

Word Order in Paraguayan Guaraní

A Senior Honors Thesis

Presented in Partial Fulfillment of the Requirements for Graduation with Research
Distinction in Linguistics in the Undergraduate Colleges of The Ohio State University

by
Erika Colijn

The Ohio State University
June 2007

Project Advisor: Professor Judith Tonhauser, Department of Linguistics

Abstract

My honors research project examines word order in Paraguayan Guaraní, i.e. the order in which the subject, object, and verb of a sentence are realized. No study of word order has been conducted on Paraguayan Guaraní, and my quantitative study brings Guaraní to bear on cross-linguistic studies of word order. The main research questions I am pursuing are: What word orders are possible in Guaraní and is there a basic word order? Do the factors of grammatical function, animacy, and discourse status affect the placement of arguments? If so, how?

Knowledge of the basic word order of a language is important because it has been found to correlate with other grammatical properties of the language. For example, linguists have observed that there is a relation between basic word order and other properties like the position of adjectives relative to the noun and the type of adposition a language will have: verb-initial languages generally have prepositions, while verb-final languages generally have post-positions.

Grammatical function, animacy, and discourse status have been shown in previous studies to influence the realization of word order. English word order, for example, is strongly constrained by grammatical function; the subject typically comes before the verb, while the object normally follows. In Tojolobal, a Mayan language, word order is influenced by animacy. Of the six possible word orders, there are several that are possible only if certain arguments have a certain type of animacy, such as human or animal. Discourse status, which we measure by looking at definiteness, has also been found to constrain word order. Cayuga, an Iroquoian language, has been shown to realize indefinite noun phrases before definite noun phrases.

This quantitative study overall finds Subject-Verb-Object to be the basic word order of Paraguayan Guaraní, based on the criteria of frequency and disambiguation. I also find that grammatical function and discourse status affect the placement of arguments, while animacy by itself is not much of an influence. When paired together, however, discourse status and animacy do affect the realization of subjects in Paraguayan Guaraní.

Acknowledgments

I would like to thank the Ohio State University Colleges of the Arts and Sciences Honors Committee for awarding me a scholarship with which I was able to complete my research and thesis.

I would also like to thank my advisor, Professor Judith Tonhauser, for the many hours of help and encouragement she gave me for both the research and the writing of the thesis.

Glosses

3	3 rd person
3.pron	3 rd person pronoun <i>ha'e</i>
A1sg	1 st person singular set A crossreference marker
A2sg	2 nd person singular set A crossreference marker
A3	3 rd person set A crossreference marker
ACC	Accusative case
Adj	Adjective
AT	cotemporaneity connector
B1sg	1 st person singular set B crossreference marker
CAUS1	causative prefix <i>mbo-</i>
CF	counterfactual modality marker
COMPL	completive aspect marker
DATIVE	dative case
DIM	diminutive suffix
EMPH	emphatic marker
GUI	causative marker/ablative case
JE	reflexive and middle prefix
LA	definiteness/cleft/noun phrase marker (borrowed from Spanish)
MA	perfect aspect suffix
N	Noun
NEG...NEG	clausal negation circumfix
NOM	nominative case
O	Object
PASSIVE	passive marker
PAST	past tense marker
PE	marker of non-A arguments and spatiotemporal locations
PL	plural marker
PURP	purposive/benefactive marker
QU	interrogative marker
RC	relative clause marker
RE	“in”/ “for” postposition
S	Subject
SAY	reportative evidential
TOP	topic

(most borrowed from: Tonhauser 2006)

Table of Contents

Title Page	1
Abstract	2
Acknowledgments	3
Glosses.....	4
ToC.....	5
§1 Introduction.....	6
§2 Previous Research and Motivation	8
§2.1 Motivation.....	8
§2.2 Definitions of Basic Word Order.....	10
§2.3 Problems with Determining a Basic Word Order.....	14
§2.4 Factors Influencing Word Order.....	15
§2.5 Previous Research on Guaraní Word Order.....	19
§3 Methodology	21
§3.1 Introduction to Guaraní.....	21
§3.2 The Corpus.....	22
§3.3 Annotation Schema.....	24
§3.4 Coding Decisions.....	27
§4 Results	30
§4.1 The Realization of Subjects and Objects.....	30
§4.2 Which Factors Affect Pre-/Post-Verbal Realization?.....	31
§4.3 Realization of Arguments of Transitive and Intransitive Verbs.....	33
§4.4 Pairing Features.....	35
§4.5 Does Guaraní Have a Basic Word Order?.....	38
§4.6 Relating to Previous Research on Guaraní.....	40
§5 Conclusion and Future Research.....	41
References	43

1.0 Introduction

This is a study of the word order of Paraguayan Guaraní, a Tupí-Guaraní language spoken by about 4 million people in Paraguay and the surrounding areas. The study is based on a corpus of seven texts consisting of approximately 2,000 Guaraní words. I consider naturally occurring data rather than elicited data because it is likely to be more representative of how the language actually functions. Research is done on word order because in addition to the intrinsic scientific knowledge that can be gained, finding the basic word order has been found to correlate with other features of the language, like placement of the adjectives, adpositions and subordinate clauses. I am interested in Guaraní word order specifically because Paraguayan Guaraní is an underrepresented language. This means that little is known about it, including about its word order, and for the first time there is a corpus that allows for this kind of study.

My research questions ask if there is a basic word order in Paraguayan Guaraní, and if and how grammatical function, animacy and discourse status (in the guise of definiteness) affect the placement of subjects and objects. I look at these three factors because they have been shown in other languages to have an impact on the way in which word order is realized. Basic word order can be defined as the least marked order in the language, the order used to disambiguate, or the most frequent order. There can be some difficulties, however, when investigating the basic order, because some languages may not have an order that is sufficiently more frequent than the others to be called the basic order or there may be conflicting results from the different criteria. In my study, I have found that there does seem to be a basic word order in Guaraní: Subject-Verb-Object. I reached this conclusion on the basis of both the frequency and disambiguation criteria,

which are discussed in §2.2. With regard to the second research question, it seems that grammatical function, animacy, and discourse status all have an effect on Guaraní word order.

The thesis continues with the following sections. In §2, I discuss the previous research done on the subject and the motivation for this study, specifically how basic word order has been defined, problems with and factors influencing word order, my motivation for this study, and previous research on Guaraní languages. In §3, I give a background of Paraguayan Guaraní, discuss the corpus, describe the annotation schema used on the corpus, and some coding decisions that I made. In §4 I present my results, and in §5 I discuss the results, implications, and future research.

2.0 Previous Research and Motivation

This section discusses the motivation for this study in §2.1. §2.2 describes the three main ways to define a basic word order. §2.3 states some of the problems with determining the basic word order. §2.4 discusses the three factors of grammatical function, animacy and definiteness that affect word order, and §2.5 reviews previous research on Guaraní.

2.1 Motivation

The study of word order is important because it can make indications about other features of the language, such as the placement of adjectives, adverbs and subordinators. When looking at word order, we look at the placement of subjects and objects in relation to the verb. For this study, I refer to the argument which corresponds to Dowty's (1991)

proto-agent as the subject; essentially, they cause “an event or change of state in another participant” (Dowty 1991, p. 572). I refer to the argument which corresponds to his proto-patient as the object; they undergo a change in state or are “causally affected by another participant” (Dowty 1991, p. 572). Looking at the placement of subjects and objects leads to further information about the language and can be used to classify it typologically.

