Portuguese: Corpora, Coordination and Agreement

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

[*]University of Essex, UK

[†]Universidade Federal do Rio Grande do Sul, Brazil

1. Introduction

This paper reports some results from a corpus study of Portuguese, and explores their implications for the analysis of agreement processes involving coordinate structures (CSs), especially as regards gender agreement within noun phrases (NPs).[1] Agreement phenomena have received considerable attention in recent years, but agreement involving CSs, and NP-internal agreement processes have received less attention. As will appear, this cannot be taken as a reflection of inherent theoretical interest. Some of the data discussed here appear to be novel, and to pose a serious challenge for existing analyses of coordinate structures. One goal of this paper is to suggest how they can be overcome. More generally, the study demonstrates the value of corpus data in challenging existing analyses, requiring a more sophisticated view of phenomena, and raises some interesting methodological issues. The paper is structured as follows. Section 2 introduces some basic ideas about agreement in general, and what is standardly assumed about Portuguese. Section 3 describes the corpus study itself, and the results. The key conclusion is that Portuguese agreement is more complex than has generally been assumed hitherto. Section 4 discusses the theoretical implications, and provides a relatively theory neutral and intuitive analysis of the facts about Portuguese agreement as they emerge. The main point is that, contrary to what is assumed in most approaches to agreement, CSs must make several kinds of agreement information available *at the same time*. Section 5 summarizes the discussion and provides some brief comments of a methodological nature.

2. Background

In general terms, 'agreement' refers to the phenomenon where the form of one element (the 'agreement target') varies depending on properties of another (the 'agreement controller'). For example, the following show that Portuguese nominals control agreement for number and gender on determiners and adjectives.

(1) *o* *teto* *colorido*
the.MSG ceiling.MSG coloured.MSG
'the coloured wall'

(2) **os/a/as teto*
the.MPL/the.FSG/the.FPL ceiling.MSG
coloridos/colorida/coloridas
coloured.MPL/coloured.FSG/coloured.FPL
'the coloured wall'

Agreement phenomena in general have received considerable attention in recent years. However, the main focus has been on subject-predicate agreement, at the expense of other forms of agreement, notably head-modifier agreement. In particular, there has been relatively little work on the problems that arise in this kind of agreement when the agreement controller is a Coordinate Structure (CS). It turns out that extending analyses based on non-coordinate structures to deal with CSs raises non-trivial problems. In particular, CSs appear to be able to control agreement in a variety of different ways.

The two agreement strategies which are most widely attested crosslinguistically involve (syntactic or semantic) *resolution* and '*closest conjunct agreement*' (CCA). Resolution strategies are familiar from many languages (for discussion and references see e.g. Corbett, 1991; Dalrymple and Kaplan, 2000; Wechsler and Zlatić, 2003). Intuitively, under a resolution strategy agreement involves properties of the CS as a whole — more precisely, the agreement properties of a CS are some function of the properties of the conjuncts and the CS as a whole. In the case of Portuguese, this agreement strategy gives rise to examples like (3).

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

Here, plural agreement has been triggered on the adjective *coloridos* because the preceding CS is plural (e.g. it denotes a plurality). Masculine gender has been triggered because the CS contains a masculine conjunct (masculine is the default resolution gender in Portuguese — leaving aside cases of CCA, feminine agreement is only possible if all conjuncts are feminine).

Under a CCA strategy, by contrast, rather than agreeing with the CS as a whole, agreement targets agree with just the *closest* conjunct. CCA is perhaps less familiar than resolution, but it is nevertheless widely attested. It has been observed in, *inter alia*, Irish, Welsh, Spanish, Arabic, and Ndebele. (e.g. McCloskey, 1986; Corbett, 1991; Sadler, 1999; Camacho, 2003; Moosally, 1999; Yatabe, 2004).

Though it does not seem to have been much discussed in the theoretical literature on Portuguese, the existence of this strategy has been noted in descriptive grammars of Portuguese. de Almeida Torres (1981) gives examples like (4):

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

Here we see that the postnominal adjective is feminine and singular, like the last conjunct, even though it semantically modifies the whole preceding CS (which contains a masculine noun, and so might be expected to trigger masculine agreement).

