Portuguese: Corpora, Coordination and Agreement

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

[†]Department of Language and Linguistics,
University of Essex, UK

{louisa,doug}@essex.ac.uk

[*]Instituto de Informática,
Universidade Federal do Rio Grande do Sul, Brazil

avillavicencio@inf.ufrgs.br

Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives,
Tübingen, February 2006

Outline

Aims

- Results of a corpus-based study of agreement in Portuguese:
  - Coordinate Structures (CSs)
  - Inside NP.
- Implications for understanding of agreement processes;
- Outline of formal analysis (HPSG);
- Methodological issues.

Outline (1)

1 Background

Agreement

'Systematic covariation of linguistic forms'.

(1) a parede colorida (Portuguese)
the$_{fsg}$ wall$_{fsg}$ coloured$_{fsg}$

(2) o teto colorido
the$_{msg}$ ceiling$_{msg}$ coloured$_{msg}$

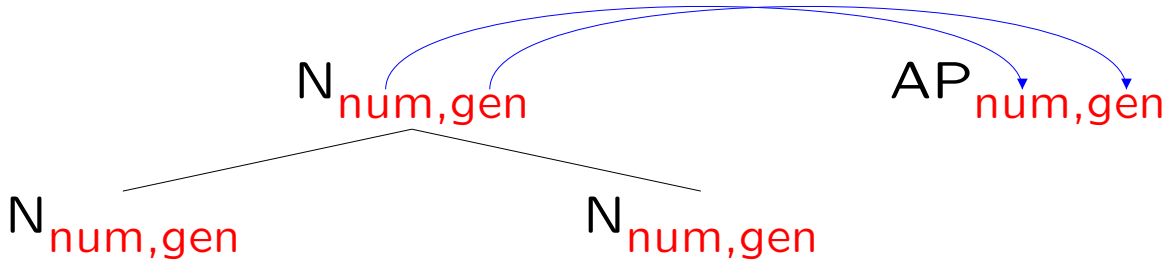
(3) a/*as parede colorida/*coloridas
the$_{fsg/*fpl}$ wall$_{fsg}$ coloured$_{fsg/*fpl}$

(4) o/*os teto colorido/*coloridas
the$_{msg/*mpl}$ ceiling$_{msg}$ coloured$_{msg/*mpl}$

Agreement in Coordinate Structures

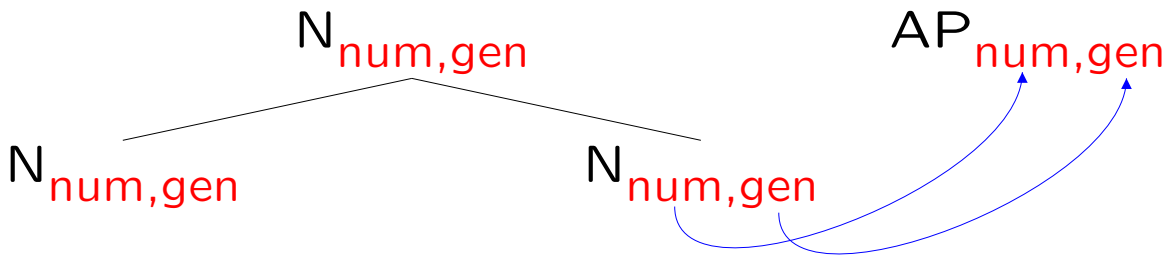
Crosslinguistically, two agreement strategies are observed:

Resolution: Agreement marking on modifiers is some function of the marking on conjuncts (e.g. modifiers are masculine if any conjunct is masculine);



Closest Conjunct Agreement (CCA): modifiers agree with one conjunct — the closest:

- McCloskey (1986); Corbett (1991); Sadler (1999); Moosally (1999); Yatabe (2004) . . .
- Irish, Welsh, Arabic, Ndebele, Spanish, Portuguese,. . .



One approach to CCA would be 'resolution-like' — a CS would still have a single set of agreement properties calculated from properties of the conjuncts, but the calculation would (trivially) return values from one conjunct.

But this sort of approach is problematic:

- Different agreement processes can target the resolved and closest conjunct features at the same time (Sadler, 2003, 1999).
- Our study suggests that the same agreement process can target different features at the same time.
- CSs must make several kinds of agreement information available simulataneously.

Outline (2)

2 Agreement with Portuguese Coordinate Structures

A corpus study was undertaken to investigate the range and approximate frequency of agreement patterns associated with Portuguese coordinate structures (cf. Villavicencio et al., 2005).

The study focused on NP-internal agreement patterns with post-nominal plural adjectives (aimed at investigating gender controlled by the closest conjunct).

Method

Examples were obtained by Google searches for occurrences of coordinated NPs followed by plural adjectives:

‘<ART> * e <ART> * <ADJ>’

- ART = Portuguese (definite and indefinite) articles
- ADJ = Adjectives
  - extracted from the 1,528,590 entry NILC Lexicon (http://www.nilc.icmc.usp.br/nilc/index.html);
  - only adjectives that overtly reflect gender distinctions
    - 9,915 masculine, and
    - 9,811 feminine
- Only exact matches.

Results were manually inspected to remove noise (e.g. cases where adjective scopes over only one noun).