Greenberg (1966) discusses 45 “universal” tendencies relating to word order that carry across all languages. These universals relate to basic word order typologies, syntax, and to morphology. Some of Greenberg’s universals that are relevant to the present study are:

- Universal 1: In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.
- Universal 3: Languages with dominant VSO order are always prepositional.
- Universal 14: In conditional statements, the conditional clause precedes the conclusion as the normal order in all languages
- Universal 17: With overwhelmingly more than chance frequency, languages with dominant order VSO have the adjective after the noun.
- Universal 27: If a language is exclusively suffixing, it is post-positional;; if it is exclusively prefixing it is prepositional.
- Universal 41: If in a language the verb follows both the nominal subject and nominal object as the dominant order, the language almost always has a case system. (Greenberg 1966, pp. 77, 78, 84, 85, 93, 96)

These universals were pioneering and influential in the study of word order because they allow us to make inferences about the properties of a language based on its word order.

There are six total possible permutations of subject, verb and object: SOV, SVO, OVS, OSV, VSO and VOS. These six permutations can really be condensed into two main categories, OV (SOV, OVS, OSV) and VO (SVO, VSO, VOS), because the orders in each category follow the same general trends for features like placement of adpositions, auxiliary verbs, adjectives, etc. (Dryer n.d.). Verb-initial languages pattern

with SVO languages in general, except for the order of the genitive and the noun: verb-initial languages are usually noun-genitive, while SVO languages can either be noun-genitive or genitive-noun. Other than this single feature, verb-initial languages pattern identically with SVO, allowing them to be combined into a single category of VO. The same can be said for OVS and OSV languages collapsing into a category of VO with SVO languages (Greenberg 1966; Dryer n.d.). In comparing VO languages with OV languages, we can easily see the relationship between word order and these other features. These are patterns and tendencies, rather than absolute rules, but they are accurate enough to help identify patterns in a language by its basic word order. Dryer (n.d.) gives an overview of many of these correlations as the following:

OV	VO
postpositions	prepositions
genitive – noun	noun – genitive
manner adverb – verb	verb – manner adverb
standard – marker	marker – standard
standard – adjective	adjective – standard
final adverbial subordinator	initial adverbial subordinator
adpositional phrase – verb	verb – adpositional phrase
main verb – auxiliary verb	auxiliary verb – main verb
predicate – copula	copula – predicate
final question particle	initial question particle
final complementizer	initial complementizer
noun – article	article – noun
subordinate clause – main clause	main clause – subordinate clause
relative clause – noun	noun – relative clause
noun – plural word	plural word – noun (Dryer p. 52)

2.2 Definitions of Basic Word Order

When studying of basic word order, researchers restrict their focus to the simplest clauses (Brody 1984). Here simplicity has two components to which a basic word order (BWO) clause must adhere: being a main clause, and being declarative. Word order can

often be more rigid in subordinate clauses and nondeclarative clauses like interrogatives or imperatives. Brody (1984) cites Steele (1978) as arguing:

that this restriction of the corpus [to main declarative clauses, EC] gives it a desirable ‘coherence’ with regard to order phenomena by eliminating certain characteristic variations, since, in many languages, subordinate clauses ‘not uncommonly exhibit more rigid word order than main clauses’, and nondeclarative sentences ‘have certain word order requirements which are not found in declarative sentences’ (p. 713).

Because of these word order restrictions that can be found in subordinate and nondeclarative sentences, main declarative clauses are considered simpler, i.e. more basic or fundamental. Brody (1984) states that “The general criterion of simplicity for BWO sentences appears, for most writers, to have two components: that the BWO sentence be (1) in a main clause, and (2) declarative” (p. 713). In order to adhere to this condition I disregarded subordinate clauses as well as interrogatives and imperatives in my study of Guaraní word order.

Basic word order has been defined in many different ways, according to sets of criteria like markedness, disambiguation and frequency, each discussed here. The first way to define basic word order is identifying it as the least marked order (Brody 1984; Dryer n.d.). There are several respects in which an order may be marked, including phonologically, morphologically and syntactically, and semantically, stylistically, and pragmatically. Phonological marking includes special intonational stress or emphasis, while the requirements for being morphologically and syntactically unmarked are less clear. Some studies require that the nouns and verbs of the BWO are the least marked for

definiteness and plurality for nouns, and voice, aspect and mood for verbs. Other studies may call for the least syntactically marked sentences to be able to occur in the largest variety of syntactic contexts in the language. Having the greatest distribution means that these least marked sentences are not restricting; the marking that they have is compatible with and can occur in a wide range of contexts. Different authors utilize different terms with regard to markedness, but Brody (1984) claims that there is overlap of the concepts of semantic, stylistic, and pragmatic marking. This type of least marked sentence consists of sentences that have the least restrictions on where they may be used and do not indicate topicalization, emphasis, or a change in focus (Brody 1984). Examples of all three types of marking are seen in (1).

- (1) a. I LIKE cake.
 b. Nego-ga sakana-o tabeta
 cat-NOM fish-ACC eat:PAST
 ‘The cat ate the fish’ (Tallerman 2005, p. 191)
 c. Sakana-ga neko-ni tabe-rare-ta.
 fish-NOM cat-DATIVE eat-PASSIVE-PAST
 ‘The fish was eaten by the cat.’ (Tallerman 2005, p. 191)
 d. John was seen by Mary.

(1a) is an example of phonological marking from English. The typical English sentence that comes out of the blue, with no prior context, has a pitch accent on the last content word. (1a) is marked because the pitch accent, here indicated with small caps, comes on the word “like”. (1c) is an example of the passive construction in Japanese. This is morphological marking because the passive morpheme, *rare*, gives the verb greater morphological complexity than is found in the active voice construction in (1b). Syntactic marking is found in (1d), another example from English. Constructions in the passive voice in English are marked in that the sentence has an auxiliary verb and

participle, “was seen”, instead of a finite main verb, “saw”. The passive construction is also marked because the semantic argument is realized by an adjunct, “by Mary”.

Some studies like Dixon (1972) and Pullum (1977) have defined basic word order as the order required to avoid ambiguity. Consider the following example: Russian has noun classes in which the nominative and accusative cases have the same endings. While both SVO and OVS orders are possible in the language, Russian disallows the OVS order in sentences with nouns from the classes with identical case endings (Brody 1984).

- (2) a. Mat' ljubit doč
 mother-NOM loves daughter-ACC
 ‘The mother loves the daughter’
- b. Doč ljubit mat'
 Daughter-NOM loves mother-ACC
 ‘The daughter loves the mother’
- c. *Doč ljubit mat'
 Daughter-ACC-TOP loves mother-NOM
 *‘The mother loves the daughter’ (Bloom 1999, p. 64)

Examples (2a, b) show that ‘mother’ and ‘daughter’ are two of the nouns that have identical forms in the nominative and accusative cases. Because of the ensuing ambiguity, SVO is the required order and the OVS order of (2c) is not allowed.