These examples involve postnominal agreement, which is what we focus on here. However, a few words about the behaviour of prenominal adjectives and determiners are in order. As regards gender, it seems that in Portuguese CCA is *required* for prenominal modifiers and determiners modifying coordinated nominals. For example, in (5) the presence of a masculine conjunct in the CS is not sufficient to permit masculine agreement on the prenominal adjective and noun, which must agree with the closest conjunct in gender.

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

As regards number, matters are less clear, and proper discussion would take us too far from the focus of this paper. Part of this complexity arises from the existence of 'single entity' readings of CSs (as in cases like *my friend and colleague*) which are semantically singular. Even leaving cases like this, there seems to be evidence of both CCA and resolution for number in Portuguese. Example (6) shows resolved number — a plural determiner and adjective with a CS which is semantically plural, though it consists of singular nominals (*prováveis* ('probable') is plural, but is not marked morphologically for gender); and (7) shows CCA for number — a singular determiner with a CS that is again semantically plural

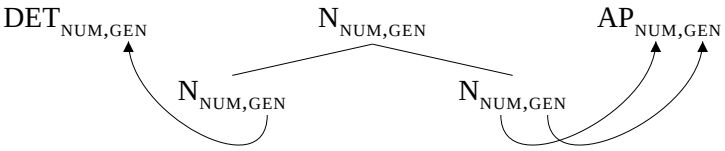
(6) *Os prováveis diretor e ator principal são*
the.MPL probable.PL director.MSG and actor.MSG principal.MSG are
Gus Van Sant e Johnny Deep, respectivamente
Gus Van Sant and Johnny Deep respectively
'the likely director and main actor are, respectively, Gus Van Sant and Johnny Deep'

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

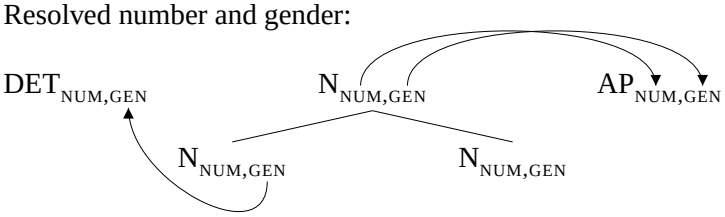
However, the issue is complex and somewhat controversial, and not essential to the main point of this paper, and we will not pursue it.[2]

To summarize: NP internally, Portuguese shows clear evidence of two agreement strategies involving CSs: CCA (postnominally, and prenominally as regards gender), and *resolution* (postnominally, and perhaps also prenominally for number). Leaving aside the matter of prenominal number, these might be represented schematically as in (8) and (9), respectively.

(8) CCA for number and gender:



(9) Resolved number and gender:



The existence of two patterns raises an obvious question about their relative frequency. As we have noted, CCA in Portuguese has not been much discussed in the literature, and one might wonder if this is because it is rare or marginal. In order to investigate this, a corpus study was undertaken, which will be described below, and whose quantitative results give a clear answer to this question (CCA is not rare or marginal). As it turns out, this study also raises (and answers) an interesting qualitative question, which has not previously been considered: are these the *only* patterns of agreement that are found? As will appear, some of the examples produced by the study seem to show the existence of 'mixed' agreement strategies, whose existence has not been previously noticed, and which have significant implications for the analysis of agreement with CS.

3. Corpus Study

This section reports the results of a corpus based study into the agreement strategies used for NP internal agreement involving CSs, focusing especially on gender agreement for post-nominal dependents.

In order to estimate the approximate frequencies with which the agreement strategies are used, a Web based corpus investigation was performed by means of searches using the Google API service.[3] Occurrences of coordinated nominals followed by adjectives were found by posing Google queries of the general form (10).

(10) `"<ART> * e <ART> * <ADJ>"`

Here `ART` stands for instances of the Portuguese (definite and indefinite) articles, `ADJ` stands for instances of Portuguese plural adjectives, and `e` is the Portuguese conjunction *e* ('and'). The adjectives were extracted from the 1,528,590 entry NILC Lexicon.[4]

Because we were interested in the correlation between the gender of each of the nominals 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 returned by the queries were manually inspected to remove noise — in cases of putative CCA this entailed removing all cases where, in the judgement of a Portuguese native speaker, the adjective should be interpreted as modifying only the the closest nominal, rather than the CS as whole.