Results

In combination with examples from descriptive grammars, the study shows that mixed gender coordinate structures can trigger different agreement patterns on different targets inside NP.

Less surprisingly:

- Postnominally:
  - Resolution (for NUMBER and GENDER);
  - CCA (for NUMBER and GENDER);
- Prenominally (Determiners and Prenominal Adjectives)
  - CCA (at least for GENDER).

More surprisingly:

- CCA pre- and post- nominally *with different effects*
- Postnominally, resolution for NUMBER, CCA for GENDER.

Postnominal Resolution for NUM and GEN

(5) o teto e a parede coloridos
the$_{msg}$ ceiling$_{msg}$ and the$_{fsg}$ wall$_{fsg}$ coloured$_{mpl}$
the coloured ceiling and wall

(6) o homem e a mulher solteiros/modernos
the$_{msg}$ man$_{msg}$ and the$_{fsg}$ woman$_{fsg}$ solitary$_{mpl}$/modern$_{mpl}$
the solitary/modern man and woman

(Resolution to feminine seems not to occur, as one would expect.)

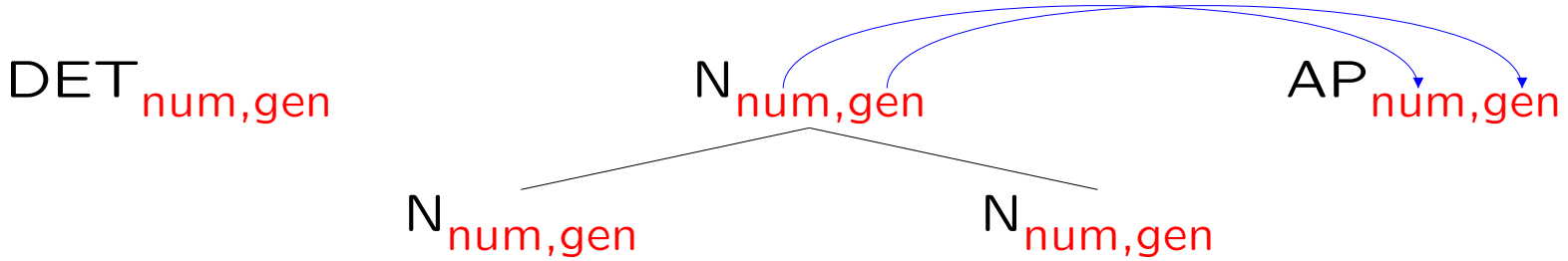


Figure 1: Resolution for Number and Gender

Postnominal CCA

(7) estudos e profissão monástica
studies$_{mpl}$ and profession$_{fsg}$ monastic$_{fsg}$
monastic studies and profession

(8) no povo e gente hebreia
on the$_{msg}$ people$_{msg}$ and persons$_{fsg}$ hebrew$_{fsg}$
(de Almeida Torres, 1981)

- The existence of CCA is briefly noted in descriptive studies.
- CCA is more widespread and common than generally thought.
- There seem to be no cases of 'furthest' conjunct agreement.

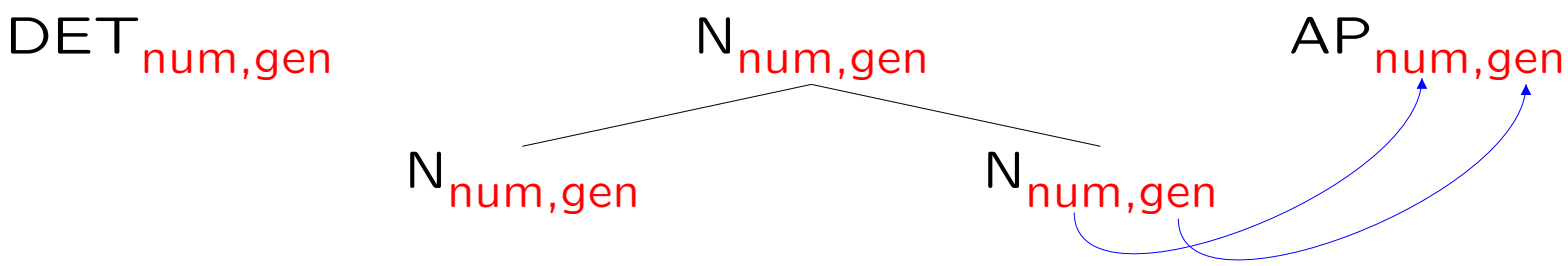


Figure 2: CCA for Number and Gender

Prenominal CCA

(9) suas próprias reações ou julgamentos
his$_{fpl}$ own$_{fpl}$ reactions$_{fpl}$ or judgements$_{mpl}$
his own reactions or judgements

(10) diversas secções ou subgrupos
diverse$_{fpl}$ sections$_{fpl}$ or subgroups$_{mpl}$
various sectors or subgroups

(11) a correcta gestão e preservação
the$_{fsg}$ correct$_{fsg}$ management$_{fsg}$ and conservation$_{fsg}$

- CCA for GENDER (i.e. no resolution for GENDER).
- Possibly CCA for NUMBER.

