higher (74%) if possessive noun phrases and *la*-marked ones, which realize more objects than subjects, are excluded. This is expected for possessive noun phrases since they are much more likely than other discourse-/hearer-old noun phrase forms to have inanimate denotations (among subjects, for example, 56% of possessive noun phrases denote inanimate entities, compared to 7% for discourse-/hearer-old bare noun phrases and 1% for pronouns, cf. section 3.1). For *la*-marked noun phrases the difference in grammatical function realized suggests that they are not entirely equivalent to discourse-/hearer-old bare noun phrases (even though both are translated by English definite noun phrases with *the*).

Looking at the figures in Table 6 in a different way, we see that only 6 of 291 (2%) subjects denote hearer-new entities, while objects denote such entities much more frequently (28 of 130 (22%)).

	<i>petẽ</i> -marked NPs		Discourse-/hearer-old bare NPs					
		Hearer-new bare NPs		Pronouns	Demonstrative NPs	Names	Possessive NPs	<i>lã</i> -marked NPs
Subject	4	2	56	70	13	19	27	4
Object	15	13	15	31	5	2	41	8

Table 6: Noun phrase form and grammatical function

In conclusion, we find that grammatical function, animacy and discourse status (in the guise of noun phrase form) affect the placement of noun phrases in Guaraní, although to varying degrees. Subjects and objects differ in their placement preferences (objects exhibit a strong preference for post-verbal placement), and also in that subjects tend to be human/animate and discourse-/hearer-old, while objects tend to be inanimate and are more likely to realize hearer-new entities than subjects. Since subjects and objects exhibit a striking difference in placement, we explore constraints on noun phrase placement separately for subjects and objects in the following two sections.

3.1 Placement of subject noun phrases

We find that a variety of factors constrain the placement of subject noun phrases, especially transitivity and discourse status. Subjects overall show no particular realization preference (recall that 55% are pre-verbal) since different types of subject noun phrases exhibit different sets of constraints. A first indication of this (cf. Table 7) is that only intransitive subjects exhibit a pre-verbal preference (68 of 111 (61%) are pre-verbal) while transitive subjects in fact have a slight preference for post-verbal realization (44 of 84 (52%) are post-verbal). (In Table 7, ‘tr’ stands for transitive and ‘it’ for intransitive.) The data in Table 7 also show that transitive and intransitive subjects differ with respect to animacy: transitive subjects are more likely to be human/animate than intransitive subjects (79 of 84 (94%)) versus 87 of 111 (78%). Since human/animate entities are more likely to be topical than inanimate entities (e.g. Croft 1990), we can hypothesize that transitive subjects are more likely to be topical than intransitive ones. (We investigate this hypothesis below using Givón’s topicality measure.)

	Human/Animate		Inanimate	
	it	tr	it	tr
Pre-verbal	52	37	16	3
Post-verbal	35	42	8	2
	87	79	24	5

Table 7: Animacy and (in)transitive subject placement

The data in Table 8 show that it is not possible to identify a pattern for the placement of noun phrases according to (i) discourse status or (ii) form. Regarding (i), consider for example intransitive subject noun phrase types that denote discourse-/hearer-old entities: some exhibit a pre-verbal preference (bare noun phrases, pronouns and possessive noun phrases) while others exhibit a post-verbal preference (names) or no preference at all (demonstratives). Likewise, only some transitive subject noun phrases that denote discourse-/hearer-old entities exhibit a post-verbal preference (pronouns, possessive noun phrases), while others exhibit no preference (bare noun phrases) or a pre-verbal preference (names). Even if we consider particular noun phrase forms (ii) no pattern emerges: Pronouns, for example, exhibit a pre-verbal preference when they realize intransitive subjects and a post-verbal pattern when they realize transitive subjects, and the pre-verbal preference of discourse-/hearer-old bare intransitive subject noun phrases is not reflected for their transitive counterparts.

	<i>peter</i> -marked NPs		Hearer-new bare NPs		Discourse-/hearer-old bare NPs		Pronouns		Demonstrative NPs		Names		Possessive NPs		<i>la</i> -marked NPs	
	it	tr	it	tr	it	tr	it	tr	it	tr	it	tr	it	tr	it	tr
Pre-verbal	0	0	0	2	22	12	25	13	5	2	2	8	13	3	1	0
Post-verbal	4	0	0	0	9	13	14	18	5	1	4	5	5	6	2	1

Table 8: Subject form and placement

Rather, the data in Table 8 reveal that the overall trend reported above for transitive and intransitive subjects is due to the placement of a particular subset of noun phrase forms: Intransitive subjects overall have a preference for pre-verbal realization because of the strong pre-verbal preference of discourse-/hearer-old bare noun phrases (22 of 31, 71%), pronouns (25 of 39, 64%) and possessive noun phrases (13 of 18, 72%). Among transitive subjects, the slight overall preference for post-verbal realization is due to the post-verbal preference of the same set of noun phrase types, namely discourse-/hearer-old bare noun phrases (13 of 25, 52%), pronouns (18 of 31, 58%) and possessive noun phrases (6 of 9, 67%). The three noun phrase forms that affect the overall placement of transitive and intransitive subjects make up for more than three quarters of the subject noun phrases (153 of 195, 78%). Since discourse status does not emerge as a predictor for noun phrase placement when all noun phrases are considered together, we examine the effect of topicality on noun phrase placement for these three noun phrase forms individually.