The overall results found are displayed in Table 1, where 'Frequency' indicates the number of hits returned by Google for the searches, and 'N1', 'N2' and 'ADJ' refer to the gender of the first conjunct, second conjunct, and adjective, respectively.[5]

	Frequency	N1	N2	ADJ	Interpretation
(a)	0	f	m	f	(Resolve to f)
(b)	4054	m	f	m	(Resolve to m)
(c)	626	f	m	m	(CCA/Resolve to m)
(d)	550	m	f	f	(CCA)
total	5230				

Table 1. Frequency of Masc vs Fem Adjectives Modifying Mixed Gender Coordinations of Nominals.

The first thing to notice here is that there are *no* instances of a feminine

nominal conjoined with a masculine triggering feminine agreement (row (a)). That is, no instances of the form (11), which would be instances of resolution to feminine, or perhaps 'furthest conjunct agreement'. This is not particluarly surprising, but it supports our implicit assumption that cases of feminine gender agreement where a CS contains a masculine conjunct are indeed cases of CCA, and not some special 'resolution to feminine' strategy.

(11) [N_F conj N_M] ADJ_F

Similarly, row (b) is unsurprising. This row reports the count of cases which are schematically of the form in (12), where a conjoined masculine and feminine trigger masculine agreement. Leaving aside the possibility of 'furthest conjunct agreement', these are unambiguously cases of resolution to masculine, and they are very frequent (almost 80% of cases).

(12) [N_M conj N_F] ADJ_M

The cases counted in row (c), which are schematically like (13), are ambiguous — they might be either cases of resolution to masculine, or CCA with the masculine conjunct.

(13) [N_F conj N_M] ADJ_M

The most interesting case is row (d), which gives the number of cases of the form in (14). These are unambiguously cases of CCA (resolution would produce a masculine agreement on the adjective).

(14) [N_M conj N_F] ADJ_F

The interesting point is that they are not at all infrequent. Even on the narrowest interpretation, disregarding all ambiguous cases from row (c), CCA for gender is evidently widespread: the ratio of (d) cases to the total is 550/5230, or slightly over 10%. If these data are representative, the odds on speakers using CCA are better than 1 in 10. We can conclude that while resolution is the dominant strategy for postnominal gender agreement, CCA is by no means rare or marginal.

Apart from this quantitative finding, the study also threw up some unexpected qualitative results. Among these results were examples such as (15), which is schematically something like (16).

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

(16) DET_M [N_M conj N_F] ADJ_F

What this shows is CCA for gender both prenominally and postnominally,

with different effects. In this example, prenominal CCA has produced masculine agreement (recall that CCA for gender appears to be obligatory in Portuguese, so this cannot be a case of resolution to masculine on the determiner), at the same time, postnominal CCA has produced feminine agreement (resolved agreement would have made the adjective masculine).

Given that a language exhibits CCA, and has both prenominal and postnominal dependents, it is perhaps not surprising that this should occur. However, the possibility seems not to have been previously considered, and its existence is a significant result, with important theoretical implications, which we will take up below.

A second kind of case which appears not to have been previously noticed is exemplified in (17) to (21), which are schematically of the form (22).

(17) *todo o constrangimento e a dor*
all.MSG the.MSG embarrassment.MSG and the.FSG pain.FSG
sofridas
suffered.FPL
'all the embarrassment and pain suffered'

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

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

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

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

(22) [N_{MSG} conj N_{FSG}] ADJ_{FPL}

What these examples seem to show is postnominal CCA for gender (the adjective is feminine, like the last conjunct, even though the CS contains a

masculine nominal) combined, simultaneously, with resolution for number (the individual conjuncts are singular, but the adjective is plural). These cases raise an interesting theoretical issue, because not only has the existence of such cases not been previously noticed, it seems even to have been considered as a possibility. We will consider the theoretical implications of this in Section 4, below. Such cases also raise an interesting methodological issue, because though these are all attested examples, some native speakers of Portuguese are uncomfortable with them (not including the present author who is a native speaker of Brazilian Portuguese). In this context, it is worth looking at some other quantitative results.

