An HPSG account of closest conjunct agreement in NP coordination in Portuguese

Aline Villavicencio
Universidade Federal do Rio Grande do Sul, Brazil

Louisa Sadler
University of Essex

Doug Arnold
University of Essex

Proceedings of the 12th International Conference on
Head-Driven Phrase Structure Grammar

Department of Informatics, University of Lisbon

Stefan Müller (Editor)

2005

Stanford, CA: CSLI Publications

pages 427–447

Villavicencio, Aline, Louisa Sadler & Doug Arnold. 2005. An HPSG account of closest conjunct agreement in NP coordination in Portuguese. In Stefan Müller (ed.), *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, 427–447. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2005.25.

Abstract

This paper discusses the NP-internal agreement strategies observed in an empirical (corpus based) study of Portuguese, and proposes an analysis which is formalized in the framework of Head-Driven Phrase Structure Grammar (HPSG). The empirical study suggests that what were previously thought to be rare or non-existent strategies occur with surprising frequency. Capturing these strategies poses problems for many standard approaches to agreement. The formalization shows how they can be captured with a relatively conservative extension of the existing HPSG theory of agreement.

1 Introduction

This paper discusses the NP-internal agreement strategies observed in an on-going empirical study of Portuguese, and proposes an analysis which is formalized in the framework of Head-Driven Phrase Structure Grammar (HPSG). In particular, we focus on the behaviour of determiners and attributive adjectives which modify coordinate structures, such as can be seen in (1). As will appear, the agreement strategies observed pose a challenge for most existing approaches to coordination and agreement.

(1) *Esta canção anima os corações e mentes brasileiras.*
This song animate the.MPL hearts.MPL and minds.FPL Brazilian.FPL
'This song animates Brazilian hearts and minds'

The paper is structured as follows. Section 2 provides some background on the way agreement is handled in HPSG, including some brief references to the literature. Section 3 describes the different agreement strategies that appear to be employed in Portuguese in relation to coordinated nouns and NPs. We will suggest that, in addition to the widely attested 'resolution' agreement strategy, Portuguese also uses a crossliguistically less familiar (but still widely attested) 'closest conjunct' agreement strategy for NP internal agreement. Perhaps more surprisingly, we will suggest that Portuguese also permits 'mixed' strategies, for example, using one strategy for prenominal dependents and another for postnominal dependents, in the same NP (in fact, this possibility is exemplified in (1)), and even allowing the use of one strategy for number with another for gender. In Section 4 we will present corpus data which show that these 'alternative' strategies are more widespread than has been generally assumed. Section 5 presents the HPSG formalization: the central idea will be that three kinds of agreement information must be recorded — information about the leftmost and rightmost conjuncts, as well as information

[†]We have benefitted from discussion with many people, but special thanks are due to Mary Dalrymple, Irina Nikolaeva, and participants at the HPSG 2005 Conference in Lisbon. Remaining unclarities and errors are purely our fault, of course.

This research was supported by the AHRB Project *Noun Phrase Agreement and Coordination*, MRGAN10939/APN17606.

about the coordinate structure as a whole. Section 6 provides a conclusion and notes some open questions.

2 Agreement in HPSG

Agreement phenomena have received considerable attention within HPSG since Pollard and Sag (1994) laid the foundations (see, for example, Kathol, 1999; Moosally, 1999; Wechsler and Zlatić, 2001, 2003; Abeillé, 2004; Yatabe, 2004).

Pollard and Sag (1994, Ch2), distinguished two main kinds of agreement: 'index-based' agreement, and 'syntactic' agreement.[1] A typical instance of syntactic agreement (or 'concord') is agreement for case between a noun and a determiner or attributive adjective. One way of modelling this kind of agreement in HPSG is to assume that nouns, determiners, and attributive adjectives carry a feature CONCORD, containing attributes such as CASE and GENDER. NP-internal agreement is then the result of requiring token identity between the CONCORD feature on nouns, determiners and adjectives. Index agreement is more semantic. The idea is that nominal expressions are associated with indices, which correspond roughly to discourse variables — so, for example, a pronoun and its antecedent will share the same index. Indices are taken to be feature structures, specified for attributes like NUMBER, GENDER, and PERSON, whose values relate to the referential/semantic possibilities of the associated nominal. Agreement for person, number, and gender between a pronoun and its antecedent is then an automatic consequence of coindexation. Subject-verb agreement can be handled by having verbs select subjects with a certain kind of index — for example, a third person singular verb like *walks* will require that its subject's INDEX be third person and singular.

This provides an account for a wide range of intricate agreement phenomena, including 'hybrid nouns' (Corbett, 1991), which can trigger different kinds of agreement on different targets within the same clause. For example, in Spanish the title *Majestad* ('Majesty') is feminine, so it triggers feminine agreement on attributive adjectives and determiners. However, if it refers to a male individual, it triggers masculine agreement on a predicative adjective, and requires masculine anaphora:

(2) *Su* *Majestad.*I *Suprema* *esta contento.* *(Éli.*I *...)*
Pron.FEM Majesty Supreme.FEM is happy.MASC. (He.MASC ...)
His Supreme Majesty is happy.

This is easily dealt with in this approach, by allowing CONCORD and INDEX values to differ (cf. Kathol, 1999; Wechsler and Zlatić, 2003). As used in an example like

[1] Pollard and Sag (1994) also discuss a third kind of agreement, 'pragmatic' agreement, which we ignore here. Pragmatic agreement is exemplified by honorific agreement in Korean. The idea is that certain kinds of marking convey background information about social relationships (e.g. between speaker and addressee), and this information must be consistent on expressions which co-occur. Instances of pragmatic agreement failure do not involve violation of grammatical constraints *per se*, so they are infelicitous, rather than strictly ungrammatical.