In Givón (1983), topicality is a scalar property which is measured in terms of the recurrence and continuity of the denotation of the noun phrase. The two measurements of topicality are *referential distance* and *topic persistence*. Referential distance measures the thematic continuity of the denotation of the noun phrase; it is defined as the average distance in number of clauses between the present mention and the first mention preceding the present mention. The lower the referential distance of a discourse participant, the higher its degree of thematic continuity, i.e. the more topical.¹⁶ Topic persistence measures the number of contiguous subsequent clauses in which the denotation of the noun phrase is mentioned. More continuous, hence more topical, participants have a larger topic persistence.

Of the three subject noun phrase forms we examined, pronouns, not surprisingly, emerge as the most topical, with a referential distance (RD) of 2.37 and a topic persistence (TP) of 5.66 (70 tokens). Discourse-/hearer-old bare noun phrases were less topical than pronouns (RD = 6.23, TP = 4.29, 56 tokens), but still more topical than possessive noun phrases (RD = 8.93, TP = 1.89, 27 tokens). As expected from the data in Table 8, no particular position (pre- or post-verbal) or transitivity type (transitive or intransitive subject) emerged as realizing topical subjects when all three noun phrase forms were examined together. We therefore examine each noun phrase form in turn to identify constraints on its placement.

¹⁶Following Givón (1983), a value of 20 was assigned for referential distance if no previous mention was found within 20 clauses; new mentions were also assigned a referential distance of 20.

Pronouns Given that subject pronouns code highly topical entities (compared to other subject noun phrase types) it is not surprising that almost all subject pronouns in the corpus refer to human/animate entities:

	Human/Animate		Inanimate	
	it	tr	it	tr
Pre-verbal	25	12	0	1
Post-verbal	14	18	0	0

Table 9: Animacy and (in)transitive pronoun subjects

As mentioned above, transitive and intransitive subject pronouns differ in their placement. They also differ in their relative topicality: transitive subject pronouns (RD = 1.38, TP = 5.73, 31 tokens) are more topical than intransitive subject pronouns (RD = 2.99, TP = 5.52, 39 tokens). Furthermore, post-verbal subject pronouns (RD = 2.00, TP = 6.29, 32 tokens) are more topical than pre-verbal pronouns (RD = 2.66, TP = 5.04, 38 tokens), similarly to Velázquez-Castillo’s (1995) findings for objects. The figures for the individual subject pronoun types are given in Table 10.

	RD	TP	# of tokens
pre-verbal intransitive	3.46	4.88	25
pre-verbal transitive	1.29	5.36	13
post-verbal intransitive	2.53	6.67	14
post-verbal transitive	1.59	6.00	18

Table 10: Referential distance and topic persistence of pronoun subjects

We see that post-verbal transitive pronoun subjects are among the most topical pronoun subjects (together with pre-verbal transitive ones), while pre-verbal intransitive pronoun subjects code the least topical entities. The following sections show that other kinds of subject noun phrases pattern differently.

Discourse-/hearer-old bare noun phrases Just like pronouns, discourse-/hearer-old bare noun phrases hardly ever denote inanimate entities:

	Human/Animate		Inanimate	
	it	tr	it	tr
Pre-verbal	21	12	1	0
Post-verbal	6	13	3	0

Table 11: Animacy and (in)transitive discourse-/hearer-old bare subjects

In contrast to subject pronouns (and objects), discourse-/hearer-old pre-verbal bare noun phrases are more topical than post-verbal ones (pre-verbal: RD = 5.48, TP = 4.61 (34 tokens) versus post-verbal: RD = 6.98, TP = 3.77 (22 tokens)), but transitive and intransitive discourse-/hearer-old bare noun phrase subjects do not show a difference in topicality since transitive ones have a smaller referential distance but also a lower topic persistence (transitive: RD = 5.51, TP = 4.03 (25 tokens), versus intransitive: RD = 6.52, TP = 4.48 (31 tokens)). Overall, post-verbal intransitive bare noun phrases code the least topical entities and pre-verbal intransitive and post-verbal transitive bare noun phrases both code topical ones:

	RD	TP	# of tokens
pre-verbal intransitive	4.14	4.91	22
pre-verbal transitive	7.93	4.07	12
post-verbal intransitive	12.33	3.44	9
post-verbal transitive	3.27	4.00	13

Table 12: Referential distance and topic persistence of pre- and post-verbal definite bare subjects

Possessive noun phrases As reported above, noun phrases of this type were the least topical of the three types examined, and, as expected, we find that they are more likely to denote inanimate entities than the other two types:

	Human/Animate		Inanimate	
	it	tr	it	tr
Pre-verbal	3	3	10	0
Post-verbal	1	5	4	1

Table 13: Animacy and (in)transitive possessive noun phrase subject

As illustrated in Table 14, no particular possessive noun phrase subject emerges from the topicality study as coding topical entities: Those types with a low referential distance (pre-verbal transitive, post-verbal intransitive) also have a low topical persistence. Transitive and intransitive possessive noun phrase subjects do not differ in topicality, and nor do pre- and post-verbal ones.