According to this criterion, SVO is then considered the BWO because it is the order that disambiguates between the nominative and accusative nouns. Another example comes from Latin and the Romance languages. Classical Latin had a relatively free word order without ambiguity, which was possible because a well-developed case system. When the language evolved to the point of losing the distinctive case endings, however, ambiguity began to be a factor. Since many of the nouns began to have the same endings, prepositions and a set word order were required in order to understand phrases like *the*

mother of John slept, where the subject and possessor could previously have been identified in any order because of the case marking on each noun. The basic word order of this evolving language that sprang from avoiding ambiguity can be seen in the modern Romance languages: SVO (Marazzini 2002).

Third, basic word order can be defined as the most frequent word order in a language (Dryer n.d.; Mithun 1992; Brody 1984). As Dryer (n.d.) says, “Where languages allow alternative orders, one order is often overwhelmingly more frequent” (p. 10). This is obvious in English, where an order like OSV is possible, as in *John, she hated*, though it is vastly less frequent than the SVO order: *She hated John*. When an order is the most frequent in a language, it often falls into the categories delineated by the above definitions of basic word order, i.e. the most frequent order is often the least marked and avoids ambiguity. Brody (1984) says that “if a sentence type is excluded as a BWO candidate because it occurs only in certain marked discourse or pragmatic situations, that sentence type will probably occur with low frequency in the language anyway” (p. 717). In other words, when defining the BWO as the most frequent word order in the language, the definition in reality includes other factors, like being the least marked or disambiguating. In §4.5, I examine whether Paraguayan Guaraní has a BWO using each of the three criteria discussed here.

2.3 Problems with Determining a Basic Word Order

Determining the basic word order of a language can be difficult when considering languages with a seemingly free word order. Establishing the BWO of a language like English is fairly simple; SVO is by far the most common order, and other orders tend to

be fairly marked by intonation. Conversely, there are languages in which all possible constituent orders occur with enough regularity that declaring an order the most frequent and therefore the basic word order is difficult. There may be an order that is more frequent than the others, but as it is not significantly more frequent than the other orders there is no reason to pronounce it the basic word order (Mithun 1992). Some studies have claimed that a language may either have more than one BWO or have a word order that is affected not by syntactic categories but instead by pragmatic conditions or other features like animacy or definiteness (Brody 1984). These issues have a significant effect on studies of word order in languages like Guaraní. There may not be a word order that is sufficiently more frequent than the others to be called the basic order, or there may be conflicting results from the application of different criteria of basic word order. For my study, this means that I may or may not be able to reach a definite conclusion with the use of the previously discussed criteria.

2.4 Factors Influencing Word Order

The three main factors that have been identified in the literature to determine word order are grammatical function, animacy and definiteness (Brody 1984; Mithun 1992; Rude 1992). I discuss each in turn. Tojolobal, a Mayan language, the animacy hierarchy constrains possible word orders. The typical animacy hierarchy considers humans to be more animate than other animate entities such as animals, with inanimate entities lowest on the hierarchy (eg. Silverstein 1976):

- (3) Humans > (non-human) animate > inanimate

Brody (1984), in her discussion of Tojolobal, makes use of a slightly different animacy hierarchy, which additionally distinguishes humans referred to by proper names and independent pronouns from other humans:

- (4) 1-proper names and independent pronouns
 > 2-other humans
 > 3-animate
 > 4-inanimate (Brody 1984, p. 720)

This hierarchy, contrary to the more typical animacy hierarchy, mixes reference form and the animacy of the referent. Humans referred to by proper names and pronouns are level 1 and are considered more animate than humans referred to by other kinds of nominal expressions, level 2. Animate entities, level 3, are lower than humans, and inanimate entities, level 4, are again lower than animate entities.

Of the six logically possible word orders, all are possible in Tojolobal, but SOV, VSO, OSV, and OVS are only allowed when the subject is level 1 or 2 and the object is level 3 or 4 on the animacy hierarchy (cf. the SOV strings *The man the rock kicked*, *John the dog scratched*, etc.). In (5a), we can see that the OVS order requires the object to be of lower animacy as well as focused.

- (5) a. miyuk chenek' wa s-chon-o-Ø ja migel-i
 NO BEAN pro 3e-SELL-tvm-3a det MIGUEL-term
 'No, Miguel sells BEANS.'
 b. ja Roberto-i Ø-s-mak'-a-Ø ja julio-i
 det ROBERTO-term com-3e-HIT-tvm-3a det JULIO-term
 'Roberto, he hit Julio.' (Brody 1984, p. 722)

In Tojolobal, the SVO seen in (5b) differs from the others because it can have a subject of level 1, 2, or 3 and an object of equal or lower animacy (cf. the strings *Mary hugged Joe*, *The dog licked the cake*, *The snake ate the mouse*, etc.). We can consider these word orders marked because they are acceptable only with certain features, here, the animacy

of the referents of the subject and object. Since the SVO order can have a subject and object of equal animacy levels instead of requiring the object to be of lower animacy, it is a less marked order than the previous four. The VOS of Tojolobal is overall the least marked order of the language and could therefore be considered the most basic word order of Tojolobal. A problem arises, however from the fact that while VOS is the least marked order, SVO is the most frequent order, which hence could also be considered a basic word order (Brody 1984).

Discourse status, which I measure by measuring definiteness or givenness, is another factor that can influence word order (Heim 1982; Mithun 1992). A definite noun phrase, like *the dog*, is a phrase in which the referent of the noun phrase has previously been identified and is known by both the speaker and the hearer. An indefinite phrase, like *a cat*, refers to an individual who has not yet been mentioned and is new information to the hearer. Word order determined by definiteness in languages like Mandarin and Russian, noun phrases that realize old information precede noun phrases that indicate new information, as seen in the following Russian example.

- (6) a. Èto plat'e šila Inna
 this dress-TOP sewed Inna
 'This dress-TOP Inna sewed.'
- b. Staruju lodku my prodali
 old boat-TOP we sold
 'We-TOP sold the old boat-TOP.'
- (Bloom 1999, p. 13)

In (6) we can see that the old or known information, marked by TOP 'topic', precedes all of the other elements in the sentence. Bloom (1999) also notes that "the topic position in Russian is not restricted to arguments which bear a particular grammatical role to the predicate; subjects, objects, and more oblique arguments can all appear in the topic position" (p. 14).

There are other languages that have the opposite ordering: Cayuga, an Iroquoian language spoken in Ontario, Ngandi, an Australian aboriginal language, and Coos, an Oregon language all place indefinite noun phrases before definite, and so new information before old information, as we can see in (7) in Cayuga.

- (7) a. Katsihwá' k̄hsa:s.
 hammer I.seek
 'I am looking for a hammer.'
 b. To: ti' nika:no.:' n̄:kȳ katsihwa'?
 how then so.it.costs this hammer
 'How much does this hammer cost?' (Mithun 1992, p. 28)

The hammer is indefinite and new information in (7a), and therefore at the beginning. In (7b), 'this hammer' is definite, and so it comes at the end of the sentence.

Grammatical function has also been shown to play a role in word order, especially in English. The word order of English encodes which argument is the subject and which is the object; the basic word order is SVO, so the pre-verbal argument is the subject and the post-verbal argument is the object.

