

Portuguese: Corpora, Coordination and Agreement

Louisa Sadler[†], Doug Arnold[†], and Aline Villavicencio^{*}

[†]University of Essex, UK

^{*}Universidade Federal do Rio Grande do Sul, Brazil

1 Introduction

This paper explores the implications of data from a corpus study of Portuguese for our understanding of agreement processes involving coordinate structures (CSs). These data indicate serious limitations of most existing analyses of coordinate structures: this paper will suggest how they can be overcome. It demonstrates how corpus data can require a more sophisticated view of phenomena, and raises some interesting methodological issues.

Agreement phenomena have received considerable attention in recent years, but extending analyses based on non-coordinate structures to deal with CSs presents non-trivial problems. In particular, CSs appear to be able to control agreement in a variety of different ways. The two which are most widely attested crosslinguistically involve (syntactic or semantic) *resolution* and ‘*closest conjunct agreement*’ (CCA). In the former, the agreement properties triggered on an agreement target are some function of the properties of the conjuncts (for example, a CS will trigger masculine agreement if any conjunct is masculine, and feminine agreement only if all conjuncts are feminine). Under CCA, the modifier will agree with just the closest conjunct (so, e.g., post-nominal modifiers will be feminine when the last conjunct is feminine). Resolution is familiar from many languages; CCA has been observed in, *inter alia*, Irish, Welsh, Portuguese, Spanish, Arabic, and Ndebele. (e.g. McCloskey, 1986; Sadler, 1999; Corbett, 1991; Moosally, 1999; Yatabe, 2004)

Resolution can be modelled by a grammatical mechanism which ‘calculates’ the set of resolved agreement features to be associated with the coordinate structure as a whole: this set of resolved features then controls agreement on agreement targets (Dalrymple and Kaplan, 2000). In principle, one might treat CCA in a similar fashion — by associating the agreement features of a single, distinguished, conjunct with the CS as a whole (as in e.g. Moosally, 1999). However, as noted in Sadler (2003, 1999) this is not appropriate in CCA languages such as Welsh, where different agreement processes can target both the resolved and the ‘closest’ conjunct agreement features, suggesting that both resolved and ‘closest’ agreement features are associated with the CS as a whole.

The data discussed here provide further and different support for this point: CSs must make several kinds of agreement information available for NP internal agreement at the same time.

2 Portuguese CSs

Portuguese nominals trigger number and person agreement in a number of different contexts in a fairly straightforward way (e.g. attributive adjectives show person and number agreement with the nouns they modify). However, agreement with coordinate structures is more complex.

To investigate this, a large scale corpus study was undertaken. Here we will summarise the key results, focusing on NP internal agreement processes (cf Villavicencio et al., 2005, for more discussion).

Several different strategies were observed. To begin with, the data showed straightforward cases of resolution (for number and gender), and CCA (for number and gender): cf. (1) where a CS consisting an MSG and an FSG triggers MPL on the post-head adjective; and (2), where *monástica* (‘monastic_{FSG}’) modifies a CS containing an mpl nominal, but shows fsg agreement with the closest conjunct. Prenominal CCA is exemplified in (3).

- (1) o teto e a parede coloridos
the_{MSG} ceiling_{MSG} and the_{FSG} wall_{FSG} coloured_{MPL}
the coloured ceiling and wall
- (2) estudos e profissão monástica
studies_{MPL} and profession_{FSG} monastic_{FSG}
monastic studies and profession
- (3) diversas secções ou subgrupos
diverse_{FPL} sections_{FPL} or subgroups_{MPL}
various sectors or subgroups

The existence of CCA as a possible agreement strategy in Portuguese NPs is briefly noted in some descriptive studies (e.g de Almeida Torres, 1981), but has received little analytic attention. The corpus study reported here indicates that it is far more widespread than generally supposed (perhaps as many as 1 in 10 cases of CS agreement), and is suggestive of some factors which may control its applicability.

However, the most striking result of the study is the existence of two other kinds of case, neither of which has been previously noted in the literature. The first shows that CCA operating simultaneously on pre- and post- head modifiers, *with different effects*.

In (4) an MPL+FPL coordination triggers masculine agreement on the determiner and feminine on the postnominal adjective:

- (4) Esta canção anima os corações e mentes brasileiras.
 This song animate the_{MPL} hearts_{MPL} and minds_{FPL} Brazilian_{FPL}

The second kind of case appears to involve CCA for gender, with resolution for number: the elements of the CS are singular, but it triggers plural agreement on the modifier *sofridas* (‘suffered_{FPL}’), presumably because the CS denotes a plurality of some kind. This is consistent with a resolution strategy. But we see at the same time that *sofridas* is feminine, agreeing with the closest conjunct *a dor* (‘the pain’), despite the presence of a masculine conjunct.