(2), a partial description of the HEAD value of *Majestad* might be as in (3):

(3)
$$\begin{bmatrix} \text{CONCORD} & \begin{bmatrix} \text{PER} & \textit{3rd} \\ \text{NUM} & \textit{sg} \\ \text{GEN} & \textit{fem} \end{bmatrix} \\ \text{CONTENT} \mid \text{INDEX} & \begin{bmatrix} \text{PER} & \textit{3rd} \\ \text{NUM} & \textit{sg} \\ \text{GEN} & \textit{masc} \end{bmatrix} \end{bmatrix}$$

That is, it will be CONCORD | GEN *fem*, but (as one would expect, given that its referent is male) INDEX | GEN *masc*. The behaviour of a noun like *Majestad* follows if agreement between a noun and an attributive adjective is concord (syntactic agreement, involving the value of CONCORD) whereas agreement between an NP and a predicative adjective involves the NP's INDEX value.

Though the general approach works well for non-coordinate structures, extending it to coordinate structures raises some interesting problems. In particular, predicting the agreement properties of a coordinate structure from the properties of the individual conjuncts turns out to be non-trivial. In cases where conjuncts differ in some agreement property, two strategies are widely attested crosslinguistically (although not, of course, to the total exclusion of other strategies):

Syntactic Resolution: agreement marking on agreement targets is the result of some computation over the properties of (all) the individual conjuncts — e.g. in many languages a coordinate structure will trigger feminine agreement only if all the conjuncts are feminine (e.g. Dalrymple and Kaplan, 2000; Wechsler and Zlatić, 2003);

Closest Conjunct Agreement: agreement marking on an agreement target depends on the properties of only one conjunct — the closest one (Corbett, 1991; Moosally, 1998; Sadler, 1999; Moosally, 1999; Sadler, 2003; Yatabe, 2004).

Closest conjunct agreement (CCA, also known as 'single conjunct', or 'partial' agreement)[2] is quite widespread crosslinguistically, and is found in typologically diverse languages including Romance, Celtic, Semitic and Bantu languages. Most theoretical work to date on these agreement patterns has dealt with closest conjunct predicate-argument agreement (e.g. agreement between a verbal head and its subject and object).

For example, from an HPSG perspective Moosally (1999) proposes an account of single conjunct predicate-argument agreement in Ndebele. Her treatment takes this to be a case of index-agreement, and involves a relation between the INDEX feature of the (coordinate-structure) sign and the INDEX features of the CONJ-DTRS. Moosally's CCA constraint is essentially as in (4):

[2]Strictly speaking, CCA, 'single conjunct' and 'partial' agreement are different concepts — for example, single conjunct agreement should also cover cases of *furthest* conjunct agreement. However, in fact, most cases of single conjunct agreement are cases of CCA.

(4) $\begin{bmatrix} \text{SYNSEM} \mid \text{INDEX} & \boxed{1} \\ \text{CONJ-DTRS} & \left\langle \ldots, \left[\text{INDEX} \quad \boxed{1}\right] \right\rangle \end{bmatrix}$

This constraint requires the INDEX value of the coordinate structure to be token-identical with that of the final conjunct daughter: agreement between a verbal head and a nominal coordinate structure (subject or object) then proceeds in the normal way. While this seems satisfactory for the Ndebele which Moosally discusses, it is inappropriate in very many languages with closest conjunct agreement, in which some agreement processes can be seen to target the features of a single conjunct, but where there is good evidence that the INDEX of the coordinate structure is resolved. For example, in Welsh, predicate-argument agreement is controlled by the closest conjunct, but other agreement processes access resolved features. Thus, in (5), the predicate *dw* ('be') is first person singular, agreeing with the closest conjunct in the subject *i a Gwenllian*, but the pronominal clitic *ein* is plural, reflecting the resolved number value of the coordinate structure subject, which is overall plural (cf. it denotes a plurality).

(5) *Dw i a Gwenllian heb gael ein talu.*
be.1SG I.1SG and Gwenllian.3SG without get Cl.1PL pay
'Gwenllian and I have not been paid' Sadler (2003, (12))

Similarly, in Section 3 we will see cases inside Portuguese NPs where a single coordinate structure controls different agreement properties on different targets.

Yatabe (2004) provides an account of CCA in the context of a more general treatment of unlike categories, in particular, what he calls 'each conjunct' agreement (e.g. the situation where a predicate can occur with a coordinate structure only if it can occur with each of the conjuncts separately).[3] The basic idea is that coordinate structures bear a (*head*) feature ARGS, whose value is a list made up of the conjuncts *head* values. Rather than being 're-ified' as actual feature values, agreement properties (and other properties involved in argument selection) are accessed 'as needed' by various relations.

The case Yatabe considers is that of the verb *to be* as it occurs with *there*. Assuming that English verbs never agree directly with their complements, the agreement pattern one sees in examples like *There is/*are a dog in the garden* vs *There *is/are dogs in the garden* can be handled by assuming that this use of *be* requires its subject to agree in number with its first complement. Simplifying somewhat, Yatabe's constraint to this effect could be stated as in (6), which states that the NUM value of the subject (*there*) must be the value of the relation *num_value* applied to the head value of the first complement.

[3] As regards agreement, Yatabe's focus is on predicate-argument agreement, rather than the NP internal concord processes that are our concern here, but the approach could no doubt be extended.

(6) $\left[\begin{array}{ll} \text{SUBJ} & \left\langle \left[\text{CAT} \mid \text{HEAD} \mid \text{AGR} \quad \textit{num\_value}(\boxed{1}) \right] \right\rangle \\ \text{COMPS} & \left\langle \left[\text{HEAD} \quad \boxed{1}\right], \quad \ldots \right\rangle \end{array}\right]$

In the case where the complement is a coordinate structure (*There were two women and a man in the garden*), this would presumably give *there* the number value of the whole coordinate structure *two women and a man* (i.e. 'resolved' agreement). To deal with 'first conjunct agreement' (i.e. CCA), for cases like *There was a man and two women in the garden*, Yatabe replaces *num_value(*$\boxed{1}$*)* in (6) with *num_value(first(*$\boxed{1}$*))*, where *first(*$\boxed{1}$*)* is defined so as to return the *head* value of the first conjunct in the case of a coordinate structure (i.e. the first element of ARGS in $\boxed{1}$), and otherwise $\boxed{1}$ itself. Yatabe does not discuss the sort of data we will present in Section 3, but there is no reason to suppose that additional relations could not be formulated to handle it.