	RD	TP	# of tokens
pre-verbal intransitive	9.31	2.23	13
pre-verbal transitive	8.33	0.67	3
post-verbal intransitive	7.80	0.60	5
post-verbal transitive	9.43	2.71	6

Table 14: Referential distance and topical persistence of possessive noun phrase subjects

In conclusion, the placement of subject noun phrases with respect to the verb is subject to a variety of factors. Subjects coding hearer-new entities exhibit a strong preference for post-verbal realization, but no particular position emerges for discourse-/hearer-old subject noun phrases as a whole. Rather, transitive and intransitive subjects differ in their realization preferences, as do different noun phrase forms that realize discourse-/hearer-old entities.

3.2 Placement of object noun phrases

Only six of the 130 objects in our study occur pre-verbally. The six examples containing these objects are given in (25), with the pre-verbal object in bold-face.

- (25) a. **I-memby** *o-heja*.
 B3-child A3-leave
 ‘She left her child.’
- b. **Yvypóra-pe** *rei-pytyvõ*
 human(kind)-PE A2sg-help
 ‘You help humans.’
- c. **Ha i-compañera-kuéra-pe** *avei a-ankarga:...*
 and B3-school.colleague-PL-PE too A1sg-charge
 ‘And I also charged her school friends:...’

- d. *Juán-chi ju'i rova-ité-rehe o-maña.*
 Juan-little frog face-very-REHE A3-look
 'Juanito was looking right at the frog.' (Lit.: ...at the frog's face.)
- e. *Ndé-ngo la nde-nérvio re-mondoho-pa.*
 B2sg-EMPH the B2sg-nerve A2sg-rip-COMplete
 'You completely cut your nerves.'
- f. *Ha'e la i-kuã n-o-ñandu-vé-i lo do.*
 3.pron the B3-finger NEG-A3-feel-more-NEG the two
 'She didn't feel her two fingers anymore.'

The denotations of the six pre-verbal objects in the corpus fall into two semantic classes: individuals (and individual kinds) (*i-memby* 'her child' in (25a), *yvypóra* 'humans' in (25b), *i-compañera-kuéra* 'her school friends' in (25c)) and body parts (*ju'i rova* 'frog's face' in (25d), *la nde-nérvio* 'your nerves' in (25e), *la i-kuã* 'her fingers' in (25f)). All of these noun phrases are also unified by their discourse status since they denote discourse-/hearer-old entities. Thus, if we assume that body parts are animate (contrary to our annotation scheme), we can hypothesize that objects are realized post-verbally unless they denote discourse-/hearer-old entities that are (at least) animate. Although our corpus contains fewer pre-verbal objects than that of Velázquez-Castillo (1995), our study replicates her finding that pre-verbal objects are not continuous topics (p.571): the referential distance for these six objects is (a high) 16.83 and only the referent of (25c) is mentioned in subsequent discourse, and only in one clause (pre-verbal objects have low topic persistence).¹⁷ Thus, in our corpus, only a very particular kind of object is realized pre-verbally, namely those that denote discourse-/hearer-old but non-topical human/animate entities. (They do, however, also occur post-verbally.)

3.3 Summary

The main conclusions about constraints on word order in Paraguayan Guaraní are summarized here:

- **Grammatical function:** Grammatical function is a strong predictor for the placement of direct and indirect objects, which occur post-verbally in 95% and 100% of the examples, respectively. Subjects overall only show a slight preference for pre-verbal realization (55%).

Subjects and direct objects differ in the extent to which they are overtly realized. The majority of subjects (34%) are not overtly realized, while the majority of direct objects (79%) are overtly realized.

¹⁷Velázquez-Castillo (1995) reports that pre-verbal object noun phrases have a referential distance of 20 and a topic persistence of 0, while post-verbal object noun phrases have a referential distance of 13.01 and a topic persistence of 1.16.

- **Animacy:** Inanimate noun phrases are more likely to be realized post-verbally than human/animate noun phrases, and only direct objects that denote human/animate (and discourse-/hearer-old) entities are realized pre-verbally. The majority of inanimate noun phrases realize objects while human/animate noun phrases are more likely to be subjects.
- **Discourse status:** Noun phrases denoting discourse-/hearer-old entities are more likely to be realized in pre-verbal position than those denoting hearer-new entities, which are almost exclusively realized post-verbally. Only non-topical objects that denote human/animate discourse-/hearer-old entities are realized pre-verbally.
- **Subjects:** Unlike objects, subjects overall do not show a particular pattern of realization since the placement of different types of subjects depends on different sets of constraints. **Transitivity** affects the placement of subjects: intransitive subjects exhibit a preference for pre-verbal realization while transitive subjects exhibit a post-verbal preference; transitive subjects are more likely to be human/animate than intransitive subjects. The vast majority of subjects denote discourse-/hearer-old entities, and the **discourse status** of the denotation of different noun phrase forms affects their placement. The clearest results emerged for pronouns where post-verbal transitive pronouns realize the most topical referents while pre-verbal intransitive ones code the least topical ones.