- (8) S V O
 a. A boy saw the man.
 b. The man saw a boy.
 c. A stone hit the man.
 d. The dog licked the man.

All of the sentences in (8) are possible SVO sentences, and they also show that definiteness and animacy play comparatively less of a role in English word order. (8a) would be impossible in languages like Russian in which the old information/definite noun phrase is supposed to come before the indefinite, while (8b) would be ruled out in Cayuga, where the definite noun phrase is supposed to come after the indefinite. (8c) and (8d) would not be allowable in languages like Tojolobal that have restrictions on word

order due to animacy, such as requiring the subject to be of lower animacy than the object.

In summary, animacy, definiteness/discourse function and grammatical function have all been claimed to constrain word order across languages. In my quantitative study, I explore the extent to which these three factors play a role in Guaraní word order.

2.5 Previous Research on Guaraní Word Order

There has been little research done on word order in Guaraní. Tonhauser (2006) does not claim a BWO for Guaraní, but asserts that the word order is generally free, determined by the discourse context, and that all possible orders are represented in the language (Tonhauser 2006, p. 144).

Dooley (1982) has made a number of claims about the word order of Brazilian (Mbya) Guaraní. He asserts that the unmarked order of the syntactic constituents of Mbya Guaraní is setting-subject-verb-object-adjunct, where the setting is temporal or spatial location and the adjunct refers to the instrument, manner, referential, function or beneficiary. Dooley states that the word order is affected by pragmatic factors: “the elements ... are often arranged in a marked order on pragmatic grounds, with the most informative element given final position”, i.e. given information comes before new information (Dooley 1982, p. 320). Given information is not the most informative element because both interlocutors would already know to whom or what the noun refers. A definite noun phrase like *the dog* is less informative than an indefinite phrase like *a dog* because when discussing *the dog* both speakers know that there is a dog and to which dog the speaker is referring. In contrast, the phrase *a dog* is more informative because

introduces the idea that there is a dog. Dooley would therefore likely predict that definite noun phrases would precede indefinite phrases. He also states that “in general, the verb does not occupy clause-initial position when there are three or more clause elements” (Dooley 1982, p. 322). This claim suggests that Guaraní is unlikely to realize sentences as VSO or VOS. In sum, Dooley (1982) makes three claims about the word order of Mbya Guaraní, which I will examine for Paraguayan Guaraní:

- (9) 1. The unmarked word order is SVO.
2. Old/given information precedes new/unknown information.
3. Verb-initial clauses with three or more elements are generally not realized. (pp. 321-322)

That the basic order of Paraguayan Guaraní is SVO is also claimed by Gregores and Suárez (1961). They state that the “preferred” order is the most frequent word order: Subject-Verb-Indirect Object-Object-Adverbial attribute (p. 305). Gregores & Suárez (1961) also say that of all the elements, the position of the indirect object is the most inflexible: “it occurs either immediately after or immediately before the verbal phrase” (p. 305). Gregores and Suárez (1961) preface their claims, however, with the admission that they “are to be understood therefore only as very rough approximations, based on impressionistic evaluations of what is more frequent” (p. 304). Though their claims are impressionistic, I use a quantitative study to examine if there is a word order that can be considered the most frequent or the “preferred” order, as well as if indirect objects always appear immediately before or after the verb.

3.0 Methodology

This section consists of an introduction to Paraguayan Guaraní in §3.1, information about the corpus that I compiled and worked with in §3.2, information about the annotation schema in §3.3, and decisions I made during the coding process in §3.4.

3.1 Introduction to Guaraní

Paraguayan Guaraní is a Tupí-Guaraní language with approximately 4.65 million speakers in Paraguay and an additional 200,000 in the surrounding countries. Along with Spanish, Paraguayan Guaraní is an official language of Paraguay. According to the Ethnologue, approximately 52% of rural Paraguayans are monolingual in Guaraní, but around 90% of the people in and around the capital of Paraguay, Asunción, speak a mixture of Spanish and Guaraní called Jopará (Gordon 2005).

Paraguayan Guaraní can be characterized as a head-final language, contrasting with English, a head-initial language. In English, we can see that the head of a phrase precedes its argument:

- (10) a. The dog *bit* the man.
 b. The dog bit the man *in* the knee.
 c. The dog bit the man *because* it was hungry.

In (10a) the verb *bit* precedes its argument, the noun phrase *the man*. The preposition *in* is the head in (10b), preceding the argument of *the knee*, and finally in (10c), the subordinator *because* precedes its argument, the sentence *it was hungry*. In Guaraní, however, the head typically follows its argument:

- (11) a. *Mesa ari*
 table on
 ‘on the table’
 b. *O-ja-peve*
 they-hatch-until
 ‘until they hatch’

We can see in (11a) that the preposition *ari* ‘on’ follows its argument, the noun *mesa* ‘table’, and in (11b) the subordinator *peve* ‘until’ follows the sentence *o-ja* ‘they hatch’. Because of these correlations we would expect Guaraní to be a verb-final language since it has been demonstrated to be head-final in (11). The correlation is not always perfect, however, as we can see from the adjective phrases in (12). There are different types of headedness, such as syntactic headedness and semantic headedness. In (12a), the English example, the adjective *angry* is the semantic head of the phrase, but not the syntactic head. Since *angry* precedes its argument of *dog*, this sentence is head-final, differing from the other English constructions. In the same vein, (12b) shows that *vai* ‘ugly’ is not the syntactic head. Since *vai* ‘ugly’ follows its argument, *ko mba’e* ‘this thing’, this phrase is head-initial, unlike the rest of the Guaraní constructions.

- (12) a. The *angry dog* bit the man.
 b. *Ko mba’e vai*
 this thing ugly
 ‘this ugly thing’

I would still expect Guaraní to be verb-final because in general it is a head-final language, but counter-examples do exist.

3.2 The Corpus

The corpus consists of folk tales and personal narratives that were gathered by Judith Tonhauser in Paraguay. Together, the stories comprise approximately 2,000

Guaraní words. Table 1 gives an overview of the texts in the corpus and the number of words in each text.

Table 1: Overview of the texts in the corpus of Paraguayan Guaraní (Tonhauser 2006, p. 126)

Text	Description	Words
<i>michi</i> ‘small’	narrative about childhood	283
<i>ñepyru</i> ‘beginning’	narrative about parents’ life	351
<i>A boy, a dog and a frog</i>	narrated by SC	412
<i>A boy, a dog and a frog</i>	narrated by NC	247
<i>jakare</i> ‘crocodile’	story about the life of crocodiles (author unknown)	143
<i>kirikiri</i> ‘cricket’	story about a cricket’s adventure (author unknown)	196
<i>ka’i</i> ‘monkey’	story about a monkey’s adventure (author unknown)	375
<i>ypei</i> ‘duck’	story about a frog and a duck (author unknown)	82
total		2089

It was important to use data from a corpus because in a corpus the data is naturally occurring. Language does not occur in a vacuum, and if one tried to determine the basic word order from sentences elicited from a consultant, then all of the data would be missing the surrounding context. This context can be influential on word order, and without it, the data might not be representative of how Guaraní actually functions.