One striking feature of Yatabe's approach is that, unlike Moosally's, it does not associate a single agreement value, or set of values, with a coordinate structure. Rather, this use of relations to access agreement properties opens the possibility that different processes might involve different relations, and so simultaneously access different properties. Indeed, it should even be possible for a single relation to operate 'non-deterministically' — so that even under one agreement process, a single agreement controller might trigger different agreement on different agreement targets. As will appear, some flexibility of this kind seems to be necessary, but this degree of flexibility may be excessive. Our approach is at once more limited in scope (we deal only with one aspect of coordination — the behaviour of number and gender properties), and more conservative: the formalization we propose in Section 5 will use normal feature percolation principles to associate definite agreement values with coordinate structures; flexibility will be achieved by storing separately information about coordinate structures and (some) individual conjuncts.

3 Agreement and Coordination in Portuguese NPs

In non-coordinate structures, Portuguese determiners and adjectives show a simple pattern of concord in number and gender with the nouns they modify:

(7) *a* *parede* *colorida/*colorido*
the.FSG wall.FSG coloured.FSG/*coloured.MSG

(8) *o* *teto.*MSG *colorido/*colorida*
the.MSG ceiling.MSG coloured.MSG/*coloured.FSG

(9) *o* *teto.*MSG *colorido/*coloridos*
the.MSG ceiling.MSG coloured.MSG/*coloured.MPL

(10) *a parede colorida/*coloridas*
the.FSG wall.FSG coloured.FSG/*coloured.FPL

Coordinate structures on the other hand present a much wider range of agreement patterns: since coordinated nouns often jointly control agreement on determiners, adjectives and other dependents within the NP. In fact, as will appear, mixed gender coordinate structures can trigger different agreement patterns on different targets.

We will begin with a discussion of postnominal dependents (APs), and then turn to prenominal determiners and adjectives.

3.1 Postnominal APs

Postnominal APs appear to show three distinct patterns of agreement.

Firstly, as regards gender, there is the standard resolution pattern, familiar from many two gender systems, of resolution to masculine if any of the conjuncts is masculine, and to feminine only if all conjuncts are feminine. These examples also illustrate a widely attested pattern of resolution for number, whereby a collection of singular conjuncts yields a plural coordinate structure if the coordinate structure as a whole denotes a plurality.[4]

(11) *o homem e a mulher modernos*
the.MSG man.MSG and the.FSG woman.FSG modern.MPL
'the modern man and woman'

(12) *o teto e a parede coloridos*
the.MSG ceiling.MSG and the.FSG wall.FSG coloured.MPL
'the coloured ceiling and wall'

However postnominal APs can also show a second strategy, in which the adjective agrees with the closest (i.e. final) noun in the preceding coordinate phrase:

(13) *estudos e profissão monástica*
studies.MSG and profession.FSG monastic.FSG
'monastic studies and profession'

(14) *no povo e gente hebreia*
on the.MSG population.MSG and people.FSG hebrew.FSG
'on the hebrew people' (de Almeida Torres, 1981)

Notice that in these examples the postnominal AP scopes over the whole coordinate phrase, not just the final noun (this is clear for (14), even out of context). Thus, these appear to be genuine cases of CCA, where the adjectives modify an entire coordinate structure, but only agree with one of the conjuncts (the closest).

[4]Compare examples like *my friend and colleague Mr. Smith* where a coordinate structure denotes a single entity rather than a plurality; cf. also the discussion around examples (26) and (25), below.

Given that a language permits both resolution and CCA for the same agreement process (here concord between N and postnominal AP), one might wonder whether the two strategies can be uses simultaneously for different features. The following examples seem to illustrate exactly this, the third pattern that we find for postnominal APs: CCA for gender and resolution for number:

(15) *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'

(16) *o drama e a loucura vividas*
the.MSG drama.MSG and the.FSG madness.FSG lived/felt.FPL
'the drama and the madness experienced'

(17) *o aprendizado e a experiência vividas*
the.MSG learning.MSG and the.FSG experience.FSG lived/felt.FPL
'the accumulated learning and experience'

(18) *o romantismo e a morbidez profundas da*
the.MSG romanticism.MSG and the.FSG morbidity.FSG deep.FPL of the
alma alemã
soul German
'the profound romanticism and morbidity of the German soul'

(19) *uma relação entre sobrecarga do organismo e envelhecimento*
a relation between overload of the organism and aging.MSG
e morte prematuras
and death.FSG premature.FPL
'A relation between overload of the organism and premature aging and death'

There is little literature to date on agreement strategies beyond simple resolution for Portuguese coordinate structures. One detailed descriptive grammar of Portuguese (de Almeida Torres, 1981) provides some discussion and exemplification of CCA within Portuguese NPs but does not mention this mixed pattern. In Section 4 we will present data from a corpus study which indicate that these 'non-standard' strategies are relatively common.[5]

3.2 Prenominal Modifiers

The interpretation of what goes on prenominally is somewhat less straightforward. Consider first examples such as the following:

[5]We should point out that some Portuguese speakers have serious reservations about at least some of these 'mixed strategy' examples (despite the fact that they are attested rather than constructed), and it is of course possible that some of them simply represent mistakes. However, our corpus study suggests the strategy is not uncommon (it appears in 90 instances in our sample, perhaps as many as 5% of relevant cases). Clearly, the matter deserves more study.

(20) *suas próprias reações ou julgamentos*
his.FPL own.FPL reactions.FPL or judgements.MPL
'his own reactions or judgements'

(21) *as assustadoras colinas e morros de argila do*
the.FPL frightening.FPL mounds.FPL and hills.MPL of clay of the
Parque Nacional
National Park
'the frightening mounds and clay hills of the National Park'

(22) *diversas secções ou subgrupos*
diverse.FPL sections.FPL or subgroups.MPL
'various sectors or subgroups'

Notice that these examples all involve coordinations of feminine and masculine nouns, in that order, and in each case the agreement features of the prenominal elements match those of the initial conjuncts. In fact, a gender mismatch between the first conjunct and the prenominal material appears to lead to ungrammaticality, as in (23). These data appear to indicate that gender resolution is not permitted prenominally, and CCA is the only possible strategy, at least for gender.

