Subordination strategies in South America

Rik van Gijn

Summary

Keywords: subordination, South America, language typology, language contact, genetic linguistics

Human language can be distinguished from other communication systems by its recursivity: the embedding of one structure into another. Even though recursivity is supposed to be a shared feature of all human languages, individual languages may employ radically different morphosyntactic strategies to achieve this goal. Subordination (the embedding of a sentence into another) is one of the areas where we find recursivity.

My research proposal focuses on different strategies languages use to encode subordinate events in a specific regional context: South America. In spite of the high degree of linguistic diversity that characterizes the continent, there seem to be two dominant strategies, found throughout South America, across language families and geographical regions: nominalization and clause chaining. This counter-intuitive situation requires closer study and an explanation.

South American languages have always been underrepresented in samples of comparative linguistic research, but in the last decade a huge amount of data from this continent has become available by descriptive studies carried out in institutions all over the world. Now that these new resources are available for theoretical research, the time has come to harvest.

In the research I propose here, I will investigate sixty South-American languages with respect to the mechanisms they employ to encode subordinate clauses. I will ask three research questions: (1) What strategies are employed by South American languages to encode subordinate events? (2) How are they distributed over the continent? (3) Why is the distribution the way it is?

By combining the latest insights from linguistic typology and language contact with a standardized method that allows for statistical calculations and reproducibility, this research will yield highly valuable results for Americanist linguistics, theoretical linguistics, language typology, contact linguistics, historical (genetic) linguistics, comparative methodology, and population history.

Full description

In 2005, linguist Everett published a paper on the Brazilian language Pirahã, which caused an uproar in the linguistic community. Everett claimed that the language lacked subordination (the embedding of a clause-like structure into another clause). Rather than embedding a clause into another, Pirahã speakers simply juxtapose two clauses (Everett 2005: 629):

\[(1) \quad hi\; ob-\text{áa’ái} \quad kahai \; kai\; sai\]

\[\text{he see-ATTR} \quad \text{arrow make-NOM}\]

‘He knows how to make arrows well’

Everett’s claims are surrounded by controversy, and the debate is still ongoing. In any case, the construction in (1) is rather unusual. It is perhaps not surprising that a structure like this is found in a language from South America, the least known continent in terms of language structures, with a high degree of linguistic diversity (cf. Dixon & Aikhenvald 1999:1, Adelaar with Muysken 2004:2). Until recently, it was hard to study linguistic data from this continent systematically because of the incomplete data. As a result, South American languages have been underrepresented in typological language samples. In the last decade, a dramatic change in this situation has occurred: over 40 new, high-quality descriptive grammars of South American languages have been
produced at various institutes in the world. Now that so much effort has been put in making these data available, the time has come to harvest, to confront linguistics with this vast amount of new data. The research proposed here intends to do just that.

In December 2006, I organized a workshop on subordination in South American languages in Nijmegen. From this workshop (results to be published in Van Gijn et al, in prep.) it transpired that, in spite of the dazzling linguistic diversity of the continent, two subordination strategies seem to be found in many, unrelated, South American languages: nominalization (cf. also Dixon & Aikhenvald 1999:7), and clause chaining - often in combination with switch-reference (cf. Longacre 2007, Dixon & Aikhenvald 1999, Adelaar with Muysken 2004). How is it possible that, in a continent with such a high degree of diversity (quantitative as well as qualitative), similar structures can be found across languages?

I will break down this general question into three sub-questions: (1) What kind of subordination strategies are found in South America? (2) How are they distributed over the continent? (3) Why is their distribution the way it is? I will discuss these questions, and the ways to approach them, in the next section.

1. What kind of subordination strategies are found?

Typological studies of subordination (e.g. Lehmann 1988, Cristofaro 2003, Noonan 2007, Andrews 2007, Thompson et al. 2007, Longacre 2007) show that subordination is not an easy field to delimit; the boundary with coordination (the combination of two clauses of the same morphosyntactic level) is diffuse. Moreover, it is hard to compare languages with respect to subordination, because they can employ radically different morphosyntactic structures to encode comparable meanings. To tackle these problems, I make a distinction between semantic and morphosyntactic subordination.

Following functional typology (cf. e.g. Stassen 1985, Croft 2001, 2003), I will take functions rather than structures as my point of departure. In the most thorough study of the functional aspects of subordination, Cristofaro (2003) studies 18 semantic subclasses of subordination, and comes to the conclusion that, in general terms, semantic characteristics of subordinate relations determine their form. With these results in mind, I will depart from the same 18 semantic classes studied by Cristofaro, since we know the relations between these functional domains and structures, and compare languages in terms of the morphosyntactic strategies they use for each of these domains.