I transcribed the corpus using the glosses developed in Tonhauser (2006) and edited for consistency those which had already been glossed. It is also clear that the number of Guaraní words is significantly less than the number of English words would be. Guaraní is an agglutinating language, i.e. it regularly attaches affixes to the base word, and mildly polysynthetic, i.e. it has many morphemes (smallest units of language with meaning) within a single word. This means that there can be a lot of information transmitted in what is counted as a single word. One example is the following, in which the concepts of purpose, comparison and the preposition meaning ‘after’ are all joined to the base word: the verb *karu* ‘eat’.

- (13) *karu-rire-guá-icha*
 eat-after-PURP-like
 ‘as if for dessert’

Thus, (13) is counted as one word in the Guaraní corpus, but would be four words in an English corpus.

3.3 Annotation Schema

In order to study Guaraní word order on the basis of the corpus, I developed an annotation schema to annotate the main declarative clauses with word order information. I annotated the 390 main declarative clauses for the three features that have been claimed to play a role in word order across languages, as discussed in §2.4: grammatical function, animacy, and definiteness, which is claimed to reflect discourse status. Transitivity was also annotated to be able to separate transitive and intransitive sentences. I also annotated copula constructions separately from the other types of clauses. A copula links the subject of a sentence with its predicate, essentially equating the two. An example of a copula in English is *John is a doctor*. In this sentence, *John* is the subject, *a doctor* is the predicate, and *is* is the copula verb linking the two. Guaraní has copula constructions as well, but they do not occur with a copula verb, as in Example (14). To annotate these clauses, I marked the definiteness of the subject and of the predicate (definite (DN), indefinite (IN), or unmarked (N)) and separated the two with **cp** for copula.

- (14) *Jakare peteĩ mymba*
 crocodile one animal
 N- cp- IN
 ‘The crocodile is an animal’

There were 12 copula constructions in the corpus, but since they are not the type of sentence construction that I am interested in, I ignore these 12 clauses for the rest of the study.

Table 2 shows the four features for which I annotated the corpus, as well as the abbreviations that were used in the annotation.

Table 2: Annotated Features

Feature	Overview
grammatical function	Subject (S), Object (O), Verb (V)
transitivity	Transitive (t), Intransitive (i)
animacy	Human (h), Animate (non-human) (a), Inanimate (n)
definiteness	Definite (DN), Indefinite (IN), Unmarked (N)

In what follows, I illustrate each of the annotated features separately with an example from the corpus. For grammatical function, I labeled constituents as being a subject (S), verb (V) or an object, which was further divided into four categories. The first object category consisted of quotes and complement clauses (Oq). Examples (15) and (16) show a transitive verb followed by an Oq, a quoted object in (15) and a complement clause in (16).

- (15) o-sapukái: **“A-vy’a-iterei!”**
 A3-shout: **A1sg-happy-very**
 Vt- **Oq**
 ‘He shouted: **“I’m very happy!”**’
- (16) Nd-oi-kuaa-i **mba’e-pa la oi-ko-va**
 NEG-A3-know-NEG **thing-QU LA A3-pass-RC**
 Vt- **Oq**
 ‘He didn’t know **what was going on.**’

The second category was verbal objects (Ov), and in example (17) we see a transitive verb followed by a verbal object, *o-purahéi* ‘to sing’:

- (17) o-ñepyrũ **o-purahéi**
 A3-begin **A3-sing**
 Vt- **Ov**
 ‘He began **to sing.**’

The third category contains the indirect objects (O2):

- (18) *tapia* o-japo **hese** ko’ẽ porã-mboyve
 always A3-do **to.3** dawn pretty-before
 Vt- **O2hDN**
 ‘He always did it **to her** before every dawn.’

Example (18) shows the indirect object ‘to him/her’ following the transitive verb *o-japo* ‘he did’. The final category covers the regular direct objects (O), and we can see in the following example the direct object, *yvotytỹ-re* ‘garden’, following the transitive verb *re-ñangareko* ‘you take care of’:

- (19) re-ñangareko **yvotytỹ-re**
 A2sg-take.care.of **garden-RE**
 Vt- **OnN**
 ‘You take care of **the garden.**’

I also annotated for transitivity; the subject or verb was either transitive (t), “She hit X”, or intransitive (i), “He slept”. As mentioned above, transitivity was encoded to enable me to separate the transitive clauses from the intransitive. The third feature that I annotated was animacy. The three choices for animacy were human (h), as seen in example (20), non-human animate (a), as in example (21), and inanimate (n), as in example (22):

- (20) I-memby o-heja
 3-child A3-leave
 OhDN- Vt
 ‘She left her child’
- (21) Kururu nd-o-vy’a-i-ete
 frog NEG-A3-happy-NEG-very
 SiaN- Vi
 ‘The frog was not very happy’

- (22) Kuarahy o-mimbi
 sun A3-shine
 SinN- Vi
 ‘The sun was shining’

The final feature that I annotated was definiteness. The subject or object could be definite (DN), as in example (23), indefinite (IN), as in example (24), or unmarked (N), as in example (25).

- (23) Ha’e o-pu’a voi-eterei
 3.pron A3-get.up early-very
 SihDN- Vi
 ‘She got up very early’

- (24) o-topa peteĩ itã
 A3-find one stone
 Vt- OnIN
 ‘He found a stone’

- (25) ju’i mombyry o-po
 frog far A3-jump
 SiaN- Vi
 ‘The frog jumped far’

Included in the definite category were phrases with possessives, demonstratives, the article *la* borrowed from Spanish, pronouns and proper names. The indefinite noun phrases were indicated by the use of *peteĩ* ‘one’. In Guaraní, nouns can also occur without any kind of article; these are the nouns that are counted as unmarked. Annotating the definiteness of the subjects and objects allowed me to determine if discourse status plays a role in Guaraní word order.

3.4 Coding Decisions

At various times during the annotation process, a coding decision had to be made. They are recorded here so that future research can be adequately compared:

- Spatial and temporal locative phrases were treated as adjuncts, and hence not coded.

Disregarding the locative phrases means that verbs like ‘go’, ‘leave’, and ‘climb’ are all considered to be intransitive, as in example (26).

- (26) ha o-jupi la yvyra-ñeno-ari
 and A3-climb LA tree-lie-on
 Vi
 ‘And he climbed onto the tree trunk’

An exception to this decision is the necessary locative with verbs like ‘put’, which were coded as indirect objects:

- (26) O-moĩ ka’i i-jyva-guý-pe
 A3-put monkey 3-arm-under-PE
 Vt- OaN- O2nDN
 ‘She put the monkey under her arm’

- There were many sentences in the corpus with one explicit subject and two or more verbs. A distinction was made between them depending on the presence of the word *ha* ‘and’. If the sentence was of the form ‘S V and V’, I annotated it as S-V, V, as in example (27), rather than including the subject with both verbs and annotating it as S-V, S-V.

- (27) Umi jakare-ra’y ... i-sy o-gueraha ha o-mbo-jepoi ý-pe
 those crocodile-young 3-mother A3-take and A3-CAUS1-thrown water-PE
 OaDN- StaDN- Vt Vt
 ‘Their mother takes the crocodile young and throws them in the water.’