(23) **suas próprias julgamentos ou reações*
*his.FPL own.FPL judgements.MPL or reactions.FPL
'his own reactions or judgements'

However, in these examples all the conjoined nouns are plural, so they cannot be used to see whether CCA is also being used for number, or whether there is a mixture of CCA in gender with resolved agreement for number. Investigating this requires coordinations involving singular conjuncts. Unfortunately, further issues arise with singular conjuncts, which complicate matters.

On the one hand, there appear to be some clear cases of number resolution in prenominal modifiers, as can be seen from examples where there is a difference in number between prenominal modifiers and first conjunct:

(24) *Os prováveis diretor e ator principal são Gus*
the.MPL probable.PL director.MSG and actor.MSG principal.MSG are Gus
Van Sant e Johnny Deep, respectivamente
Van Sant and Johnny Deep respectively
'the likely director and main actor are, respectively, Gus Van Sant and Johnny Deep'[6]

To see whether CCA for number is also possible requires examples where a singular determiner precedes a coordinate structure denoting a plurality. King and Dalrymple (2004) suggest that this is impossible. They claim that the singular determiner *o/a* ('the') cannot modify conjoined singular nouns which referring to

[6] *prováveis* ('probable') is plural, but not marked for gender (like many other adjectives in Portuguese).

more than one individual. They contrast (25) with (26), which is acceptable, but receives a interpretation such that it refers to a single individual:

(25) **o cachorro e gato*
the.MSG dog.MSG and cat.MSG
'the dog and cat' King and Dalrymple (2004, 91)

(26) *o presidente e diretor da Air France*
the.MSG president.MSG and director.MSG of Air France
'the president and director of Air France' King and Dalrymple (2004, 92)

One interpretation of this, which would be consistent with (24), would be that the determiner-noun agreement involves resolved number, closely tied to the semantics (a singular determiner is only possible with a coordinate structure which denotes a singular entity).

However, further work is required to determine whether this restriction is completely robust. For example (27), an attested example, is acceptable to the author of the present paper who is a native speaker of Portuguese (Villavicencio, who speaks Brazilian Portuguese), and the (constructed) examples (28) and (29) are judged acceptable by at least some native speakers. In each case there is a singular determiner scoping over a coordination of singular nouns referring to more than one individual (notice that in (28) and (29) the verbs are plural). On the face of it, these examples cannot involve resolved number agreement, and must involve CCA for number.

(27) *a correcta gestão e preservação*
the.FSG correct.FSG management.FSG and conservation.FSG
'the correct management and conservation'

(28) *o presidente e amigo comeram juntos*
the.MSG president.MSG and friend.MSG ate.3PL together
'the president and (his) friend ate together'

(29) *o chefe e vice-chefe estavam na reunião*
the.MSG chief.MSG and vice-chief.MSG attended.3PL the meeting
'the chief and vice-chief attended the meeting'

In summary, for prenominal dependents, gender agreement with the closest conjunct is always required. As for number agreement, there appear to be cases of resolved agreement. On the other hand, there is some evidence for the existence of cases of singular determiners scoping over coordinated singular nouns, which are interpreted as denoting pluralities. If so, these are cases which exhibit CCA in number. However, we will not try to settle this matter here.

3.3 Combining Prenominal and Postnominal Modifiers

Given that CCA is available for both pre- and post-nominal dependents, one might wonder if a coordination of (say) a masculine and a feminine nominal might be

able to trigger *different* agreement on pre- and post-head dependents. Examples like the following seem to show this is possible. In the following, *corações e mentes* triggers masculine agreement on the determiner, and feminine agreement on the postnominal adjective *brasileiras*:

(30) *Esta canção anima os corações e mentes brasileiras.*
This song animate the.MPL hearts.MPL and minds.FPL Brazilian.FPL

The significance of this it is not possible to define a a single feature or set of features to contain the 'syntactic' agreement properties of a coordinate structure (in the manner of most 'standard' approaches). Rather, a coordinate structure must make available several different collections of syntactic agreement features at the same time.

3.4 Summary

In this section we have presented evidence that suggests Portuguese uses a mixture of strategies for NP internal agreement. Prenominally, we have suggested that gender agreement involves CCA, but that the range of strategies involved in number agreement is less clear. Postnominally, we have suggested there may be three strategies:

1. resolution for number and gender;
2. CCA for number and gender;
3. a 'mixed' strategy: CCA for gender and resolution for number.

We have also noted that it seems possible for different strategies to be used for pre- and post-head dependents simultaneously.

Schematically, we might represent these alternatives for postnominal agreement as in (31)-(33).[7]

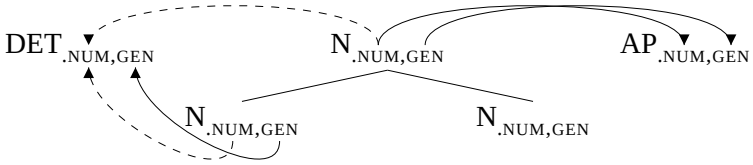
4 Data from a Corpus Study

One clear result of the preceeding discussion would seem to be that Portuguese possesses a rather rich variety of agreement strategies in relation to coordinate structures. As part of our on-going investigation into this, a corpus study was carried out to estimate the approximate frequency of different agreement strategies. The initial results of this study relate to coordinate NPs modified by postnominal adjectives. Here we will report the results of an investigation which concentrated on postnominal *plural* adjectives, and was primarily intended to investigate the occurrence of gender agreement controlled by the closest conjunct.

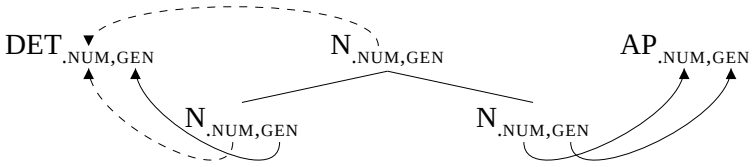
We obtained occurrences of coordinated NPs followed by plural adjectives by posing Google queries of the following general format:

[7] Notice that the representation of prenominal agreement relations is the same in each. The use of dotted lines reflects our uncertainty about the proper account prenominal number agreement.