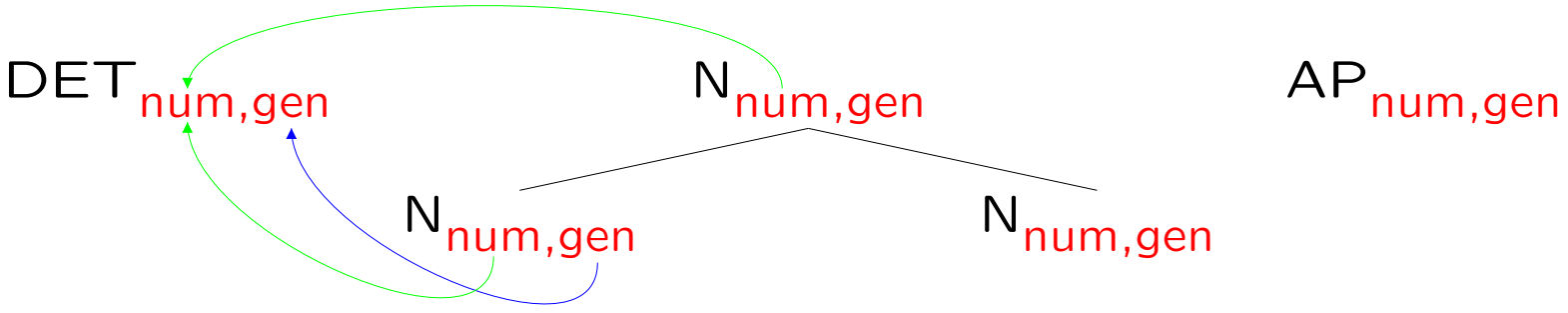


Figure 3: Prenominal CCA for Gender

Simultaneous Pre- and Post-nominal CCA

Different forms of agreement can be triggered on the prenominal and postnominal modifiers at the same time, *with different effects*.

E.g. Masculine agreement on the determiner and feminine on the postnominal adjective:

(12) Esta canção anima os$_{mpl}$ corações$_{mpl}$ e mentes$_{fpl}$ brasileiras$_{fpl}$.
This song animate the hearts and minds Brazilian

- CCA to the right (postnominally);
- CCA to the left (prenominally).

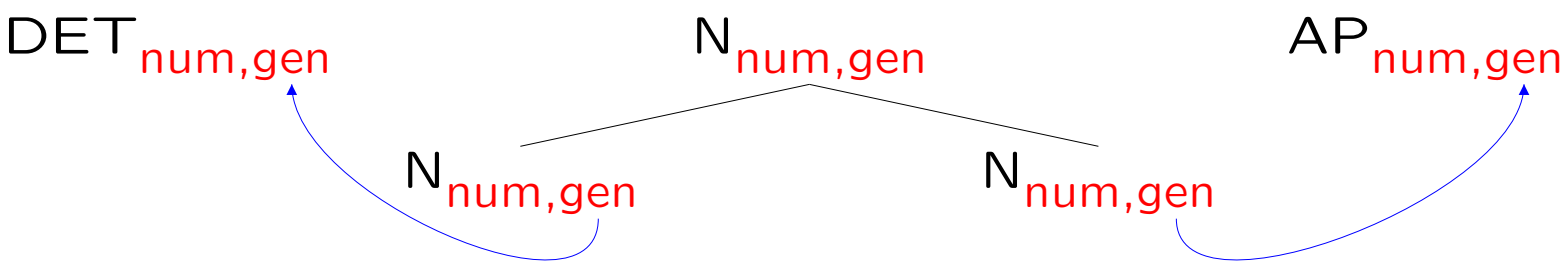


Figure 4: Prenominal *and* Postnominal CCA (for Gender)

Resolution for NUMBER, CCA for GENDER

(13) todo o constrangimento e a dor sofridas
all$_{msg}$ the$_{msg}$ embarrassment$_{msg}$ and the$_{fsg}$ pain$_{fsg}$ suffered$_{fpl}$
all the embarrasment and pain suffered

(14) o drama e a loucura vividas
the$_{msg}$ drama$_{msg}$ and the$_{fsg}$ madness$_{fsg}$ lived/felt$_{fpl}$
the drama and the madness felt

(15) . . . o aprendizado e a experiência vividas . . .
. . . the$_{msg}$ learning$_{msg}$ and the$_{fsg}$ experience$_{fsg}$ lived/felt$_{fpl}$ . . .

• This 'mixed' pattern appears not to have been observed before (and some speakers reject it).