I chose to annotate this type of sentences in this manner because it could be considered sentence coordination in which the subject of the second sentence is elided. It is possible that the underlying structure of the sentence contains a verb phrase that has two verbs, but since we cannot tell from the surface structure of the phrase whether it is sentence or verb phrase coordination, I picked sentence coordination as the more

general case and so annotated the surface string of the form, ‘S V and V’, as S-V, V, as in example (27). Guaraní has constructions known as serial verbs, which differ from the ‘S V and V’ discussed above. These constructions occur without *ha* ‘and’ as ‘S V V’. (28) shows the serial construction with the verb ‘jump’ occurring twice, but without a conjunction between them.

- (28) Kururu o-po o-po vy’á-gui ha o-sapukái
 frog A3-jump A3-jump happy-GUI and A3-shout
 SiaN- Vi SiaN-Vi
 ‘The frog happily jumped and jumped and shouted...’

The two verbs can range from designating a single event to one of the verbs modifying the other (Velázquez-Castillo 2004). Here the sentence is not broken up by the word *ha* ‘and’, cannot clearly be interpreted as two sentences syntactically, refers to the same event, and the subject is the subject of both verbs. I annotate serial verb constructions as S-V, S-V. I could have created a new annotation for the serial verbs, but that would break up the data into even smaller categories, so I simply count the subject for both verbs, as in example (28).

- There is another construction similar to the serial verb construction that appears with the verb ‘go’. Verbs of motion are frequently grammaticalized into aspectual markers, indicating, for example, the future in English: ‘I **am going** to eat at 6:00’. For the Guaraní examples in which ‘go’ appears next to another verb, I counted ‘go’ as a verb of motion unless it contained the marker *-vo*, which is when it clearly seems to be a progressive aspect marker rather than a motion verb (Tonhauser 2006).

- (29) a-ha a-je-po-reka algun mymba
 A1sg-go A1sg-JE-hand-have some animal
 Vi Vt- OaIN
 ‘I went and caught some animal’

clauses; it separates the clauses by transitivity of the verb and number of arguments realized.

Table 3: Argument realization in the Corpus

	Intransitive	Transitive	Copula
No arguments realized	121	11	---
Subject only realized	80	3	---
Object only realized	---	112	---
Both S and O realized	---	51	---
Total clauses	201	177	12

Of the 390 clauses, 201 occur with an intransitive verb, 177 with a transitive verb, and 12 are copula constructions. We find that the majority of intransitive verb clauses occur with no arguments, and the majority of transitive verbs occur with only one argument. The main result here is that the subject and object do not need to be realized regardless of the transitivity of the verb.

4.2 Which Factors Affect Pre-/Post-Verbal Realization?

Apart from looking at argument realization for transitive and intransitive verbs, I look at the effects that grammatical function, animacy and definiteness have on the realization of arguments. Table 4 shows the realization of arguments, divided by grammatical function. From here on, the objects discussed are the direct objects only; I do not include the verbal objects, quote objects or indirect objects because they may behave differently than “regular” objects and are also not considered in other word order studies (cf. §2.2).

Table 4: Effect of Grammatical Function on Argument Placement

	Subject	Object	Subject + Object
Pre-Verbal	77	5	82
Post-Verbal	57	78	135
Total	134	83	217

From this table we can see that there is a tendency for subjects to be realized pre-verbally, 77 pre-verbal versus 57 post-verbal, but only a slight tendency. The trend is much stronger with objects; about 94% of the objects occurs post-verbally. Grammatical function, specifically object status, does in fact affect noun phrase placement.

Table 5 looks at the placement of arguments according to their animacy.

Table 5: Effect of Animacy on Argument Placement

	Human	Animate	Hum. + Anim.	Inanimate
Pre-Verbal	25	47	72	10
Post-Verbal	28	67	95	40
Total	53	114	167	50

There seems to be a slight tendency for animate and somewhat stronger tendency for inanimate arguments to be realized post-verbally, and no real tendency for pre- or post-verbal placement of human arguments. This data does not show any particularly striking effects of animacy on noun phrase placement.

Table 6 addresses the effect of discourse status/definiteness on the realization of subjects and objects.

Table 6: Effect of Discourse Status on Argument Placement

	Definite	Indefinite	Unmarked
Pre-Verbal	44	0	38
Post-Verbal	76	15	44
Total	120	15	82

This table shows an obvious effect in that all indefinite arguments, whether subject or object, occur post-verbally. There also seems to be a tendency for definite arguments to

appear post-verbally, but no real tendency for the unmarked arguments. Definiteness, specifically indefiniteness, does have an influence on the realization of noun phrases.

Though grammatical function and definiteness are already demonstrating influence on noun phrase placement, to get a clearer idea of what is happening, it is necessary to separate the arguments into transitive subjects, intransitive subjects and objects (§4.3). It is also important to pair the features, looking at animacy and definiteness together for the different types of arguments, in order to better examine the data (§4.4).

4.3 Realization of Arguments of Transitive and Intransitive Verbs

I now divide the arguments into subjects of transitive verbs, subjects of intransitive verbs, and objects in order to see if there is a difference between the argument types. Table 7 shows the realization of the intransitive subjects.

Table 7: Placement of Intransitive Subjects

Si	Human	Anim.	H + A	Inanim.	Definite	Indef.	Unmkd	Total
Pre-V	13	29	42	6	25	0	23	48
Post-V	10	19	29	3	22	3	7	32

The intransitive subjects show a general tendency toward pre-verbal realization, with only the indefinite subjects occurring more often post-verbally (0 pre-verbal, 3 post-verbal). The human and definite intransitive subjects are fairly evenly divided between pre- and post-verbal realization, and the strongest trend is in the unmarked subjects, which are more often located before the verb (23 pre-verbal, 7 post-verbal).

Table 8 presents the data on the placement of transitive subjects.

Table 8: Placement of Transitive Subjects

St	Human	Anim.	H + A	Inanim.	Definite	Indef.	Unmkd	Total
Pre-V	10	17	27	2	15	0	14	29
Post-V	6	18	24	1	11	0	14	25

The transitive subjects differ from the intransitive in that there do not appear to be any real tendencies toward pre- or post-verbal realization in any of the categories. In all the categories, the transitive subjects occur roughly just as often pre-verbally as post-verbally: Animate: 17 pre-verbal / 18 post-verbal; Unmarked: 14 pre-verbal / 14 post-verbal, etc.

Table 9 contains the data on the direct objects.

Table 9: Placement of Objects

O	Human	Anim.	H + A	Inanim.	Definite	Indef.	Unmkd	Total
Pre-V	2	1	3	2	4	0	1	5
Post-V	12	30	42	36	43	12	23	78

This table clearly shows the strong tendency that objects have toward post-verbal realization: human objects - 12 post-verbal of 14 clauses, animate objects - 30/31, inanimate objects - 36/38, definite objects - 43/45, indefinite objects - 12/12, unmarked objects 23/24. Since this is an interesting phenomenon, it could be helpful to look at the few actual examples of objects that occur pre-verbally. These are seen in (31).