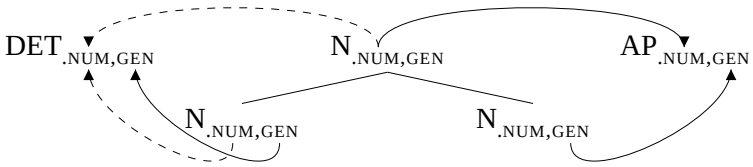
(31) Resolved number and gender:



(32) CCA for number and gender:



(33) CCA for gender, resolved number:



(34) "<ART> * e <ART> * <ADJ>"

Here ART stands for instances of the Portuguese (definite and indefinite) articles, ADJ stands for instances of Portuguese adjectives, and e is the Portuguese conjunction ('and'). The adjectives were extracted from the 1,528,590 entry NILC Lexicon.[8] As we were interested primarily in the correlation between the gender of each of the NPs and the gender of the adjective, only adjectives that overtly reflect gender distinctions were used (9,915 masculine and 9,811 feminine adjectives). The results were manually inspected to remove noise — in cases of putative CCA this entailed removing all cases in which, in the judgement of the Portuguese native speaker, the adjective should be interpreted as scoping only over the closest noun.

The results found are displayed in Tables 1 and 2, where 'Frequency' indicates the number of hits returned by Google for the searches, and 'NP1', 'NP2' and 'Adj' refer to the gender of the first conjunct, second conjunct, and adjective, respectively. Table 1 relates to coordinations of singular NPs, Table 2 relates to coordinations of singular NPs. In both cases the adjectives are all plural, however.

Several observations are worth making here. First, notice that row (d), which reports on 'masculine+feminine' coordinate structures triggering feminine agreement on a following adjective, unambiguously involves resolution for gender. As

[8] See http://www.nilc.icmc.usp.br/nilc/index.html.

	Frequency	NP1	NP2	Adj	
(a)	0	f	m	f	(resolve to f)
(b)	489	f	m	m	(cca/resolve to m)
(c)	460	m	f	f	(cca)
(d)	2317	m	f	m	(resolve to m)
total	3266				

Table 1: Frequency of Masc vs Fem Adjectives Modifying Mixed Gender Coordinate NPs (Plural).

	Frequency	NP1	NP2	Adj	
(a)	0	f	m	f	(resolve to f)
(b)	137	f	m	m	(cca/resolve to m)
(c)	90	m	f	f	(cca)
(d)	1737	m	f	m	(resolve to m)
total	1964				

Table 2: Frequency of Masc vs Fem Adjectives Modifying Mixed Gender Coordinate NPs (Singular).

one might expect, this pattern is very frequent. The agreement pattern reported in row (b) involves cases where the final conjunct is masculine, and could be instance of either gender resolution or closest conjunct gender agreement, since either would result in masculine agreement on the adjective.

On the other hand, row (c) represents cases of the 'masculine+feminine' coordinate structure triggering feminine agreement: these are instances of what we take to be CCA in Section 3. One striking result of this study is that this relatively little discussed pattern is actually not very infrequent. Notice that rows (c) and (d) correspond to those coordinate structures with final feminine conjuncts, that is, the cases in which the existence of CCA of gender could be unambiguously distinguished from other strategies. Thus, one relevant comparison is the ration of cases in (c) (apparent cases of CCA), compared to cases in (c)+(d) (that is, the total number of cases where we would be able to detect CCA if it occurred). We observe the CCA strategy in 460/2777 cases (16.56%) for plural NPs and 90/1827 (4.9%) cases for singular NPs, giving an overall frequency of some 550/4604 cases (11.9%). That is, even on the narrowest interpretation, that is, without considering additional coordinate structures with masculine final conjuncts (where CCA for gender cannot be unambiguously detected), the CCA for gender strategy is widespread, occurring in better than one in ten cases.

Second, notice that in each table, row (a) represents cases where a 'feminine+masculine' coordinate structure triggers feminine agreement — that is, what could only be cases of resolution to feminine. The fact that this is zero provides strong evidence that cases of feminine gender agreement in the presence of some masculine conjuncts as it occurs elsewhere should not be interpreted as the result

of a particular *resolution* strategy. This "unexpected" feminine gender agreement occurs *only* when the final conjunct is feminine. The zero score in (a) combined with the non-zero score in (c) is strongly suggestive that we have been correct in treating this pattern as a case of CCA.

The raw figures also display a strong and interesting bias for masculine conjuncts to precede feminine conjuncts (feminine conjuncts precede in only 626/5230 cases). This is likely to be a reflection a prescriptive bias in favour of this ordering of conjuncts.

Finally, recall that though we have reported numbers of singular and plural NPs separately, in both cases the post-nominal adjectives are plural. Thus, in Table 1, which reports numbers from plural NPs with plural adjectives, row (c) could be interpreted as showing CCA for both number and gender, or alternatively as showing the 'mixed' strategy of CCA for gender and resolution for number. However, the corresponding row in Table 2 is not open to this interpretation, the cases represented there involve singular NPs, with a plural adjective, so they can *only* be interpreted as involving a mixed strategy of CCA for gender and resolution for number. In our sample, then, this strategy is used 90 times, that is, in just under 5% of all cases involving singular NPs.

5 HPSG Analysis

To account for the cases of agreement described above, we propose an analysis that stores agreement information about the leftmost and rightmost conjuncts in two new agreement related features (i.e. in addition to CONCORD and INDEX features): LAGR for the leftmost conjunct, and RAGR, for the rightmost conjunct; CONCORD will be used to contain 'resolved' agreement information.

Like CONCORD, LAGR and RAGR are head features, defined on all sorts where CONCORD is defined (for concreteness, we assume this is at least the *head* values of nouns, determiners and adjectives), and 'normally' (e.g. in headed constructions) all three features share values. Thus, for example lexical nouns satisfy the constraint in (35). Since the features in question are *head* features, this identity carries over to N' and NP:

(35) *noun* $\wedge$ *lexical* $\rightarrow$

$$\begin{bmatrix} \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} & \begin{bmatrix} \text{LAGR} & \boxed{1} \\ \text{RAGR} & \boxed{1} \\ \text{CONCORD} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

As will appear, the idea is that determiners and prenominal adjectives agree with nouns via LAGR (at least for gender), while postnominal adjectives agree with nouns via RAGR. Since for non-coordinate structures these features have the same value, this does not produce any observable effect.