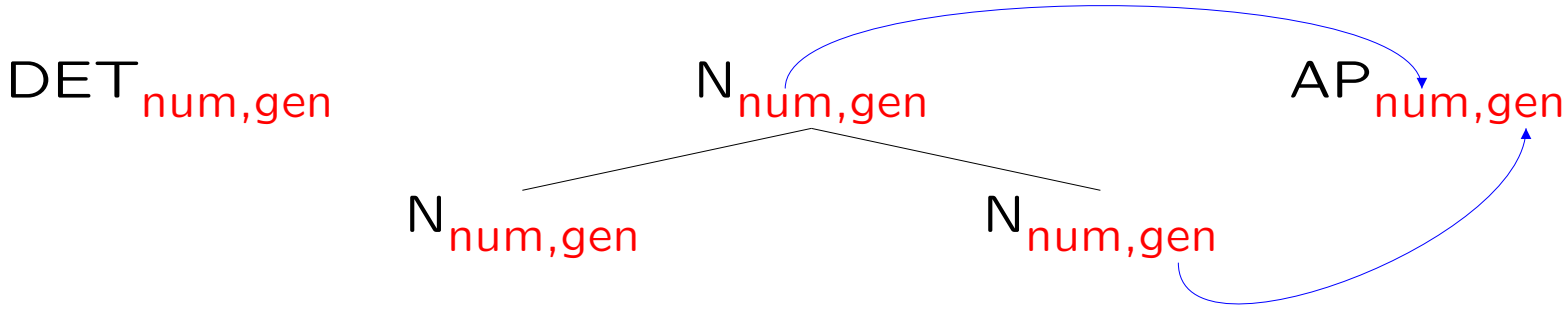
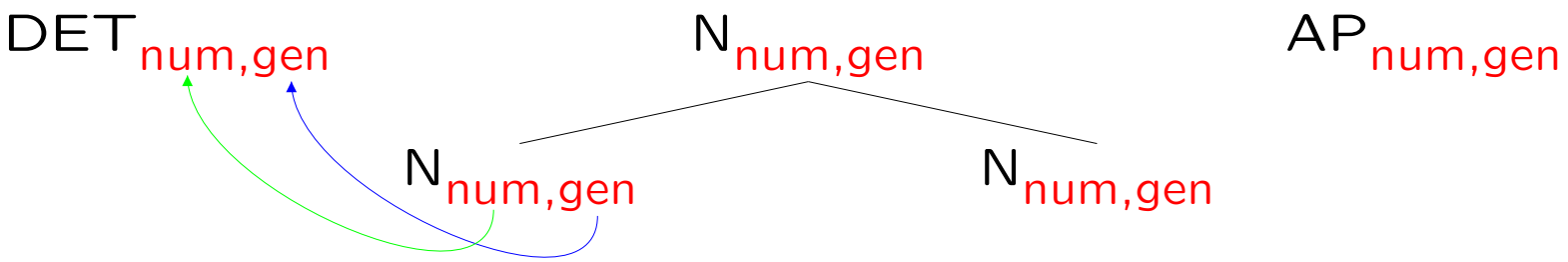


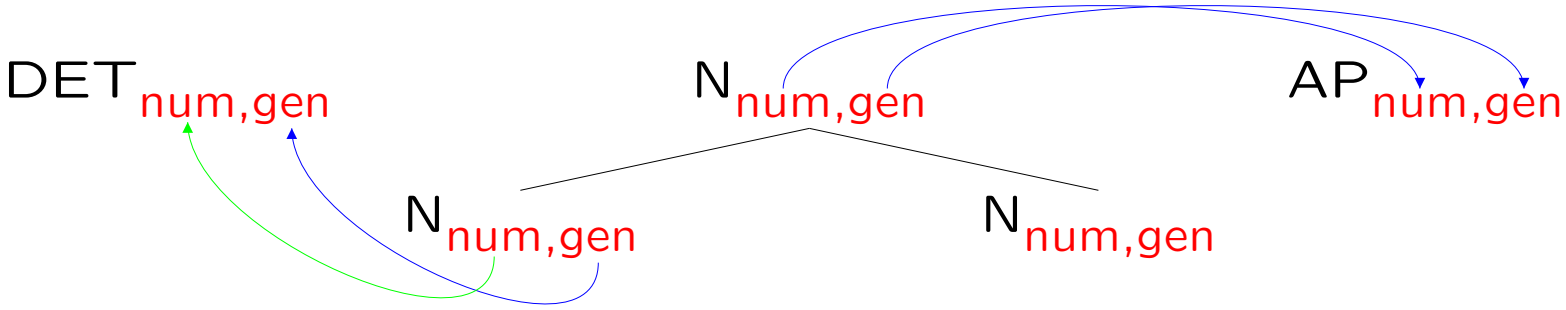
Figure 5: CCA for Gender, Resolved Number

Summary

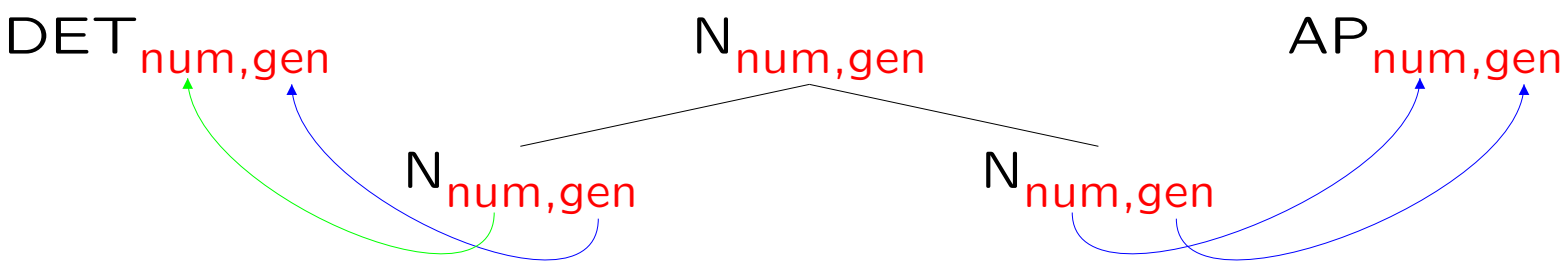
Prehead CCA for Gender (?Number):