- (31) a. I-memby o-heja
 3-child A3-leave
 ‘She left her child’
 -human
- b. ko-va-re nda-che-po-pa’ã-mo’ã-i
 this-RC-RE NEG-B1sg-hand-grab-CF-NEG
 ‘I won’t get my hands on this thing’
 -inanimate (demonstrative with animate referent)
- c. ha vokoi-nte Pirulo-pe-nte la o-jura
 and so-only Pirulo-PE-only LA A3-grab
 ‘and so he only grabbed Pirulo’
 -animate

- d. Che-rãi-ko nd-a-guerekó-i a-karu-ha-gua-nte
 B1sg-teeth-EMPH NEG-A1sg-have-NEG A1sg-eat-NOM-PURP-only
 ‘I don’t have these teeth just to eat’
 -inanimate (body part)
- e. yvypora-pe rei-pytyvõ
 mankind-PE A2sg-help
 ‘You help mankind’
 -human (group noun)

When we look at the five specific examples of pre-verbal objects, some interesting results come up. I decided to annotate both demonstratives and body parts as inanimate, because even though their referents or the whole to which they belong may be human or animate, I decided not to place these noun phrases in that category. However, when we look at the five objects in (31), (31a) is human, (31b) refers to a frog (animate), (31c) is animate, (31d) is a body part, and (31e) is a group noun referring to humans. It is also true that none of the subjects in these examples are realized overtly. It is possible that there is a constraint in Guaraní regarding the placement of objects such that objects are only allowed to be pre-verbal if the subject is not realized and the object is or refers to a human or animate entity.

At this level, the categories are broad enough that though they lead to results of their own, they can also miss finer distinctions present in the data. Because of this, I pair the features of animacy and definiteness and look at the results for each argument type.

4.4 Pairing Features

In order to look more closely at the realization of these subjects and objects, I cross-paired animacy and definiteness. Tables 10-13 show the data for intransitive and transitive subjects, all subjects together, and objects. The categories begin with human (h), animate (a), and inanimate (n) definite (DN), followed by human, animate and

inanimate indefinite (IN) and unmarked (N). Table 10 presents the data when the intransitive and transitive subjects are combined.

Table 10: Placement with Paired Features of All Subjects

S	hDN	aDN	nDN	hIN	aIN	nIN	hN	aN	nN
Pre-V	21	14	5	0	0	0	2	32	3
Post-V	9	21	3	2	1	0	5	15	1

Once the features are paired, some interesting results come out. There are differences based on both animacy and definiteness for human and animate definite and unmarked subjects. When looking at the differences in definiteness, human definite are more pre-verbal (21 of 30), but human unmarked are more post-verbal (5 of 7), while animate definite are more post-verbal (21 of 35) and unmarked are more pre-verbal (32 of 47). When considering animacy, human definite are more pre-verbal (21 of 30), while animate definite are more post-verbal (21 of 35), and human unmarked are more post-verbal (5 of 7) while animate unmarked are more pre-verbal (32 of 47). This shows that when the features of animacy and definiteness are paired, there are tendencies that are not seen in the broader categories.

Table 11 begins with the placement of intransitive subjects in view of both animacy and definiteness.

Table 11: Placement with Paired Features of Intransitive Subjects

Si	hDN	aDN	nDN	hIN	aIN	nIN	hN	aN	nN
Pre-V	13	8	4	0	0	0	0	21	2
Post-V	6	14	2	2	1	0	2	4	1

From this table we can see that intransitive subjects show the same tendencies that are present in Table 10; the behavior of just the intransitive subjects is the same as the behavior of the subjects when intransitive and transitive are grouped together. These

placement differences can occur close together in discourse, as we can see in this excerpt from the story called *ka'i* ‘monkey’.

- (32) a. Upe javé-je o-ñe-mboja kuña-karai ha o-sapukái
 that while-SAY A3-JE-approach woman-gentleman and A3-yell
Vi- SihN
 ka'i-pe
 monkey-PE
 ‘Right then, they say, a lady approached and yelled at the monkey’
- b. “Yma-ite-gui-vé-ma che-mbo-py’a-rasy rei-kó-vo.”
 Long.time-only-GUI-more-MA B1sg-CAUS1-stomach-sick A2sg-be-AT
Vt
 ‘“For such a long time now you’ve been bothering me.”’
- c. “Hasy-pe-ve re-‘a-mi che-ñuhá-me nde-teko-ve
 difficult-PE-more A2sg-fall-DIM B1sg-trap-PE B2sg-life-more
Vi
 tie’ỹ”
 unclean
 ‘“Finally you fell into my trap, you lowlife.”’
- d. Ka’i o-ñe-mo-mano-ite ta’anga araity-ári
 monkey A3-JE-CAUS1-die-only figure wax-on
SiaN- Vi
 ‘The monkey pretended to be dead on top of the wax figure.’
- e. Rei re-ñe-mbo-tavy nde, mba’e chavi
 in.vain A2-JE-CAUS1-stupid B2sg thing insignificant
Vi- SiaDN
 ‘“You’re pretending in vain, you little gnat.”’

Within five sentences, there is a post-verbal human unmarked noun phrase, a pre-verbal animate unmarked noun phrase, and a post-verbal animate definite noun phrase. This excerpt exemplifies the data shown in Table 11.

Table 12 shows the data for the subjects of transitive verbs.

Table 12: Placement with Paired Features of Transitive Subjects

St	hDN	aDN	nDN	hIN	aIN	nIN	hN	aN	nN
Pre-V	8	6	1	0	0	0	2	11	1
Post-V	3	7	1	0	0	0	3	11	0

Pairing the features for transitive subjects differs from the results seen in Tables 10-11, in that there are no striking tendencies toward pre- or post-verbal realization. Table 12

shows again the fact that there are no indefinite transitive subjects present in the corpus, which is discussed first in §4.2. The definite human subjects are realized more frequently before the verb, but none of the other pairings has any particular tendencies. The lack of obvious tendencies in the data on transitive subjects is an interesting result in itself. It shows that the transitive subjects differ from the intransitive subjects, and the results in Table 10 are essentially due to the intransitive subjects.

Table 13 presents the results for the direct objects.

Table 13: Placement with Paired Features of Objects

O	hDN	aDN	nDN	hIN	aIN	nIN	hN	aN	nN
Pre-V	1	1	2	0	0	0	1	0	0
Post-V	9	15	19	1	5	6	2	10	11

This table shows again the strong correlation between object status and post-verbal realization seen in Table 9. None of the feature pairings seems to be especially more pre-verbal than the others, so Table 13 does not really show anything not seen in Table 9.

In sum, Table 10 shows trends in the realization of all the subjects that are also present in the intransitive subjects in Table 11. The trends are not present, however, when just the transitive subjects are observed, as in Table 12. Pairing the features for objects, Table 13, does not bring to light any new results.

4.5 Does Guaraní Have a Basic Word Order?

When determining the most frequent word order of a language, previous studies only consider clauses realized with both a subject and an object (Mithun 1992; Brody 1984). In the Guaraní corpus, there were 20 clauses that were realized with both a subject and direct object. Table 14 shows the word orders present in these 20 clauses.