Table 2 summarizes the number of examples found which involved coordinations of *singular* nominals (that is, a strict subset of the examples summarized in Table 1). Since all results feature plural adjectives, these are all cases of number resolution. The cases showing CCA for gender — that is, at least the cases in row (d) — thus show this 'mixed' agreement strategy of resolved number and CCA for gender. As the table shows, this strategy appears in 90 cases, which is approximately 4.6% of all the cases counted in Table 2, and about 4.9% of all the cases that could show this effect (i.e. all the cases where the final conjunct is feminine, i.e. rows (b) and (d)). This seems to us to be an interestingly large number, which combined with their acceptablity to some speakers means that the phenomenon deserves theoretical attention, and should not be dismissed out of hand (however, we will say a little more about the methodological issue raised here in Section 5).

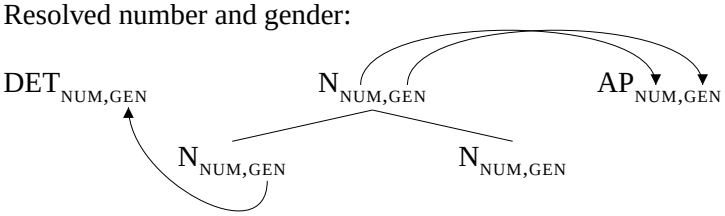
	Frequency	N1	N2	ADJ	Interpretation
(a)	0	f	m	f	(Resolve to f)
(b)	1737	m	f	m	(Resolve to m)
(c)	137	f	m	m	(CCA/Resolve to m)
(d)	90	m	f	f	(CCA)
total	1964				

Table 2. Frequency of Masc vs Fem Adjectives Modifying Mixed Gender Coordinations of *Singular* Nominals.

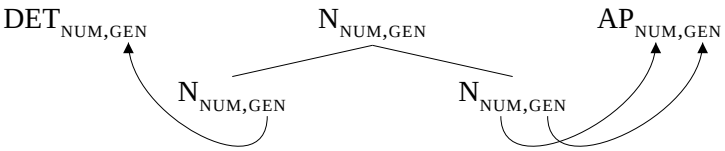
To summarize the results of this section: we have shown (a) that while gender resolution is the dominant agreement strategy postnominally, CCA is by no means infrequent or marginal, and (b) that Portuguese agreement is more complex that has been previously assumed. In particular, in addition to 'pure' resolution with prenominal CCA for gender, we also see prenomi-

nal and postnominal CCA operating independently, and a mixed postnominal strategy that involves CCA for gender with resolved number. Schematically, these strategies may be represented as in (23) to (25). The following section will consider the theoretical implications of this.

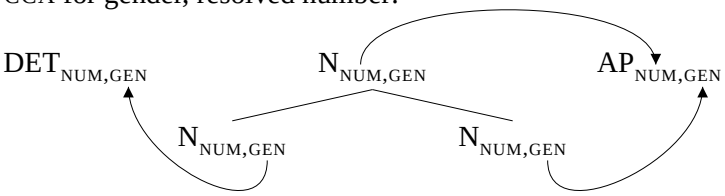
(23) Resolved number and gender:



(24) CCA for number and gender:



(25) CCA for gender, resolved number:



4. Linguistic Analysis and Theoretical Implications

In this section, we will consider some of the theoretical implications of the Portuguese data presented above, showing how an account of the data can be formulated. In the interests of generality, we will keep the presentation as intuitive and framework-neutral as possible.[6]

We will begin with resolution. In general, 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 (e.g. Dalrymple and Kaplan, 2000). So far as we can see, it is reasonable to assume that

number resolution in Portuguese is simply a matter of semantics: CS are plural just in case they denote a plurality or group of some kind. This is expressed in (26).[7]

(26) The number value on a CS resolves to plural just in case the CS denotes a plurality.

As regards gender, it seems safe to assume that masculine is the default resolution gender, or to put it another way, the resolved gender value on a CS is masculine if it contains one or more masculine conjuncts, and feminine only if all conjuncts are feminine:[8]

(27) The gender value on a CS resolves to feminine iff all conjuncts are feminine, otherwise it is masculine.