For the classification of the different morphosyntactic strategies, I will base my analysis on Lehmann (1988), who introduces a number of parameters in the form of continua, along which clause linkage strategies can be defined:

I. AUTONOMY vs. INTEGRATION
   a. hierarchical downgrading
      parataxis ←→ governed clause
   b. grammatical integration
      2 full sentences ←→ single (complex) word

II. EXPANSION vs. REDUCTION
   a. decrease of sentence properties
      clause ←→ verbal noun
   b. grammaticalization of matrix verb
      independent predicate ←→ grammatical affix

III. ISOLATION vs. LINKAGE
    explicitness of linking
    no subordinator ←→ subordinator morpheme ←→ subordinator morpheme
    morpheme without lexical content ←→ with lexical content

These continua can be illustrated with the following example sentences (subordinate clauses are between square brackets):
The examples from Pirahã (2a) and Huallaga (2b) represent opposite points on continuum Ia, Pirahã displaying a paratactic structure with two independent sentences, Quechua a governed clause (indicated by the structural case marker on the dependent predicate). This pair of examples also illustrates the opposition between the two extremes on continuum Ila: where in Pirahã the (notionally) subordinate clause has all the characteristics of a sentence, with a verb that has the full potential of verbal inflection, the dependent verb in Huallaga carries no verbal inflection. Instead, it is inflected for case, which is nominal inflection. Continuum Ib can be illustrated by contrasting examples (2a) and (2c): in Pirahã, the relation between the two clauses is expressed by two full sentences; in Yurakaré the two predicates share the negation and the subject participant, which are marked only once. Moreover, no constituent can come in between the predicates of (2c), and they are pronounced in a single intonation contour. In other words: the predicates in the Yurakaré example are almost fused into a single, complex predicate. The final example (2d) shows a construction containing an explicit subordinator morpheme with lexical content (salo ‘because’), whereas in all of the other examples such a marker is absent (continuum III).

A balanced language sample of 60 South American languages (representing 26 different language families, with 11 unclassified languages) of which we have good descriptions and/or access to specialists for additional information, will be investigated for Cristofaro’s 18 semantic domains, and classified along the continua presented above. These continua will be broken down into smaller yes-no questions (referring to e.g. the expression of participants, inflectional (verbal and nominal) possibilities), which allows for positioning constructions on discrete points on the continuum, and subsequent statistical calculations of the correlations between semantic type and morphosyntactic form, as well as between the different components of the morphosyntactic form, using multidimensional scaling techniques. The descriptive results will be made available in a publicly accessible database.

2. How are the different subordination strategies distributed over the continent?

The research with respect to the first question will result in a number of language types in terms of the constructions they use (typological distribution). This typological distribution will then be plotted onto the geographical distribution of the different languages in the sample. Using geostatistical methods, which take into account not only distance in kilometres, but also, for instance, pathways (e.g. rivers) and barriers (e.g. mountains), geographical distance between languages can be measured, and consequently correlated to typological distribution.

3. Why is the distribution the way it is?

In this phase of the research I will have a typological distribution, a geographical distribution, and a correlation between them. If this is a strong correlation, it provides a
good argument for saying that this is the result of diffusion of features.

If the geography-typology correlation is not very strong, I expect to find a stronger correlation between the typological distribution and the known language family histories of the region. This requires a more in-depth study of the language families of South America. An entire study of the language families falls outside the time-frame of this research. However, I will look at two well-studied language families in more detail: Quechuan and Arawak. These two families are represented in the basic 60 language sample by 2 and 4 languages, respectively. I will expand the sample with 6 Quechuan and 4 Arawak languages, making a total of 70 investigated languages. Quechuan and Arawak are interesting since they represent languages from two (cf. Dixon & Aikhenvald 1999:9-10) language areas: the lowlands (Arawak) and the Andes (Quechuan). Furthermore, both families span large territories, which allows for a study of the families in relation to different linguistic neighbours.

With respect to diffusion and inheritance, subordination is particularly interesting. Language contact studies suggest that the more entrenched in the grammar an item (a word, an abstract structure), the more resistant it is to borrowing (cf. e.g. Curnow 2001). When linguistic items were put on a borrowability scale, subordination would be somewhere in the middle: not as easily borrowed as lexical items like nouns, but more easily borrowed than abstract structures functioning on lower levels, such as adjective inflection (cf. Ross 2001:146). In this respect, Lehmann's continuua discussed above under I and II present an interesting background to the question of borrowability, as they can be read as continua of relative entrenchment in the grammatical system, the structures on the right side of the continua being more entrenched.

As the first systematic study of subordination strategies in South America, the research will provide new insights with respect to linguistic creativity, historical contact between groups, migratory movements, and linguistic families. The results are of great importance to Americanist studies, theoretical linguistics, contact linguistics, historical (genetic) linguistics, and population history.

The research combines the latest advances in various linguistic subdisciplines into a new approach. Both genetic linguistics and areal linguistics have mainly been concerned with holistic comparisons of languages. Linguistic typology, on the other hand, has a feature-driven approach (i.e. it looks at the possible forms of a linguistic feature across languages). With the publication of Nichols (1992), typology sought to incorporate areal and genetic linguistics to a greater extent, by shifting the focus in typological research from addressing the question What is possible? to the question What is where why? (Bickel 2007). This young sub-branch of typology has looked at very general, global patterns. This study is unique in that it looks at a smaller geographical area in a more detailed way.

New to the typological study of subordination is the aim of this project to translate the patterns found in the languages into smaller, discrete, parts, used for statistical analysis. Moreover, the whole process, from the language data to the matrices used for statistics, will be available in an open-access database.

Finally, the research will contribute to the recursivity debate started by Everett by placing Pirahã (so far discussed in isolation) in a wider linguistic and geographical context.

References

van Gijn, Rik, Katharina Haude & Pieter Muysken, eds. (in prep.) *Subordination in native South-American languages*. Amsterdam/Philadelphia: John Benjamins [TSL series].