Table 14: *Word Orders Attested in the Corpus*

SVO	SOV	OVS	OSV	VSO	VOS	Total
11	1	0	0	4	4	20

SVO is the most common order found in the corpus, 11 of 20 clauses, with the verb-initial orders being the next most common: VSO - 4, VOS - 4, total - 8. By the frequency criterion, SVO is the basic word order of Paraguayan Guaraní. The problem with this initial determination is that of the 390 clauses in the corpus, there were only 20 such clauses realized. The data in Table 14 and the discovery of SVO as the most frequent order is representative of these 20 clauses, but these 20 clauses are not necessarily representative of the language as a whole. I also need to take into consideration the other definitions of basic word order, least-markedness and disambiguation.

When applied to Paraguayan Guaraní, the results from the definition of least-markedness are not especially helpful. There is no intonational information in the corpus, so possible phonological marking cannot be taken into account for this study. There is also not much of a voice system in Guaraní, so overall, applying the definition of basic word order as the least-marked order is not useful.

The last definition of basic word order is the order used to disambiguate, as in (2), the Russian example in §2.2. SVO is found to be the disambiguating word order in Paraguayan Guaraní, as seen in (33).

- (33) tuju-ry o-jagara-pa la ij-ao
 mud-juice A3-grab-COMPL the 3-cloth
 S- V- O
 ‘The mud got all over his clothing’
 #‘His clothing got all over the mud’

Both the subject and the object in (33) are indefinite third person nouns, namely ‘mud’ and ‘cloth’. The noun is marked for the third person, but there is ambiguity because this

would agree with both of the arguments. The sentence, however, is interpreted as SVO, ‘the mud got all over his clothing’ and not as OVS, ‘his clothing got all over the mud’. This means that the SVO order is the order used to disambiguate between possible word orders, and is the more basic word order.

4.6 Relating to Previous Research on Guaraní

The determination of SVO as the basic word order in Paraguayan Guaraní agrees with the claim in Dooley (1982) that the unmarked order in Brazilian and Paraguayan Guaraní is SVO. He also states that in Mbya Guaraní old/given information precedes new/unknown information. This would suggest that definite noun phrases would tend to be realized pre-verbally while indefinite noun phrases would tend to be post-verbal. This theory is supported by the data in my study to the extent that the indefinite noun phrases, or new information, exclusively occur post-verbally. The old/given information, however, does not seem to show the tendencies claimed by Dooley (1982). Of the 120 clauses with definite noun phrases, 44 of those occur pre-verbally and 76 occur post-verbally. The third claim Dooley (1982) makes is that verb-initial clauses that have at least 3 elements are generally not realized. This is contradicted by my study; the verb-initial orders are in fact the most common orders other than SVO. Of the 20 clauses, the three most common orders are SVO (11), VOS (4) and VSO (4).

Gregores & Suárez (1961) also claim that SVO is the basic word order of Guaraní, though only based on impressions, that the preferred order of Guaraní is SVO, which is in accordance with the findings of my study. They also claim that the position

of the indirect object in Paraguayan Guaraní is inflexible; it occurs either immediately before or immediately after the verb. The corpus contains one counterexample, (34).

- (34) O-moĩ ka'i i-jyva-guý-pe
 A3-put monkey 3-arm-under-PE
 Vt- OaN- O2nDN
 'She put the monkey under her arm'

In (34), the second object is a locative “under her arm,” and is not realized next to the verb, but instead the direct object is realized immediately following the verb, with the indirect object following the direct object.

5 Conclusion and Future Research

In conclusion, we have seen that SVO emerges as the basic word order of Guaraní, based on both frequency and disambiguation. Grammatical function and discourse status (definiteness) have both been shown to have an effect on argument placement. Also, when considered in conjunction, animacy and discourse status have an effect on the placement of subjects.

In future research, there are a number of topics I would like to address. The first is the status of unmarked nouns in Guaraní. At this point, it is unclear how speakers of Paraguayan Guaraní interpret unmarked nouns. The speakers may interpret them as definite or indefinite and that could affect how the unmarked phrases are realized in a sentence. It is possible that since Guaraní does not have a definite determiner, but does have an indefinite determiner, *peteĩ* ‘one’, the unmarked noun phrases are interpreted as definite. A second topic I would like to deal with is focus in Guaraní. English realizes focus by using a different pitch accent or placing the pitch accent on a different word in the sentence, as we can see in (1) in §2.2, or by moving the focused word to the front of

the sentence, as in *CAKE, I like!* I would like to find out how focus is realized in Guaraní.

A third interesting question that I could investigate is the marking on transitive verbs when one argument is realized. It is possible for the verb in Guaraní to be marked either for the subject or the noun, depending on animacy, specifically if either argument is the first or second person? When only one argument is realized with a transitive verb, which argument is more likely to be referenced on the verb?

References

- Beavers, J. (2003). More Head and Less Categories: A New Look at Noun Phrase Structure. In S. Müller (Ed.) *Proceedings of the HPSG03 Conference*. East Lansing, MI: CSLI Publications.
- Bloom, D. B. (1999). Case Syncretism and Word Order Freezing in the Russian Language. Stanford University. Masters Thesis.
- Brody, J. (1984). Some problems with the concept of basic word order. *Linguistics*, pp. 711-736.
- Dooley, R. A. (1982). Options in the Pragmatic Structuring of Guarani Sentences. *Language*, 58 (2), pp. 307-331.
- Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67 (3), pp. 547-619.
- Dryer, M. S. (n.d.) Word Order. University at Buffalo. Manuscript.
- Gordon, R. G., Jr. (Ed.) (2005). *Ethnologue: Languages of the World* (15th ed.). Dallas, TX: SIL International. Online version: <http://www.ethnologue.com/>

- Greenberg, J. H. (1966). Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In J. Greenberg (Ed.) *Universals of Language* (pp. 73-113). Cambridge, MA: The M.I.T. Press.
- Gregores, E. & Suárez, J. A. (1961). *A description of colloquial Guaraní*. Cornell University. Doctoral dissertation.
- Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases*. University of Massachusetts. Doctoral dissertation.
- Marazzini, C. (2002). *La lingua italiana: Profilo storico* (3rd ed.). Bologna, Italy: il Mulino.
- Mithun, M. (1992). Is Basic Word Order Universal? In D. Payne (Ed.) *Pragmatics of word order flexibility* (pp. 15-61). Philadelphia: J. Benjamins Pub. Co.
- Payne, D. L. (1992). Nonidentifiable Information and Pragmatic Order Rules in ‘O’odham. In D. Payne (Ed.) *Pragmatics of word order flexibility* (pp. 137-166). Philadelphia: J. Benjamins Pub. Co.
- Rude, N. (1992). Word Order and Topicality in Nez Perce. In D. Payne (Ed.) *Pragmatics of word order flexibility* (pp. 193-208). Philadelphia: J. Benjamins Pub. Co.

- Silverstein, M. (1976). Hierarchy of features and ergativity. In R. Dixon (Ed.) *Grammatical Categories in Australian Languages*. Canberra: Australian Institute of Aboriginal Studies.
- Tallerman, M. (2005). *Understanding Syntax* (2nd ed.). London: Hodder Arnold.
- Tonhauser, J. (2006). The Temporal Semantics of Noun Phrases: Evidence from Guaraní. Stanford University. Doctoral dissertation.
- Velázquez-Castillo, M. (2004). Serial Verb Constructions in Paraguayan Guaraní. *International Journal of American Linguistics*, 70 (2), pp. 187-213.