4 Does Guaraní have a basic word order?

Having identified constraints that govern the order of expressions at the sentential level in Guaraní, we turn to the question of whether the language has a basic word order. As mentioned in section 1, Gregores and Suárez (1967) claim that SVO, as the most frequent word order, is the basic word order of Guaraní. The finding that the vast majority of direct objects are realized after the verb (cf. also Velázquez-Castillo (1996)) provides partial support for this claim, but frequency is not the only criterion that is employed to determine the basic word order of a language. In Tojolabal and Yucatec Maya, for instance, the most frequent word order (SVO) is more pragmatically marked than a verb-initial word order, which could hence be considered the basic word order on the basis of pragmatic markedness (Brody 1984, Durbin and Ojeda 1978). In this section, we discuss whether Guaraní has a basic word order on the basis of three criteria: frequency, markedness, and disambiguation.

4.1 Frequency

According to this criterion, the basic word order of a language is the one that is most frequent. Table 15 gives an overview of the word orders of the 41 clauses in the corpus that realize a verb together with its subject and direct object; the most frequent word order is SVO, with 18 instances.

Word order	SVO	SOV	OVS	OSV	VSO	VOS	Total
# of instances	18	3	0	0	11	9	41

Table 15: Word orders attested in the corpus

While one might take the finding that the SVO word order is the most frequent one as support for the claim that SVO is the basic word order of Guaraní, the 18 examples (44%) do not make for an overwhelming frequency overall. However, all VO word orders taken together (38 of 41 examples, 93%) suggest that Guaraní is a VO language on the basis of the frequency criterion. This coincides with the finding that 94% of objects in the whole corpus are realized post-verbally. In the 41 examples, subjects occur to the right of the verb almost as often as to the left of the verb (20 cases of VS word order, 21 cases of SV word order), again coinciding with the placement pattern of all subjects in the corpus (53% of all subjects occur pre-verbally). Hence, the frequency criterion does not find support for a particular position of the subject with respect to the verb. There does, however, seem to be a preference for subjects to precede objects (rather than follow them) as 32 of the 41 (78%) examples have an SO word order. In eight of the nine examples with an OS word order, a pronoun realizes the post-verbal object (the ninth object is a very short word *vai* ‘bad, ugly, ruin’). We can conclude that subjects precede objects unless the post-verbal object is a pronoun or a very short word.

4.2 Markedness

According to this criterion, the least marked word order is the basic word order. There are several respects in which an order may be (un)marked; since we do not have phonological or stylistic information available for the corpus, we focus on morphological, syntactic and pragmatic markedness.

Sentences are morphologically marked if a particular word order requires the arguments or the verb to be marked in a certain way (e.g. definiteness or plurality for arguments; voice, aspect or mood for verbs). We were not able to discern any requirements on verbal marking for any order, in particular since many (26)

of the 41 verbs do not contain any marking and realized all four of the attested orders. Although negated verbs (5 tokens), verbs marked with aspect/mood markers (9) and causativized verbs (2)¹⁸ do not occur with all four word orders, each marked verb type is attested in sentences with an SV and a VS word order. Hence, morphological markedness of the verb does not identify a basic position for the subject with respect to the verb. We do, however, observe a morphological markedness restriction for objects since there are no pre-verbal *peteĩ*-marked or discourse-new bare noun phrase objects in the corpus. Since noun phrase form is related to discourse status, it is likely that this is a pragmatic markedness restriction according to which object noun phrases denoting hearer-new entities are restricted to post-verbal position (cf. section 3.2). This restriction does not appear to be categorical for subjects since two pre-verbal hearer-new bare noun phrase subjects are attested (and both are transitive subjects in clauses where the object is also realized). Since hearer-new bare subject noun phrases occur pre- and post-verbally, and different types of highly topical discourse-/hearer-old subjects do not show a uniform pattern of realization either (cf. section 3.1), no conclusion can be drawn about a pragmatically unmarked position for subjects with respect to the verb.

Word orders are syntactically marked if they can only be realized by syntactically marked, e.g. bi-clausal, constructions. As mentioned in section 2, we consider in this study only main, declarative sentences to make the sentences under consideration as homogenous as possible, and excluded identificational/specificational constructions from our study, as well as sentences where an argument was overtly focused (e.g. using the intensifying marker *-nte* ‘only/just’). Since our corpus already consists of the syntactically least marked sentences, an examination of the 41 transitive sentences with both the subject and the object realized did not lead to the identification of additional constraints on word order.

4.3 Disambiguation

This criterion regards as the basic word order the order that is preferred in potentially ambiguous sentences (cf. Chomsky 1965, Dixon 1972, Pullum 1977). We illustrate this criterion with Russian, a language that has noun classes in which the nominative and accusative cases have the same endings. Two nouns in these classes are *mat’* ‘mother’ and *doč’* ‘daughter’. As indicated by the glosses in (26), these two forms are indistinguishable when realized in the nominative (NOM) and accusative (ACC) cases. Jakobson (1963) reports

¹⁸One of the 41 examples has both a causative and an aspectual marker, which is why the total number of examples listed here adds up to 42.

that although both SVO and OVS are generally possible word orders in Russian, transitive sentences like (26) which realize two such nouns cannot be interpreted according to the OVS word order.