However, in non-headed constructions, in particular, in coordinate structures, the identity between these values breaks down. Instead, the value of LAGR comes

from the LAGR of the leftmost daughter, and the value of RAGR from the RAGR of the rightmost daughter, while the CONCORD value reflects the resolved agreement features of the coordinate structure. To begin with, coordinate phrases which are defined for LAGR and RAGR (e.g. nominal-coordinated-phrases, *ncph*) satisfy the following constraint:

(36)
$$\textit{ncph}\left[\begin{array}{l}\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\quad\left[\begin{array}{l}\text{LAGR}\quad\boxed{1}\\ \text{RAGR}\quad\boxed{2}\end{array}\right]\\ \text{CONJ-DTRS}\quad\left\langle\left[\ldots\text{HEAD}\,|\,\text{LAGR}\quad\boxed{1}\right],\ \ldots,\ \left[\ldots\text{HEAD}\,|\,\text{RAGR}\quad\boxed{2}\right]\right\rangle\end{array}\right]$$

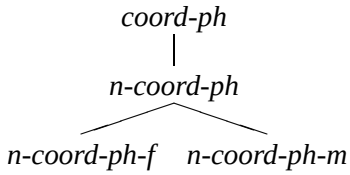
In words: the value of LAGR on a nominal coordinate phrase comes from the LAGR of the first/leftmost daughter of the phrase, RAGR comes from the RAGR of the last/rightmost daughter.

The value of CONCORD on the mother reflects resolved GENDER and NUMBER values computed from the values on the conjunct daughters. As regards NUMBER, we assume (in the absence of contradictory data) that resolution is simply a matter of semantics: (i) the value of INDEX | NUM on a nominal (coordinate or not) is plural whenever the nominal denotes a plurality; and (ii) the value of CONCORD just reflects this. As regards (ii), this means that all *head* values (including those on coordinate structures) satisfy (37):

(37)
$$\textit{head}\left[\begin{array}{l}\text{CONTENT}\,|\,\text{INDEX}\,|\,\text{NUM}\quad\boxed{1}\\ \text{CONCORD}\,|\,\text{NUM}\quad\boxed{1}\end{array}\right]$$

In words, as regards NUMBER, CONCORD and INDEX are always identical.[9]

Resolution for GENDER is slightly more complex. To deal with it, we introduce two subtypes of nominal-coordinate-phrase (which is itself a subtype of coordinate phrase): one for coordinate phrases that resolve to masculine, and one for those that resolve to feminine.

(38)


The relevant constraints on these sorts are as follows:

[9] As stated, this is a 'hard' constraint. It predicts that one should not find divergences of INDEX | NUMBER and CONCORD | NUMBER analogous to the divergence of gender observed with nouns like *Majestad* 'Majesty', mentioned in Section 2. That is, there should not be cases nouns denoting pluralities which trigger singular agreement, or *vice versa*. This is incorrect. For example, Wechsler and Zlatić (2003) discuss a class of collective nouns in Serbo-Croat (the '*deca*-type') which trigger singular agreement inside NP. The constraint should be only a default. However, nothing else in the analysis hangs on this.

(39) *n-coord-ph-f* →

$$\left[\begin{array}{l}\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\,|\,\text{CONCORD}\,|\,\text{GEND}\ \boxed{3}\\ \text{CONJ-DTRS}\quad\left\langle\left[\text{SS}\,|\,\text{LOC}\,|\,\text{HEAD}\,|\,\text{CONCORD}\,|\,\text{GEND}\ \boxed{3}\,\textit{fem}\right]^{*}\right\rangle\end{array}\right]$$

(40) *n-coord-ph-m* →

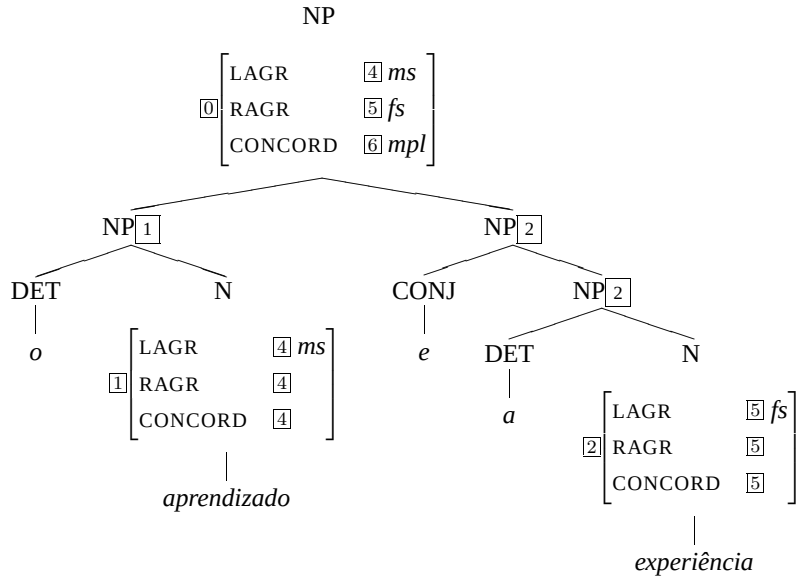
$$\left[\begin{array}{l}\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD}\,|\,\text{CONCORD}\,|\,\text{GEND}\ \boxed{3}\\ \text{CONJ-DTRS}\quad\left\langle .^{*},\ \left[\text{SS}\,|\,\text{LOC}\,|\,\text{HEAD}\,|\,\text{CONCORD}\,|\,\text{GEND}\ \boxed{3}\,\textit{masc}\right],\ .^{*}\right\rangle\end{array}\right]$$

In words, (39) says that a coordinate structure is feminine just in case all its daughters are feminine — intuitively, if its conjunct daughters list is of the form "`fem*`", a list of zero or more feminines; (40) says that a coordinate structure is masculine if it contains a single masculine daughter, that is, if it consists of a masculine daughter preceded and followed by zero or more other daughters — if it is "`.*,masc,.*`", so to speak. These constraints are stated using regular expressions over conjunct daughters, which seems natural, but they could clearly be stated in many other ways, e.g. using list membership predicates.