In principle, one might try to treat CCA in a similar fashion to resolution — a CS would have a single set of agreement properties calculated from properties of the conjuncts, but rather than involving calculations reflecting principles like (26) and (27), the calculation would simply return values from one designated conjunct (the last one, say). Such an approach might be unproblematic in a language which has only CCA. In a case like Portuguese which has both resolution and CCA, one might try to give every CS a single number and gender value, but allow the values to be calculated in one of two ways: either (a) by a resolution method, or (b) by a CCA method copying the value from (say) the last conjunct. Most existing approaches to agreement involve some kind of 'single feature' approach like this. Notice that such an approach predicts that all agreement processes will involve the same set of features.[9]

The Portuguese data clearly indicate that this sort of approach cannot be correct in general. First, the fact that prenominal CCA for gender is obligatory, while postnominally either CCA or resolution are possible indicate that CSs cannot be assigned a *single* agreement value: they need at least two sets of values, one for CCA, and one for resolution based agreement. Moreover, as we have seen, examples such as (15), repeated here, show CCA operating both prenominally and postnominally, with different effects. So we cannot manage with just one set of 'CCA agreement' features — we need two, one to for prenominal CCA, and one for postnominal CCA.

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

It seems the simplest way to approach this is to assume three sets of number/gender features: one reflecting resolved values, one for 'leftwards' CCA (i.e. CCA on prenominals), and one for 'rightwards CCA' (CCA on postnominals). Let use call these 'RESOL', 'LAGR' and 'RAGR'. The behaviour of these features will be governed by principles such as the following:

(29) The RESOL values of a CS are calculated from the features of the conjuncts, according to principles such as (26) and (27).

(30) The LAGR values of a CS come from its leftmost conjunct.

(31) The RAGR values of a CS come from its rightmost conjunct.

The existence of such features on CSs raises the question of what agreement features the conjuncts have, when they are not themselves CSs. One can imagine two approaches. The first would define features like LAGR, RAGR and RESOL only on CSs — 'normal' nominals would have only normal agreement features. But this is unattractive: it would complicate the statement of normal agreement principles, which would have to be different depending on whether the agreement control was a CS or not. It would also complicate the statement of the agreement percolation principles inside CSs. If instead, we assume that these features are defined on nominals of all kinds, a much simpler picture emerges.

To begin with, we need some principle like the following, to capture the fact that non-coordinate structures exhibit only one kind of agreement behaviour:

(32) In non-coordinate nominal structures the values of RAGR, LAGR and RESOL are identical.

(One way of implementing this would be to make it a lexical requirement of nouns, which is inherited by nominal projections; if noun phrases are analyzed as DPs, it would be stated as a requirement on Ds and their associated Ns that is inherited by DPs).

Now, (30) and (31) can be stated more precisely, and with complete generality, as (33) and (34):

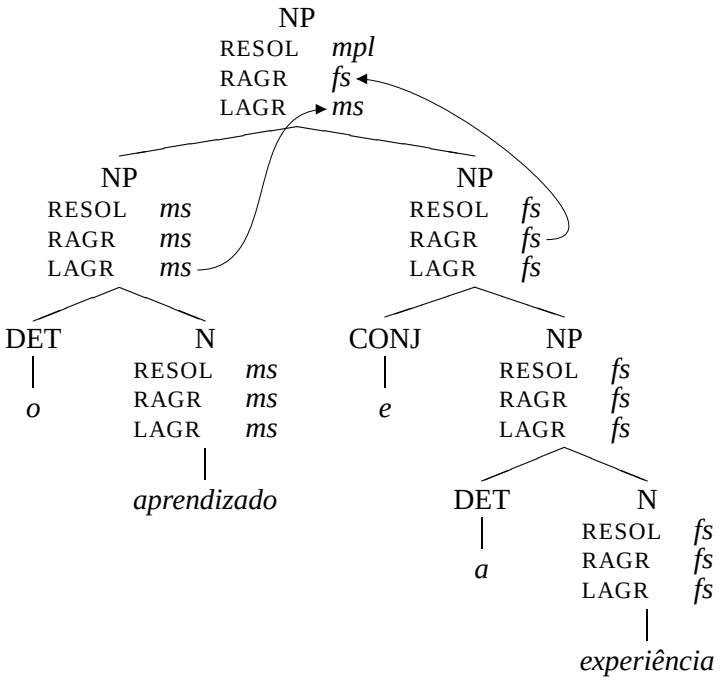
(33) The LAGR values of a CS are the LAGR values of its leftmost conjunct.