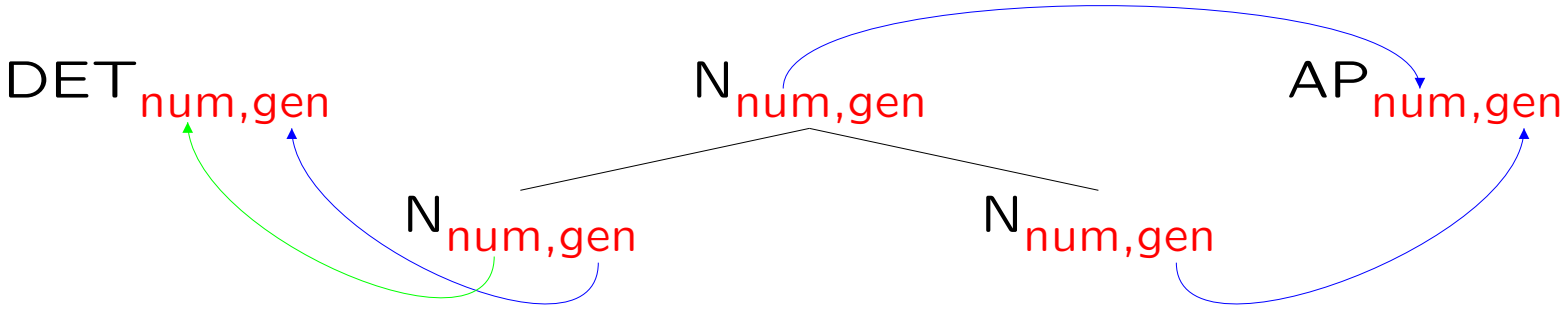
Resolved Number and Gender:



CCA for Number and Gender:



CCA for Gender, Resolved Number:



Some Numbers

- In all cases the adjective has been judged to scope over both nouns.
- All cases show number resolution (the postnominal adjective is plural, even when both NPs are singular).

	Frequency	NP1	NP2	Adj	
(a)	0	f	m	f	(resolve to f)
(b)	626	f	m	m	(cca/resolve to m)
(c)	550	m	f	f	(cca)
(d)	4054	m	f	m	(resolve to m)
total	5230				

Table 1: Frequency of Masc vs Fem Adj Modifying Mixed Gender Coordinate NP.

- Even on the narrowest interpretation, CCA for gender is widespread:
  (c)/total = 550/5230 $\approx$ 10%
  — if these data are representative, the odds on speakers using CCA are better than 1 in 10.

The figures for cases of *singular* NPs are particularly interesting with respect to the 'mixed strategy'.

	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 Adj Modifying Mixed Gender Coordinate NP (Singular NPs).

In Table2, the conjuncts are singular, but the adjectives are plural (i.e. we have resolved number). Thus, row (c) shows CCA for gender, with resolved number — the 'mixed' strategy (90 cases, $\approx$ 4.6% of the total)

Implications

- Agreement with CSs is more complex that has been thought.
- These data are problematic for most existing analyses of agreement.
- CSs do not possess a single set of agreement features (we need *both* resolved and CC features).
- As regards CCA, we need information about the conjuncts at *both* ends of the CS.

Outline (3)

3 Formalization

CSs make available:

- agreement properties from left conjunct;
- agreement properties from right conjunct;
- resolved agreement properties.

Features

Agreement information is stored in the values of three HEAD features

LAGR: for the leftmost conjunct;
RAGR: for the rightmost conjunct.

Plus:

CONCORD: for 'resolved' information.

LAGR and RAGR are defined on all sorts where CONCORD is defined; and 'normally' (e.g. in headed constructions) all three features share values:

(16) *noun* $\wedge$ *lexical* $\rightarrow$ $\left[\begin{array}{ll} \text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\,|\,\text{HEAD} & \left[\begin{array}{ll} \text{LAGR} & \boxed{1} \\ \text{RAGR} & \boxed{1} \\ \text{CONCORD} & \boxed{1} \end{array}\right] \end{array}\right]$

In non-headed constructions, in particular, in coordinate structures, this identity breaks down. Instead:

LAGR/RAGR

- LAGR comes from the LAGR of the leftmost daughter,
- RAGR from the RAGR of the rightmost daughter;

CONCORD

- is 'resolved' from the CONCORD values of the daughters.