(26) a. *Mat' ljubit doč*
 mother-NOM/ACC loves daughter-NOM/ACC
 'The mother loves the daughter.'
 (Not: 'The daughter loves the mother.')

b. *Doč ljubit mat'*
 Daughter-NOM/ACC loves mother-NOM/ACC
 'The daughter loves the mother.'
 (Not: 'The mother loves the daughter.')

(Jakobson 1963:269)

Thus, for sentences that are potentially ambiguous, SVO is the disambiguating word order, and would be considered the basic word order of Russian according to the disambiguation criterion.

Since at least one argument is cross-referenced on a transitive verb in Guaraní, only transitive sentences with two third person arguments are potentially ambiguous. They can be disambiguated by a variety of factors, including animacy, suffixes such as the (in)direct object marker *-pe*, and word order. Of the 41 transitive sentences in the corpus that realize both the subject and the object noun phrase, four are unambiguous because one of the arguments is a non-third person. Of the remaining 37 potentially ambiguous examples, 19 are disambiguated by animacy in combination with world knowledge. (27) is such an example: here, the verb *nohẽ* 'remove' is realized with the noun phrases *umi karia'y* 'those young men' and *h-embu'u ha i-sapatu-kuera* (B3-food and B3-shoe-PL) 'their food and shoes'. Since it is world knowledge that humans gather food and shoes (and not the other way around), animacy (in combination with world knowledge) identifies the noun phrase directly following the verb as the subject, and the other noun phrase as the object.

(27) *O-nohẽ sapy'a umi karia'y h-embu'u ha i-sapatu-kuera.*
 A3-remove suddenly those young.man B3-food and B3-shoe-PL

'Finally, the young men gathered their food and shoes.'

[C]

Another 16 of the 37 potentially ambiguous examples are disambiguated by a suffix on the non-subject noun phrase argument. The examples in (28) illustrate two such suffixes, namely *-pe* in (28a) and *-rehe*, which is subcategorized for by the verb *maña* 'look' (28b).

- (28) a. *Ha upéi o-hecha sapy'a Juán-chi ha Pirúlo ju'í-pe.*
 and then A3-see suddenly Juan-little and Pirulo frog-PE
 'And then suddenly Juanito and Pirulo saw the frog.' [C]
- b. *Juán-chi ju'i rova-ité-rehe o-maña*
 Juan-little frog face-very-REHE A3-look
 'Juanito was looking right at the frog.' (Lit.: ...at the frog's face.) [C]

Finally, two of the 41 examples with both the subject and the object noun phrases realized are not disambiguated by person marking, animacy (in combination with world knowledge) or a suffix on the object noun phrase. The word order of these two examples is VSO (29a) and SVO (29b).

- (29) a. *O-mbo-kakuaa karai pe mitã'í.*
 A3-CAUS-grow gentleman that child
 'The gentleman raised the child.' [C]
- b. *Tuju-ry o-je-agara-pa la ij-ao.*
 mud-liquid A3-JE-grab-COMPLETE the B3-cloth
 'Mud got all over his clothing.' [C]

In both of these examples, the subject precedes the object, which suggests that the order SO is a disambiguating order in Guaraní, and hence a basic word order according to this criterion.¹⁹ A basic position of the subject with respect to the verb cannot be determined.

These strategies for disambiguating potentially ambiguous transitive sentences are confirmed by the judgments of our Guaraní consultants when presented with examples such as (30a) and (30b), and the other five logically possible permutations of the verb, subject and object.

- (30) a. **Juan o-topa peteĩ ita.**
 Juan A3-find one stone
 'Juan found a stone.' [E]
- b. **Juan o-hecha Maria.**
 Juan A3-see Maria
 'Juan saw/sees Maria.' [E]

¹⁹Seki (2000:166) finds that pragmatically unmarked transitive clauses in Kamaiurá are most frequently realized in the SOV order. It is also the disambiguating word order since transitive clauses where the grammatical function of noun phrases is not determined by overt markers are interpreted according to the SOV order. Transitive clauses with only one argument realized preverbally are ambiguous. Such examples are not necessarily ambiguous in Paraguayan Guaraní: (25a), for example, cannot mean 'Her child left her'; this would be expressed as *I-memby o-heja chu-pe* 'B3-child A3-leave 3-PE'.

All permutations of the two examples in (30) were deemed grammatical by our consultants but the two sentences differed in whether the SO disambiguating word order was applied. In (30a), Juan was consistently judged the subject of the verb *o-topa* ‘A3-find’, and the stone the object, which is consistent with animacy and world knowledge determining argument linking. In (30b), on the other hand, the SO disambiguating word order was applied: Juan was judged to be the subject only if he was mentioned before Maria.

We conclude that there is much evidence in support of the claim that VO is a basic word order in Paraguayan Guaraní. Regarding Gregores and Suárez’ (1967) claim that SVO is the basic word order of the language, we cannot come to a definite conclusion. On the one hand, this is the most frequent word order of the language (but not overwhelmingly so), but other criteria do not suggest that the pre-verbal position of the subject can be considered its basic position.