- (5) todo o constrangimento e a dor sofridas
 all_{MSG} the_{MSG} embarrassment_{MSG} and the_{FSG} pain_{FSG} suffered_{FPL}
 all the embarrassment and pain suffered

3 Discussion

Taken together, these data suggest it is necessary to associate at least three kinds of agreement properties with CSs — roughly, information about the leftmost conjunct, information about the rightmost conjunct, and ‘resolved’ information about the CS as a whole. The theoretical challenge is to formulate principles which will determine the projection of this information in a way that is compatible with analyses of agreement involving non-coordinate structures. Such an analysis is provided in Villavicencio et al. (2005) — in outline: suppose three features (say) LAGR, RAGR, and AGR are defined for nominals; lexically, and in normal headed constructions, these are required to have the same value, but in coordinate structures LAGR and RAGR come from the LAGR and RAGR of the leftmost and rightmost daughters respectively, and AGR is computed from the AGR values of the conjuncts. An analysis of (4) can be obtained by allowing pre-head modifiers to get their agreement properties from the CS’s LAGR, while post-head modifiers can get theirs from the CS’s RAGR. The resolutions strategy will involve get agreement properties from the CS’s ARG value.

This leaves open a number of questions, and raises a number of methodological issues.

One relatively straightforward methodological point is that this study is of necessity based on *interpreted* corpus data: it is not enough to find appropriate sequences of CSs and modifiers, it is essential to limit attention to cases where the interpretation makes it clear that the modifier scopes over the whole CS.

A second, rather obvious, methodological point involves the value and limitations of corpus data. On the one hand, the value of corpus data comes out clearly: the existence

of examples like (4) in corpora force one to consider the possibility of CCA operating differently in different directions, which one might not have expected, *a priori*. On the other hand, getting relevant data can be extremely difficult due to various complicating factors. For example, a case like (4) can be taken as showing that determiners show CCA for number and gender, but it is consistent with them showing CCA for gender and resolved agreement for number (investigating this requires conjunctions of singular nouns with pre- and post- nominal modifiers, and these are hard to find). Here one is naturally drawn to constructing examples. But this is not straightforward, because in many cases native speakers appear uncertain about the status of some examples.

In particular, it seems that some speakers reject examples like (5). This rejection raises an important and difficult methodological point, given that such examples occur in significant numbers in naturally occurring texts (over 100 cases, in our corpus).

Acknowledgement

We are grateful for useful comments from many people, including: the anonymous referees for this workshop, numerous colleagues at Essex, participants at HPSG05 in Lisbon, and the ‘Alliance 05 Project’ Workshop held at Paris 7 in Oct 2005, and Mary Dalrymple, and Irina Nikolaeva. The research was supported by the AHRB Project *Noun Phrase Agreement and Coordination*, MRGAN10939/APN17606.

References

- Corbett, G. G. (1991). *Gender*. Cambridge University Press, Cambridge, UK.
- Dalrymple, M. and R. Kaplan (2000). Feature indeterminacy and feature resolution in description-based syntax. *Language*, **76**(4).
- de Almeida Torres, A. (1981). *Moderna gramática expositiva da Língua Portuguesa*. Martins Fontes, Sao Paulo.
- McCloskey, J. (1986). Inflection and conjunction in Modern Irish. *Natural Language and Linguistic Theory*, **4**(2):245–282.
- Moosally, M. J. (1999). Subject and object coordination in Ndebele: and HPSG analysis. In S. Bird, A. Carnie, J. D. Haugen, and P. Norquest, eds., *Proceedings of the WCCFL 18 Conference*. Cascadilla Press, Somerville, MA.
- Sadler, L. (1999). Non-distributive features in Welsh coordination. In M. Butt and T. H. King, eds., *Proceedings of LFG 1999*. CSLI Publications, Stanford, CA.

- Sadler, L. (2003). Coordination and asymmetric agreement in Welsh. In M. Butt and T. H. King, eds., *Nominals: Inside and Out*, pp. 85–118. CSLI Publications, Stanford, CA.
- Schütze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. University of Chicago Press, Chicago, IL.
- Villavicencio, A., L. Sadler, and D. Arnold (2005). An HPSG account of closest conjunct agreement in NP coordination in Portuguese. In S. Müller, ed., *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*, Lisbon. CSLI Publications, Stanford, CA.
- Yatabe, S. (2004). A comprehensive theory of coordination of unlikes. In S. Müller, ed., *Proceedings of the HPSG04 Conference*, pp. 335–355. CSLI Publications, Katholieke Universiteit Leuven.