Percolation

LAGR, RAGR

Coordinate phrases which are defined for LAGR, RAGR and CONCORD (e.g. nominal-coordinated-phrases) satisfy:

(17)

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

CONCORD (Resolution)

Different principles operate for

- NUMBER, and
- GENDER

NUMBER Resolution

Number resolution is just semantics:

- INDEX|NUM is plural whenever an NP denotes a plurality;
- CONCORD just reflects this

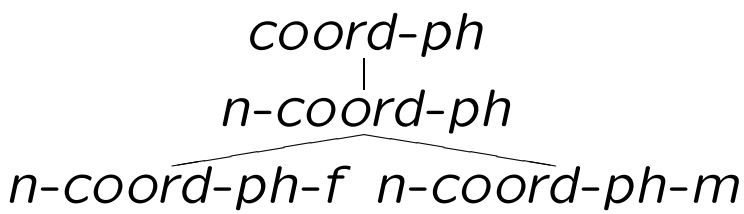
All *head* values (including those on coordinate structures) satisfy:

(18)
$$\textit{head}\begin{bmatrix}\text{CONTENT}\,|\,\text{INDEX}\,|\,\text{NUM} & \boxed{1}\\ \text{CONCORD}\,|\,\text{NUM} & \boxed{1}\end{bmatrix}$$

GENDER Resolution

Resolution for GENDER is more complex.

(19)



n-coord-ph $\rightarrow$

$$\textit{ncpf}\left[\begin{array}{ll}\text{SS}\,|\,\text{L}\,|\,\text{CAT}\,|\,\text{HD}\,|\,\text{CONCORD}\,|\,\text{GEND} & \boxed{3} \\ \text{CONJ-DTRS} & \left\langle\left[\text{SS}\,|\,\text{L}\,|\,\text{HD}\,|\,\text{CONCORD}\,|\,\text{GEND} \quad \boxed{3}\ \text{fem}\right]^{*}\right\rangle\end{array}\right]$$

$$\vee$$

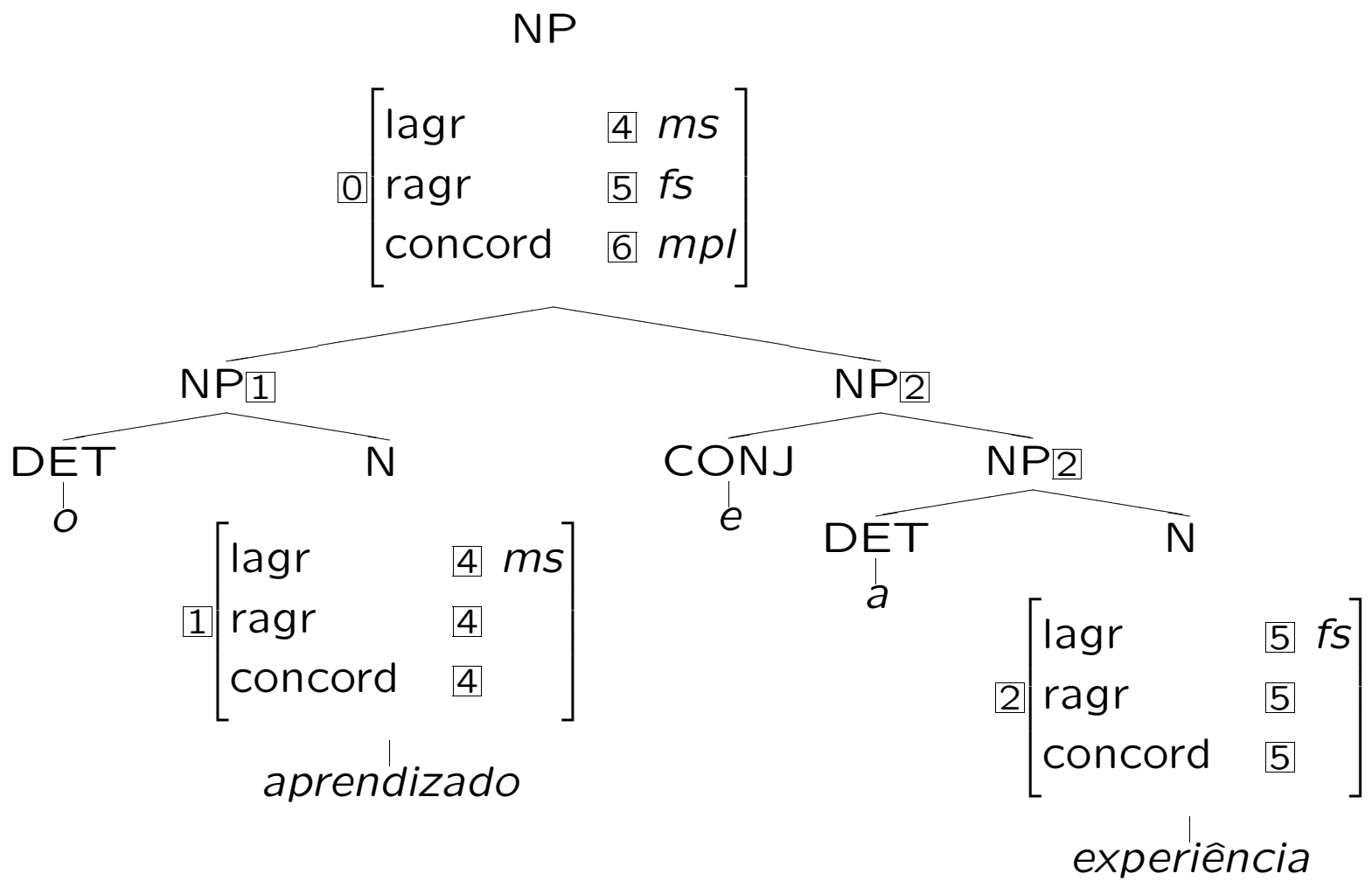
$$\textit{ncpm}\left[\begin{array}{ll}\text{SS}\,|\,\text{L}\,|\,\text{CAT}\,|\,\text{HD}\,|\,\text{CONCORD}\,|\,\text{GEND} & \boxed{3} \\ \text{CONJ-DTRS} & \left\langle .^{*},\ \left[\text{SS}\,|\,\text{L}\,|\,\text{HD}\,|\,\text{CONCORD}\,|\,\text{GEND} \quad \boxed{3}\ \text{masc}\right],\ .^{*}\right\rangle\end{array}\right]$$