5 Word order in other Tupí-Guaraní languages

We conclude by comparing the results about basic word order and constraints on word order in Paraguayan Guaraní to those of other Tupí-Guaraní languages. (See Rodrigues (1984) for a classification of the family, and Jensen (1999) for a general introduction.) Since we cannot do justice here to the wealth of descriptive and theoretical work on these languages, this comparison is not meant to be comprehensive; some shorter descriptions we consulted did not contain any claims about word order (constraints).

It is striking that almost all of the authors who discuss word order in Tupí-Guaraní languages make observations similar to those made above for Paraguayan Guaraní: (i) Noun phrases need not be overtly realized in many discourse contexts (e.g. Seki (2000) for Kamaiurá, Rose (2003) for Emerillon, González (2005) for Tapiete), (ii) While one word order might be considered basic, (all) other possible permutations are also attested (e.g. Rose (2003) for Emerillon, Harrison (1976) for Asuriní, Gomes (2002) for Kayabí, Dooley (1982) for Mbyá),²⁰ and (iii) Word order is affected by the discourse context (e.g. Dobson (1997) for Kayabí, Dooley (1982) for Mbyá, Seki (2000) for Kamaiurá, Nicholson (1978) and Rodrigues dos Santos and dos Santos Gomes (2003) for Asurini de Trocará).

²⁰SOV is the basic order of Kamaiurá (Seki 2000), of Nheengatu (Tupan-an 2000), of Tupí Antigo (Navarro 1998), of Asurini (Olson 1978), and of Kayabí (Dobson 1997, 2005). Harrison (1976) considers OVS the basic word order of Asurini. Aragão (1996) states that in Lingua Geral objects occur post-verbally but the position of subjects is rather free.

Mbyá Guaraní, Chiriguano and Tapiete are Tupí-Guaraní languages that belong to the same subgroup of the language family as Paraguayan Guaraní (subgroup 1 of Rodrigues' (1984) classification). According to Dooley (1982, 2006), the unmarked order of Mbyá Guaraní is SVO, and pragmatic factors lead to other, marked word orders. In particular, he claims that the most informative element is given in final position, e.g. a subject noun phrase whose denotation is discourse-new follows the verb (Dooley 1982:320). Martins (2003), on the other hand, maintains that both SVO and SOV are equally preferred for transitive sentences (pp.118, 154). She suggests that the more archaic OV order is still preferred by older people and women, whereas the innovative VO order is used by younger people, as a result from contact with Portuguese (p.163). Interestingly, she points out that Mbyá sentences realized in the SVO or the SOV order that are not disambiguated by animacy or nominal suffixes are ambiguous (p.156), which suggests that neither word order is basic on the basis of the disambiguation criterion. Other tribal Tupí-Guaraní languages tend to have what is believed to be the more archaic OV word order (cf. footnote 2): The basic word order of Chiriguano is SOV (Dietrich 1986:158), but objects can precede and follow the verb (p.127), and the order can be changed for stylistic reasons such as “cuando el complemento es mucho más largo que el verbo” (when the complement [object] is much longer than the verb) (p.158). Likewise, González (2005) reports that the default order of constituents in Tapiete is SOV, although other, pragmatically marked, orders are possible (p.206): for example, objects may appear after the verb if their referent is discourse-new (p.208).

We were able to identify two quantitative word order studies of Tupí-Guaraní languages: Pease (2007) for Parintintin (subgroup 6) and Rose (2003) for Emerillon (subgroup 8). As both languages are only very distantly related to (Paraguayan) Guaraní, it is not surprising that these authors report rather different word order constraints for these languages. While Pease (2007) finds that the most common word order attested in her corpus of Parintintin is SVO, the preferred word order is VSO when the subject is a pronoun (p.70). She also finds that the object may never precede the subject unless both are in front of the verb (p.70). According to Rose (2003), the most frequent word order of Emerillon is SOV when both the subject and the object are realized (p.555). While subjects are almost obligatorily realized pre-verbally (p.550), objects occurs more often post-verbally than pre-verbally (p.554). However, since post-verbal objects are more topical than pre-verbal ones according to Givón's topicality measure (which accords with Velázquez-Castillo's (1995) finding for objects in Paraguayan Guaraní), she concludes that SOV is the basic word order since it is the most frequent and pragmatically unmarked (p.556).

To conclude, grammatical function plays a strong role in determining word order in many Tupí-Guaraní

languages, although the languages differ in whether the subject or the object has a (relatively) fixed position. Givón's topicality measure has provided useful in identifying constraints on the placement of more flexibly realized arguments, as well as comparing the effect of topicality on word order among the various languages. In Paraguayan Guaraní, the topicality of different subject noun phrase forms provided one of the strongest arguments against assigning a basic position to the subject with respect to the verb.