(34) The RAGR values of a CS are the RAGR values of its rightmost conjunct.

These principles can be seen at work in (36), representing the CS in (35).[10]

(35) *o aprendizado e a experiência*
the.MSG learning.MSG and the.FSG experience.FSG

(36)



Briefly, the lexical nouns *aprendizado* ('learning') and *experiência* ('experience') are lexically specified as masculine singular and feminine singular, and these values appear for all the agreement features. These same values appear on the non-CS nodes that dominate them, as required by (32). The mother node of the CS has LAGR *ms* from its left daughter and RAGR *fs* from its right daughter (as required by (33) and (34)). Its RESOL value is masculine because one of the daughters is masculine, reflecting (27); its RESOL number is plural because it denotes a plurality, reflecting (26).

Precisely how agreement is handled, given structures like (36), will depend on assumptions about the mechanics of determiner-noun and adjective-noun agreement, but the underlying principles will be roughly as follows:[11]

(37) Post-head modifiers must share either:
   a. their agreement controller's RESOL values (resolved agreement); or
   b. their agreement controller's RARG values ('full' CCA); or
   c. their agreement controller's RESOL.NUMBER and RARG.GENDER values ('mixed' CCA/resolution).

(38) Determiners and pre-head modifiers must share their agreement controller's LAGR.GENDER (CCA for gender)

The adjective *modernos* in (39) exemplifies (37a); the adjective *monástica* in (40) exemplifies (37b); *sofridas* in (41) exemplfies (37c); and *próprias* (‘own’) and *suas* (‘his/her’) in (42) exemplify (38).

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

(40) *estudos e profissão* ***monástica***
[studies.MSG and profession.FSG] monastic.FSG
‘monastic studies and profession’

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

(42) ***suas próprias*** *reações ou julgamentos*
his.FPL own.FPL [reactions.FPL or judgements.MPL]
‘his own reactions or judgements’

Notice that these principles also apply equally, and unproblematically, in the case of a non-coordinate agreement controllers (the account of agreement is thus uniform for CSs and non-coordinate structures, as one would wish). For example with in a noun like *teto* (‘ceiling’) in (43) all the principles in (37) produce exactly the same effect, because in a non-CS all the agreement features have the same values.

(43) *o teto colorido*
the.MSG ceiling.MSG coloured.MSG
‘the coloured wall’

Notice also that, as well as being ‘uniform’ in this sense, this account is consistent with a very standard idea of locality for agreement processes: percolation of features means that agreement can always be stated as a relation between an agreement controller and its sister(s).

Principles like those above appear to account for the data, and from a descriptive point of view they are attractive — they provide a simple conceptual and descriptive vocabulary for the analysis of Portuguese and other complex agreement systems. But there is clearly a theoretical cost in terms of the introduction of features which might not be otherwise required. However, it is worth pointing out that some proliferation of features seems to be required independently, because somewhat different features are required for handling NP-internal agreement processes, like those we have examined

here, and NP-external agreement processes like subject-predicate agreement. Familiar examples of this involve so-called 'hybrid nouns' (Corbett, 1991), which can trigger different kinds of agreement on different targets. 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 (cf. e.g. Corbett, 1991; Kathol, 1999; Wechsler and Zlatić, 2003):

(44) *Su Majestad*$_i$ *Suprema esta contento.*
Pron.F Majesty Supreme.F is happy.M.
'His Supreme Majesty is happy.'

In this context, it is interesting to ask whether Portuguese CCA might be 'NP-bounded' — a purely NP-internal process, which might limit the number of features required. Examples like the following suggest that it is not.