- GENDER in a coordinate structure resolves to feminine just in case all conjunct daughters are feminine,. . .
- and to masculine if there is a single masculine daughter.

Example

(20)	. . . o	aprendizado	e	a	experiência	(vividas)	. . .
	. . . the$_{msg}$	learning$_{msg}$	and	the$_{fsg}$	experience$_{fsg}$	(lived$_{fpl}$)	. . .

(21)


NP
0
lagr 4 ms
ragr 5 fs
concord 6 mpl
NP1
NP2
DET
N
CONJ
NP2
o
e
DET
N
a
1
lagr 4 ms
ragr 4
concord 4
aprendizado
2
lagr 5 fs
ragr 5
concord 5
experiência

Agreement

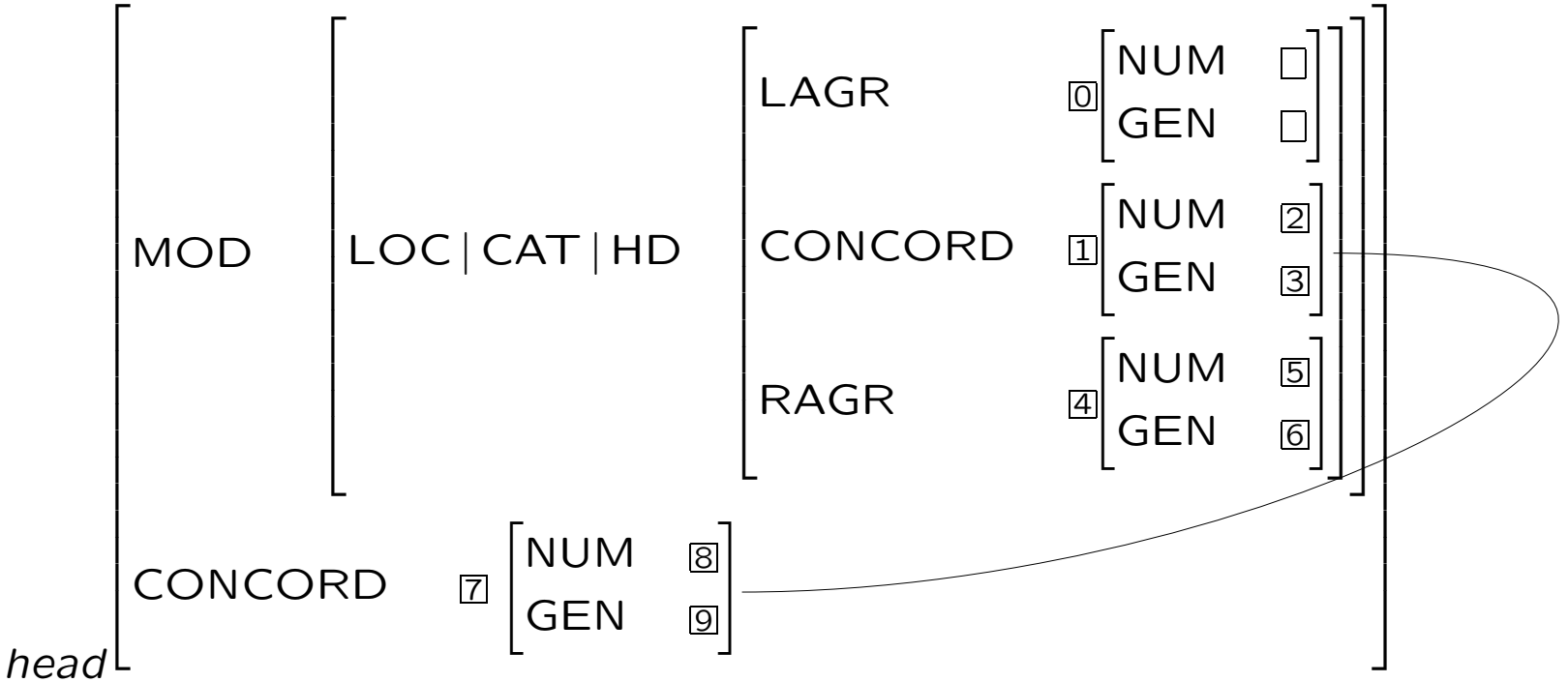
- Posthead:
  - CCA for NUMBER and GENDER; or
  - Resolution for NUMBER and GENDER; or
  - Resolution for NUMBER, CCA for GENDER.
- Prehead (Determiners and Prenominal Adjectives)
  - CCA (at least for GENDER).

