Guaraní Direct-Inverse Alignment in Lexicalist Unification Grammars

Petr Homola

phomola@uoc.edu

June 11, 2024

Abstract

Guaraní exhibits direct-inverse alignment, transitive verbs agree with the subject or the object depending on which ranks higher on the person hierarchy. It is shown how such verbs' morpholexical entries encode this type of alignment and how phrase structure rules operate on them.

1 Introduction

Guaraní is a Native American language spoken in Argentina, Bolivia, Brazil and Paraguay by 6–7 million people. It is a compositionally polysynthetic language (according to Mattissen's (2006) classification) with incorporation and rich morphology. Transitive verbs agree with the subject or the object, depending on which ranks highest on the person hierarchy[1] (Estigarribia, 2020).

As a non-configurational language, Guaraní exhibits free word order and implicit arguments. Information structure is expressed by a combination of word order and discourse particles.

2 Guaraní morphology

Noun phrases in Guaraní form one phonological word with case markers at the end of the phrase,[2] for example:

(1) *tape   porã-re*
     road   good-TRANSL
     "on the good road"

Verbs agree with the subject, the object or both, for example:

(2) *Ro-hayhu*
     1SG>2SG-love
     "I love you."

The morpholexical entry of the form *ro-hayhu* thus is

---

[1] 1 > 2 > 3

[2] This complicates the contents of morpholexical entries for adjectives whose CASE attribute refers to the adjective's head.

(3)
$$
\begin{bmatrix}
\text{LEXSEM} & \text{'-ayhu'} \\
\text{SUBJ} & \begin{bmatrix} \text{PERSON} & 1 \\ \text{NUMBER} & \textsc{sg} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PERSON} & 2 \\ \text{NUMBER} & \textsc{sg} \end{bmatrix}
\end{bmatrix}
$$

3 Direct-inverse alignment

Transitive verbs use two sets of person-number prefixes depending on whether they agree with the subject or the object. In the unmarked case, they agree with the subject, for example:

(4) *Ai-nupã   ichupe*
1SG-beat   him/her
"I beat him/her."

If the object outranks the subject on the person hierarchy, the verbs agrees with it:

(5) *Che-nupã   ha'e*
>1SG-beat   (s)he
"(S)he beats me."

Hence the morpholexical entry for *ainupã* is

(6)
$$
\begin{bmatrix}
\text{LEXSEM} & \text{'nupã'} \\
\text{SUBJ} & \begin{bmatrix} \text{PERSON} & 1 \\ \text{NUMBER} & \textsc{sg} \end{bmatrix} \\
\text{TENSE} & \textsc{smpl}
\end{bmatrix}
$$

and the morpholexical entry for *chenupã* is

(7)
$$
\begin{bmatrix}
\text{LEXSEM} & \text{'nupã'} \\
\text{OBJ} & \begin{bmatrix} \text{PERSON} & 1 \\ \text{NUMBER} & \textsc{sg} \end{bmatrix} \\
\text{TENSE} & \textsc{smpl}
\end{bmatrix}
$$

As a special case, for second person objects, both the subject and the object are restricted in the morpholexical entry:

(8) *Ne-nupã   ha'e*
>2SG-beat   (s)he
"(S)he beats you."

(9)
$$
\begin{bmatrix}
\text{LEXSEM} & \text{'nupã'} \\
\text{SUBJ} & \begin{bmatrix} \text{PERSON} & 3 \\ \text{NUMBER} & \textsc{sg} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PERSON} & 2 \\ \text{NUMBER} & \textsc{sg} \end{bmatrix} \\
\text{TENSE} & \textsc{smpl}
\end{bmatrix}
$$

As can be seen, the two types of agreement are straightforward to encode in the lexicon. Since Guaraní is a non-configurational language, the general phrase structure rule for a verb's argument is[3]

(10) S → NP, I

The corresponding unification formula for the subject and object is

(11) $[\text{SUBJ} : \Phi_{\text{NP}}] \sqcup \Phi_{\text{I}} \vee [\text{OBJ} : \Phi_{\text{NP}}] \sqcup \Phi_{\text{I}}$

4 Computational analysis

We have developed an experimental morphological analyser for Paraguayan Guaraní. The ouput of the analyser is a list of morphological entries that are used by the parser we are working on. The output of the parser is a (rather flat) phrase structure tree and a dependency tree. An example is in order:

---

[3]The comma signifies that the order of the constituents is not fixed.

(12) *Ha'e oi-mé-ne   kosiná-me*
    (s)he 3SG-be-POSB   kitchen-LOC
    "(S)he might be in the kitchen."

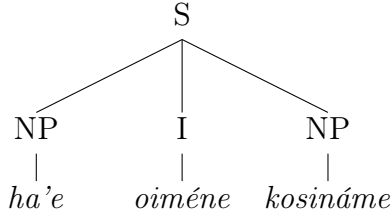The phrase structure tree of (12) is given in Figure 1.

Figure 1: Phrase structure tree of (12).

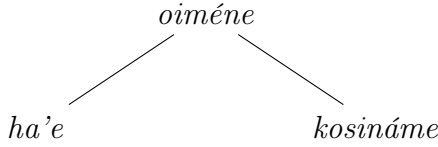The dependency tree of (12) is given in Figure 2.

Figure 2: Dependency tree of (12).

As can be seen, the parser respects lexical integrity. For example, *oiméne* is represented by one node in the trees, although it means "might be" or "possibly is". On the other hand, in (13) the verb phrase *oho kuri* "(s)he went" would be represented by two nodes in the trees, although *kuri* is a function word.

(13) *Che sy   o-ho   kuri*
    my   mother 3SG-go PST
    eskuéla-pe
    school-LOC
    "My mother went to school."

In the corresponding feature structure, though, *kuri* would not be represented by an embedded feature structure, but by a TENSE attribute.
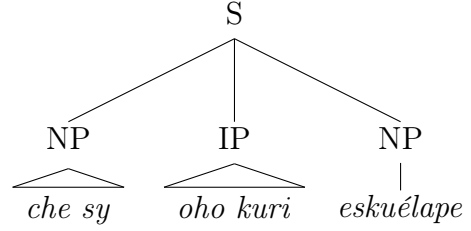

Figure 3: Phrase structure tree of (13).

(14)
$$
\begin{bmatrix}
\text{LEXSEM} & \text{'ha'} \\
\text{SUBJ} & \begin{bmatrix} \text{LEXSEM} & \text{'sy'} \\ \text{POSS} & \begin{bmatrix} \text{PERSON} & 1 \\ \text{NUMBER} & \text{SG} \end{bmatrix} \end{bmatrix} \\
\text{ADJ} & \left\{ \begin{bmatrix} \text{LEXSEM} & \text{'escuéla'} \\ \text{CASE} & \text{LOC} \end{bmatrix} \right\} \\
\text{TENSE} & \text{PAST}
\end{bmatrix}
$$

References

Bruno Estigarribia. *A Grammar of Paraguayan Guarani.* UCL Press, London, 2020.

Johanna Mattissen. On the Ontology and Diachrony of Polysynthesis. In Dieter Wunderlich, editor, *Advances in the theory of the lexicon*, pages 287–354. Walter de Gruyter, Berlin, 2006.