References

- Adelaar, Willem F. H. 1994. The nasal/oral distinction in Paraguayan Guaraní suffixes. *Revista Latinoamericana de Estudios Etnolingüísticos* 8:125–133.
- Aissen, Judith. 1992. Topic and focus in Mayan. *Language* 68(1):43–80.
- Aragão, Batista. 1996. *Gramática Tupy*. Barraca do Fortaleza: Escritor Cearense.
- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco, and Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1):28–53.
- Brody, Jill. 1984. Some problems with the concept of basic word order. *Linguistics* 22:711–736.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, and truth-theoretic semantics. In C. Li, ed., *Subject and Topic*, pages 25–55. New York: Academic Press.
- Choi, Hye-Won. 1999. *Optimizing Structure in Context. Scrambling and Information Structure*. Stanford: CSLI Publications.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Croft, William. 1990. *Typology and Universals*. Cambridge, England: Cambridge University Press.
- Dayley, Jon P. 1981. Voice and ergativity in Mayan languages. *Journal of Mayan Linguistics* 2:3–82.
- Dietrich, Wolf. 1986. *El idioma chiriguano*. Madrid: Ediciones Cultura Hispánica. Instituto de Cooperación Iberoamericana.
- Dixon, R. M. W. 1972. *The Dyirbal Language of North Queensland*. Cambridge, England: Cambridge University Press.
- Dobson, Rose M. 1997. *Gramática práctica com exercícios da língua Kayabi*. Cuiabá, Brazil: Summer Institute of Linguistics.
- Dobson, Rose M. 2005. *Aspectos da língua Kayabi*. Porto Velho, Brazil: Summer Institute of Linguistics, 2nd edn.

- Dooley, Robert. 1982. Options in the pragmatic structuring of Guaraní sentences. *Language* 58(2):307–331.
- Dooley, Robert. 2006. *Léxico guaraní, dialecto mbyá*. Cuiabá, Brazil: Summer Institute of Linguistics.
- Dryer, Matthew. 1992. The Greenbergian word order correlations. *Language* 68(1):81–138.
- Dryer, Matthew. 2005. Order of adpositions and noun phrase. In M. Haspelmath, M. Dryer, D. Gil, and B. Comrie, eds., *The World Atlas of Language Structures*, chap. 85, pages 346–349. Oxford: Oxford University Press.
- Durbin, Marshall and Fernando Ojeda. 1978. Basic word-order in Yucatec Maya. In N. England, ed., *Papers in Mayan linguistics*, no. 2 in Studies in Mayan Linguistics, pages 69–77. Columbia, Missouri: University of Missouri Miscellaneous Publications in Anthropology no. 6.
- Enrico, John. 1986. Word order, focus, and topic in Haida. *International Journal of American Linguistics* 52(2):91–123.
- Foley, William and Robert D. van Valin. 1984. *Functional syntax and universal grammar*. Cambridge, England: Cambridge University Press.
- Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.
- Givón, Talmy. 1983. Topic continuity in discourse: Quantitative cross-language studies. In T. Givón, ed., *Discourse and Syntax*, pages 1–42. Amsterdam: John Benjamins.
- Givón, Talmy. 1990. *Syntax. A Functional/Typological Introduction*. Amsterdam: John Benjamins.
- Gomes, Nataniel dos Santos. 2002. A ordem oracional em kayabí. *Revista Philologus* 23. Article 13.
- González, Hebe Alicia. 2005. *A Grammar of Tapiete (Tupi-Guarani)*. Ph.D. thesis, University of Pittsburgh.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg, ed., *Universals of Language*, pages 73–113. Cambridge, MA: The MIT Press.
- Gregores, Emma and Jorge A. Suárez. 1967. *A Description of Colloquial Guaraní*. The Hague: Mouton de Gruyter.
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2):274–307.
- Harrison, Carl. 1976. *Gramática asuriní*. Brasília, Brazil: Summer Institute of Linguistics.
- Haspelmath, Martin. 2006. Terminology of case. Manuscript, to be published in A. Malchukov & A. Spencer (eds.), *Handbook of Case*, Oxford: Oxford University Press.
- Haspelmath, Martin, Matthew Dryer, David Gil, and Bernard Comrie, eds. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

- Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge, England: Cambridge University Press.
- Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts, Amherst.
- Hock, Hans Henrich. 1986. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.
- Jakobson, Roman. 1963. Implications of language universals for linguistics. In J. H. Greenberg, ed., *Universals of Language*, pages 263–278. Cambridge, MA: MIT Press.
- Jensen, Cheryl. 1999. Tupí-Guaraní. In R. Dixon and A. Y. Aikhenvald, eds., *Amazonian Languages*, pages 125–163. Cambridge, England: Cambridge University Press.
- King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford: CSLI Publications.
- Kiss, Katalin E. 2002. *The Syntax of Hungarian*. Cambridge, England: Cambridge University Press.
- Lambrecht, Knut. 1994. *Information Structure and Sentence Form*. Cambridge, England: Cambridge University Press.
- Martins, Marci Fileti. 2003. *Descrição e análise de aspectos da gramática do Guarani Mbyá*. Master's thesis, Universidade Estadual de Campinas.
- Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4):369–415.
- Mayer, Mercer. 1967. *A Boy, a Dog and a Frog*. New York: Penguin Books.
- Ministerio de Educación y Cultura. 2004. La educación bilingüe en la reforma educativa Paraguaya. Available online at www.escuelaviva-mec.edu.py/investigaciones/60.pdf, last checked 08/30/08.
- Mithun, Marianne. 1992. Is basic word order universal? In D. L. Payne, ed., *Pragmatics of Word Order Flexibility*, pages 15–61. Amsterdam: John Benjamins.
- Navarro, Eduardo de Almeida. 1998. *Método moderno de Tupi Antigo*. Petrópolis, Brazil: Editora Vozes.
- Nicholson, Velda. 1978. *Aspectos da língua assuriní*. Brasília, Brazil: Summer Institute of Linguistics.
- Nordhoff, Sebastian. 2004. *Nomen/Verb-Distinktion im Guaraní*. Master's thesis, University of Cologne, Germany.
- Olson, Gary. 1978. *Descrição preliminar de orações wajapĩ*. Brasília, Brazil: Summer Institute of Linguistics.
- Pease, Helen. 2007. *Parintintin grammar*. Porto Velho, Brazil: Summer Institute of Linguistics. Publication of a 1968 document.
- Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness and information–status. In W. Mann and