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

Here we see the CS *o travestismo e a copulação ritual* ('the transvestism and the ritual copulation') triggering plural agreement on the predicate (the verb *são* ('be') and the participle *realizadas* ('realized')), which would be consistent with a resolution strategy for number. However, we also see that *realizadas* is marked feminine — i.e. apparently agreeing with the closest conjunct *copulação ritual*. That is, subject predicate agreement may sometimes involve CCA.

An obvious objection to the analysis we have described is that it is stipulative, and does not really capture the fact that CCA is *closest* conjunct agreement. That is, the principles we have given could equally well be phrased so as to yield *furthest* conjunct agreement, which is not observed in Portuguese. However, furthest conjunct agreement *is* observed in some languages (e.g. Slovene Corbett, 1983). Moreover, notice that any account which tries to express CCA directly, as 'closest' conjunct agreement, will be in danger of losing one of the attractions of our account — the fact that it is consistent with standard ideas of locality. For example, an attempt to formulate such an account using any kind of conventional phrase structure will require agreement relations to hold between aunts and nieces, as well as sisters. Moroever,

attempts to deal with 'closeness' in terms of purely linear adjacency of agreement controllers and targets appear problematic: several of the examples we have given above involve CCA between determiners and nouns which are not adjacent (see, e.g. (5), (6), and (17)).[12]

5. Conclusion

The foregoing has presented some novel data and conclusions about Portuguese agreement. In particular, we have presented data which suggest that CCA is more widespread than has generally been assumed. We have also presented data which that agreement involving CSs is more complex that has been assumed, in ways that challenge existing analyses of agreement. In particular, we have argued that CSs do not possess a single set of agreement features (because *both* 'resolved' and 'closest conjunct' features are needed, and because information about the conjuncts at *both* ends of a CS may be needed for CCA). We have presented an analysis which captures these facts, and is consistent with a uniform treatment of agreement involving CS and non-coordinate structures.

The discussion involves what we take to be an interesting mix of 'empirical' (e.g. corpus based) and more traditional 'theoretical' linguistic investigation and analysis, a mix which is increasingly common, and productive. It also raises a number of methodological issues which deserve brief attention.

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 in corpora, it is essential to limit attention to cases where the interpretation makes it clear that the modifier scopes over the whole CS. Not only is there no conflict between corpus methods and methods based on 'native speaker intuition' here, both are actually necessary.

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 those above 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 — notably of course the fact that even large corpora do not typically show all possible variations and combinations of the phenomena one is interested in. Here one is naturally drawn to constructing examples. But this is not straight-

forward, because native speakers are often uncertain about the status of some examples.

In particular, it seems that some speakers reject examples involving post-nominal CCA for gender with resolved number (i.e. example like (17) to (21)).[13] Of course speakers' acceptability judgements are notoriously unreliable (cf. e.g. Schütze, 1996), especially judgements of *un*acceptability. And in fact, experience indicates that this sort of conflict between corpus data and intuition is rather rare. It is much more common for exposure to corpus data to persuade speakers that their intuition are over-restrictive.[14] But this just makes the problem harder to deal with when it arises. In the case of a web-based study such ours one cannot appeal to any pre-existing quality control (e.g. that the texts have been authored and proof-read by native speakers) One may observe, as we did above, that one has many examples of the relevant kind (in our case, 90). But how many is 'many'? In the case of web-based queries, there is no useful estimate of the total number of words in the corpus, but we found 4.9% of cases that could have shown the relevant pattern did show it (cf. Table 2). Is this a significant number? We are inclined to think that it is rather a large number to be just the result of 'noise' — that is, simple mistakes and the like. On the other hand, we note that the normal standard for statistical significance is 0.05, or 5%, so one could argue that it is statistically non-significant.

Bibliography

Camacho, J. (2003). *The Structure of Coordination: Conjunction and Agreement Phenomena in Spanish and Other Languages*. Kluwer Academic Publishers, Dordrecht.

Corbett, G. (1983). *Hierarchies, Targets and Controllers: Agreement Patterns in Slavic*. Croom Helm, London.

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.

Kathol, A. (1999). Agreement and the syntax-morphology interface in HPSG. In R. Levine and G. Green, eds., *Studies in Contemporary Phrase Structure Grammar*, pp. 209–260. Cambridge University Press, Cambridge and New York.