The following will exemplify these constraints working together.[10]

(41) *...o aprendizado e a experiência (vividas) ...*
...the.MSG learning.MSG and the.FSG experience.FSG (lived.FPL) ...

(42)



The values of LAGR, RAGR, and CONCORD are the same on each lexical noun (cf. constraint (35)); because these are *head* features, these values percolate to the NPs. The value of LAGR on the coordinate structure is $\boxed{4}$*ms* (i.e. masculine,

[10] We have assumed that the conjunct *e* ('and') forms a constituent with the final conjunct. Nothing hangs on this.

singular) — the same as the leftmost conjunct daughter, as required by (36). Similarly, the value of RAGR is [5] *fs* because that is the value of RAGR on the rightmost daughter. The value of CONCORD is *mpl*: *plural* because the coordinate structure denotes a plurality, *masc* because one of the conjunct daughters is masculine — cf. the structure satisfies (40), and does not satisfy (39).

Having described the propagation of agreement features in coordinate structures, we now turn to the matter of agreement with determiners and attributive adjectives. In section 3 we suggested the following patterns exist:

- Posthead:
  - CCA for NUMBER and GENDER;
  - resolution for NUMBER and GENDER;
  - resolution for NUMBER, CCA for GENDER.
- Prehead (Determiners and Prenominal Adjectives)
  - CCA (at least for GENDER).

As standardly assumed in HPSG, attributive adjectives, like other adjuncts, have (as part of their HEAD specification) a MOD feature which expresses constraints on the sort of object the adjective can modify. Agreement between attributive adjectives and nouns can be captured by stating constraints on the relation between the value of agreement features within this MOD value, and values on the adjective itself.If we take the general conditions on adjectival modifiers to be along the lines of (43a), we can capture the different agreement patterns if we require attributive adjectives to satisfy one of the additional constraints in the disjunction in (43b).

(43) a.

$$\textit{head}\left[\begin{array}{ll}\text{CONCORD} & [0]\begin{bmatrix}\text{NUM} & [1]\\ \text{GEN} & [2]\end{bmatrix}\\ \text{MOD} & \left[\begin{array}{ll}\text{LOC}\,|\,\text{CAT}\,|\,\text{HD} & \left[\begin{array}{ll}\text{LAGR} & [3]\begin{bmatrix}\text{NUM} & [4]\\ \text{GEN} & [5]\end{bmatrix}\\ \text{CONCORD} & [6]\begin{bmatrix}\text{NUM} & [7]\\ \text{GEN} & [8]\end{bmatrix}\\ \text{RAGR} & [9]\begin{bmatrix}\text{NUM} & [10]\\ \text{GEN} & [11]\end{bmatrix}\end{array}\right]\end{array}\right]\end{array}\right]$$

b. $[0]=[6] \vee [0]=[9] \vee ([1]=[7] \wedge [2]=[11]) \vee [2]=[5]$

Let us consider these conditions in turn.

1. $[0]=[6]$ identifies the adjective's CONCORD with the CONCORD value of the nominal it modifies. This is appropriate for an adjective under a resolution strategy (for both NUMBER and GENDER) — an adjective such as *modernos* in (44):

(44) *o homem e a mulher* ***modernos***
[the.MSG man.MSG and the.FSG woman.FSG] modern.MPL
'the modern man and woman'

2. $\boxed{0}=\boxed{9}$ identifies the adjective's CONCORD with the RAGR of the nominal it modifies. This is appropriate for a post-head modifier under a CCA strategy (for both NUMBER and GENDER):

(45) *estudos e profissão* ***monástica***
[studies.MSG and profession.FSG] monastic.FSG
'monastic studies and profession'

3. ($\boxed{1}=\boxed{7} \wedge \boxed{2}=\boxed{11}$) identifies the adjective's NUMBER value with the nominal's CONCORD | NUMBER value (i.e the resolved NUMBER), and the adjective's GENDER with the GENDER of the nominal's RAGR. This is appropriate for the 'mixed' resolution/CCA strategy with post-head dependents:

(46) *o constrangimento e a dor* ***sofridas***
[the.MSG embarrassment.MSG and the.FSG pain.FSG] suffered.FPL
'all the embarrasment and pain suffered'

4. $\boxed{2}=\boxed{5}$ identifies the adjective's CONCORD | GENDER with the LAGR | GENDER of the nominal it modifies. This is appropriate for pre-head modifiers under a CCA strategy for gender:[11]

(47) *suas* ***próprias*** *reações ou julgamentos*
his.FPL own.FPL [reactions.FPL or judgements.MPL]
'his own reactions or judgements'

This formulation evades the issue of number agreement for prenominal adjuncts — in section 3 we left open the question of whether they show resolution or CCA (or indeed both) for number. If they turn out to show CCA for number, then we should replace this equation with $\boxed{0}=\boxed{3}$; resolved number can be stated as $\boxed{1}=\boxed{4}$.

We can now be slightly more precise. Supposing we have some way of picking out pre- and post-head adjectives (here we suppose there is a type distinction, but nothing hangs on this), we can say:

(48) a. *post-head-attrib-adj* $\rightarrow$
(43a) $\wedge$ ($\boxed{0}=\boxed{6} \vee \boxed{0}=\boxed{9} \vee (\boxed{1}=\boxed{7} \wedge \boxed{2}=\boxed{11})$)
b. *pre-head-attrib-adj* $\rightarrow$
(43a) $\wedge$ ($\boxed{2}=\boxed{5} \vee \ldots$)

In words, (48a) states that postnominal adjuncts can either (i) share have the resolved (i.e. CONCORD) value of the nominal, or (ii) its RAGR, or (iii) take GENDER from RAGR | GENDER, and NUMBER from the CONCORD | NUMBER of the

[11] We have highlighted *próprias* ('own') in (47), but of course *suas* ('his/her') shows the same agreement.

nominal (i.e. CCA for gender with resolved number). (48b) states that pre-head adjectives take their gender from the LAGR | GENDER of the noun (i.e. CCA for gender); '...' can be filled in with whatever conditions turn out to be appropriate for prenominal number agreement. Together (48a) and (48b) amount to a straightforward statement of the descriptive generalizations we have seen in the preceding discussion.