- S. Thompson, eds., *Discourse Description: Diverse Analyses of a Fundraising Text*. Amsterdam: Benjamins.
- Pullum, Geoffrey. 1977. Word order universals and grammatical relations. In P. Cole and J. Sadock, eds., *Syntax and Semantics 8: Grammatical Relations*, pages 249–277. New York: Academic Press.
- Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26(3):287–350.
- Rodrigues, Ayrton D. 1984. Relações internas na família lingüística Tupí-Guaraní. *Revista de Antropologia* 27/28:33–53.
- Rodrigues dos Santos, Jociney and Nataniel dos Santos Gomes. 2003. Considerações sobre o novo/velho na língua Asuriní. *Revista Philologus* 17:21–36.
- Rose, Françoise. 2003. *Morphosyntaxe de l’Emerillon: Langue Tupí-Guaraní de Guyane Française*. Ph.D. thesis, Université Lumière Lyon 2.
- Rude, Noel. 1992. Word order and topicality in Nez Perce. In D. Payne, ed., *Pragmatics of word order flexibility*, pages 193–208. Philadelphia: John Benjamins.
- Ruiz de Montoya, Antonio. 1993. *Arte de la lengua Guaraní*. Asunción, Paraguay: Centro de Estudios Paraguayos “Antonio Guasch”, first print in Madrid in 1640 edn. With an introduction and notes by Bartolomeu Melià.
- Seki, Lucy. 2000. *Gramática do Kamaiurá*. Sao Paulo: Editora da Unicamp.
- Silverstein, Michael. 1976. Hierarchies of features and ergativity. In R. Dixon, ed., *Grammatical Categories in Australian Languages*, pages 112–171. New Jersey: Humanities Press Inc.
- Stalnaker, Robert C. 1974. Pragmatic presuppositions. In M. Munitz and P. Unger, eds., *Semantics and Philosophy*, pages 197–213. New York: New York University Press.
- Steele, Susan. 1978. Word order variation: A typological study. In J. H. Greenberg, ed., *Universals of Human Language*, vol. Vol. 4: Syntax, pages 585–624. Stanford: Stanford University Press.
- Tonhauser, Judith. 2006. *The Temporal Semantics of Noun Phrases: Evidence from Guaraní*. Ph.D. thesis, Stanford University.
- Tonhauser, Judith. 2007a. Agent focus and voice in Yucatec Maya. In *Proceedings of the 39th Meeting of the Chicago Linguistics Society*, pages 540–558. Chicago: Chicago Linguistic Society.
- Tonhauser, Judith. 2007b. Nominal tense? The meaning of Guaraní nominal temporal markers. *Language* 83:831–869.
- Tupan-an, Nic. 2000. *Nheengatu-Tupí, Vocabulário e Gramática Tupi-Guarani*. Sao Paulo: STS.
- Velázquez-Castillo, Maura. 1995. Noun incorporation and object placement in discourse: The case of

- Guaraní. In P. Downing and M. Noonan, eds., *Word Order in Discourse*, pages 555–579. Amsterdam/Philadelphia: John Benjamins.
- Velázquez-Castillo, Maura. 1996. *The Grammar of Possession: Inalienability, Incorporation and Possessor Ascension in Guaraní*. Amsterdam: John Benjamins.
- Velázquez-Castillo, Maura. 2002a. Grammatical relations in active systems. The case of Guaraní. *Functions of Language* 9(2):133–167.
- Velázquez-Castillo, Maura. 2002b. Guaraní causative constructions. In M. Shibatani, ed., *The Grammar of Causation and Interpersonal Manipulation*, pages 507–534. Amsterdam: John Benjamins.
- Velázquez-Castillo, Maura. 2004a. Guaraní (Tupí-Guaraní). In G. E. Booij, C. Lehmann, J. Mugdan, and S. Skopeteas, eds., *Morphologie/Morphology: Ein internationales Handbuch zur Flexion und Wortbildung/An International Handbook on Inflection and Word-Formation*, vol. 2, pages 1421–1432. Berlin: Walter de Gruyter.
- Velázquez-Castillo, Maura. 2004b. Serial verb constructions in Paraguayan Guaraní. *International Journal of American Linguistics* 70(2):187–213.
- Zwicky, Arnold. 1992. Jottings on adpositions, case inflections, government and agreement. In D. Brentari, G. Larson, and L. MacLeod, eds., *The Joy of Grammar: A Festschrift for James D. McCawley*, pages 369–383. Amsterdam: John Benjamins.