King, T. H. and M. Dalrymple (2004). Determiner agreement and noun conjunction. *Journal of Linguistics*, **40**:69–104.

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.

Wechsler, S. and L. Zlatić (2003). *The Many Faces of Agreement*. 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.

Chapter 1
Notes

1. The research was supported by the AHRB Project *Noun Phrase Agreement and Coordination*, MRGAN10939/APN17606. We are grateful for useful comments from many people, including: the anonymous referees for, and participants at, the LingEvid2006 conference held in Tuebingen in February 2006; participants at HPSG05 in Lisbon; participants at the 'Alliance 05 Project' Workshop held at Paris 7 in Oct 2005; and numerous colleagues at Essex, Mary Dalrymple and Irina Nikolaeva. See `http://privatewww.essex.ac.uk/~louisa/agr/NPagreement.html` for more information.
2. For example, King and Dalrymple (2004) claim that a singular determiner can only modify a CS with a 'single entity' ('boolean' or 'joint') interpretation, as in *o presidente e diretor da Air France* 'the.MSG president.MSG and director.MSG of Air France', where it is assumed that the president and director are one and the same individual. On the face of it, (7) is a counter-example to this claim.
   Another complicating issue is that the presence of material between the determiner and nominal may exert an influence on acceptability. In the (acceptable) example (6) the subject noun phrase is *Os prováveis diretor e ator principal* ('the probable director and main actor') with the adjective *prováveis* ('probable') intervening between the determiner and noun. Omitting it seems to have a deleterious effect, so **Os diretor e ator principal...* ('the director and main actor...') is judged unacceptable.
3. See `http://www.google.com/apis`.
4. See `http://www.nilc.icmc.usp.br/nilc/index.html`.
5. One interesting point which we will not pursue is that the figures seem to show a strong bias for masculine conjuncts to precede feminine conjuncts (feminine conjuncts precede in only 626/5230 cases). This is probably a reflection a prescriptive bias in favour of this ordering of conjuncts.
6. For a fully worked out formal treatment, see Villavicencio et al. (2005).
7. An example of a CS which does not denote a plurality is given in note 2. A counter-example to our assumption would be a CS containing a plural

nominal that triggered singular agreement, where this could not be analyzed as a case of CCA. In fact, we have not found any examples of CSs involving plural nominals triggering singular agreement at all.

8. A counter-example would be a CS which contains a masculine conjunct, but triggers feminine agreement, where this cannot be attributed to CCA. As noted in Section 3, no such cases were found in our study.
9. In fact, the 'single feature' approach is already known to be inadequate in other languages. Sadler (2003, 1999) shows that it will not work in Welsh, where different agreement processes can target the resolved and the CCA features at the same time, indicating that a CS must be able to have both resolved and CCA features simultaneously. However, Sadler suggests that any one agreement process can only access one kind of feature. The Portuguese data suggest that this is over-restrictive.
10. This representation makes a number of assumptions about the analysis of CSs, e.g. that the conjunction forms a constituent with the final daughter, and that the CS is an NP, rather than (say) a CONJP; none of these assumptions is critical.
11. This formulation evades the issue of number agreement for prenominal adjuncts — we leave open the question of whether they show resolution or CCA for number (or indeed both). Nothing we say hangs on this.
12. It is true that what intervenes may be an adjective which also agrees with the noun, but this is irrelevant: the adjective is not the agreement controller for the determiner.
13. Notice that this is the only case that is problematic in this way. All speakers seem happy with cases of prenominal CCA with postnominal resolution, and cases where pre- and post-nominal CCA give different effects. Thus, the main theoretical claims of the paper are not affected by this issue about data.
14. The following is a simple and uncontroversial example of this. It has sometimes been claimed that *alternately* cannot be used with *or*, cf. *John was alternately hot and/*or cold*. Many speakers accept this judgement at first glance. However, a search of the British National Corpus yields several examples of *alternately . . . or* which seem to be fully acceptable to all speakers, and which lead them to revise their judgement — e.g. *[they] spent almost three hours in each other's arms, alternately making love or talking in low whispers*.