We have not so far discussed determiners in this section. But this is straightforward: it is standardly assumed that determiners carry a SPEC feature, which constrains the kind of nominal the determiner can combine with. The appropriate constraint for determiners can be obtained by replacing MOD by SPEC in (43a).[12]

Notice that though we have exemplified these constraints with coordinate nominals, they apply equally, and without modification, to cases with non-coordinate nominals — it is just that with non-coordinate nominals LAGR, RAGR, and CONCORD are all identical. It is one of the attractions of this approach that it handles agreement in cases involving non-coordinate structures with the same apparatus as cases involving coordinate structures, without any extra complication of the grammar.

6 Conclusion

This paper has presented the results of an investigation of agreement processes involving NP/noun coordinations in Portuguese. We have provided a detailed description of some aspects of the phenomena, some of which appear to have been previously neglected, and given some of the results of a relatively large scale corpus study. Here the main results appear to be that what were thought to be relatively rare or non-existent agreement patterns are attested, and in some cases fairly widespread. We have suggested a way in which the phenomena can be described, expressed in the formalism of HPSG. The suggestion is that coordinate structures make available three kinds of 'syntactic' agreement related information: agreement properties from the leftmost conjunct; agreement properties from the rightmost conjunct; and 'resolved' agreement properties. The HPSG formalization of this analysis involves the introduction of two novel features (LAGR and RAGR, distinct from CONCORD), and a number of principles governing the way these features are projected.

The paper represents on-going work, which is part of a larger project on agreement processes, and it leaves a number of question open. We will highlight three.

First, and most obviously, we have made no commitment about the way in which number agreement works for prenominals. This clearly requires further exploration. One interesting question involves interaction with the semantics of de-

[12] Van Eynde (2003) argues that adjectives and determiners should not be distinguished in the way they select the nominals they modify, in which case the same constraint(s) would be able to deal with both.

terminers — for example, our existing data suggests that CCA for number is much more readily acceptable with indefinites than with definites.

Second, though we have identified different strategies, we have not addressed the question of what factors favour the choice of one strategy over another. Our preliminary data suggest that, postnominally, animate nouns overwhelmingly favour a resolution strategy, while the majority of CCA cases involve inanimate nouns.

Finally, we have concentrated entirely on NP-internal agreement processes, but other agreement processes appear to show CCA effects, leading one to wonder about the 'syntactic persistence' of features like LAGR and RAGR outside the NP, and their availability for other agreement processes. Consider the following example, which appears to show 'predicate-argument CCA':

(49) *o travestismo e a copulação ritual são*
the.MSG transvestism.MSG and the.FSG copulation.FSG ritual be.PL
realizadas para expressar . . .
realized.FPL to express . . .
'the transvestism and the ritual copulation are realized to express . . . '

Notice that here the passive form *realizadas* ('be realized') is feminine singular, like the final conjunct of the subject (*copulação ritual* 'ritual copulation'), though the subject itself denotes a plurality, and contains a masculine noun (giving it a resolved value of masculine plural).

References

Abeillé, Anne. 2004. A lexicalist and construction-based approach to coordinations. In Stefan Müller (ed.), *Proceedings of the HPSG04 Conference*, CSLI Publications, Katholieke Universiteit Leuven.

Corbett, Greville G. 1991. *Gender*. Cambridge, UK: Cambridge University Press.

Dalrymple, Mary and Kaplan, Ron. 2000. Feature indeterminacy and feature resolution in description-based syntax. *Language* 76(4).

de Almeida Torres, Artur. 1981. *Moderna gramática expositiva da Língua Portuguesa*. Sao Paulo: Martins Fontes.

Kathol, Andreas. 1999. Agreement and the Syntax-Morphology Interface in HPSG. In Robert Levine and Georgia Green (eds.), *Studies in Contemporary Phrase Structure Grammar*, pages 209–260, Cambridge and New York: Cambridge University Press.

King, Tracy Holloway and Dalrymple, Mary. 2004. Determiner Agreement and Noun Conjunction. *Journal of Linguistics* 40, 69–104.

Moosally, Michelle. 1998. *Noun Phrase Coordination: Ndebele Agreement Patterns and Cross-Linguistic Variation*. Ph. D.thesis, University of Texas at Austin, Austin, Texas.

Moosally, Michelle J. 1999. Subject and object coordination in Ndebele: and HPSG analysis. In Sonya Bird, Andrew Carnie, Jason D. Haugen and Peter Norquest (eds.), *Proceedings of the WCCFL 18 Conference*, Cascadilla Press, Somerville, MA.

Pollard, Carl J. and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sadler, Louisa. 1999. Non-Distributive Features in Welsh Coordination. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of LFG 1999*, Stanford, CA: CSLI Publications.

Sadler, Louisa. 2003. Coordination and Asymmetric Agreement in Welsh. In Miriam Butt and Tracy Holloway King (eds.), *Nominals: Inside and Out*, pages 85–118, Stanford, CA: CSLI.

Van Eynde, Frank. 2003. On the Notion 'Determiner'. In Stefan Müller (ed.), *Proceedings of the HPSG-2003 Conference, Michigan State University, East Lansing*, pages 391–396, Stanford: CSLI Publications.

Wechsler, Stephen and Zlatić, Larisa. 2001. A Theory of Agreement and its application to Serbo-Croatian. *Language* pages 391–423.

Wechsler, Stephen and Zlatić, Larisa. 2003. *The Many Faces of Agreement*. Stanford, CA: CSLI Publications.

Yatabe, Shûichi. 2004. A Comprehensive Theory of Coordination of Unlikes. In Stefan Müller (ed.), *Proceedings of the HPSG04 Conference*, pages 335–355, CSLI Publications, Katholieke Universiteit Leuven.