Adjectives will satisfy one of the following constraints on HEAD values:

$$\textit{head}\begin{bmatrix}\text{MOD} & \begin{bmatrix}\text{LOC}\,|\,\text{CAT}\,|\,\text{HD} & \begin{bmatrix}\text{LAGR} & \boxed{0}\begin{bmatrix}\text{NUM} & \square\\ \text{GEN} & \square\end{bmatrix}\\ \text{CONCORD} & \boxed{1}\begin{bmatrix}\text{NUM} & \boxed{2}\\ \text{GEN} & \boxed{3}\end{bmatrix}\\ \text{RAGR} & \boxed{4}\begin{bmatrix}\text{NUM} & \boxed{5}\\ \text{GEN} & \boxed{6}\end{bmatrix}\end{bmatrix}\end{bmatrix}\\ \text{CONCORD} & \boxed{7}\begin{bmatrix}\text{NUM} & \boxed{8}\\ \text{GEN} & \boxed{9}\end{bmatrix}\end{bmatrix}$$

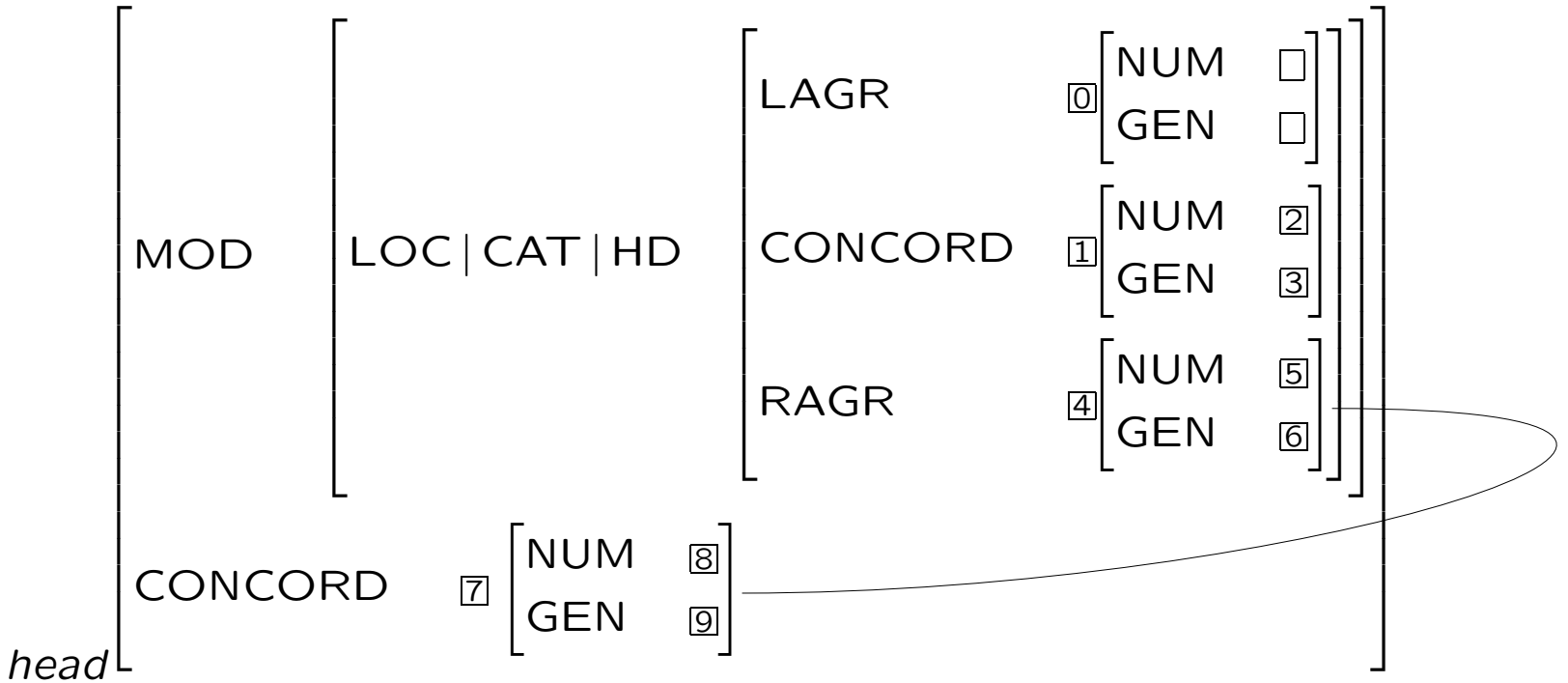
1. $\boxed{7} = \boxed{1}$ (resolution for NUM and GEN)
2. $\boxed{7} = \boxed{4}$ (CCA with last conjunct, NUM and GEN)
3. $\boxed{8} = \boxed{2}$ (resolved number)
   $\boxed{9} = \boxed{6}$ (CCA with last conjunct for gender)
4. $\boxed{7} = \boxed{0}$ (CCA with first conjunct, GEN (?and NUM))

1. $\boxed{7} = \boxed{1}$ (resolution for NUM and GEN):



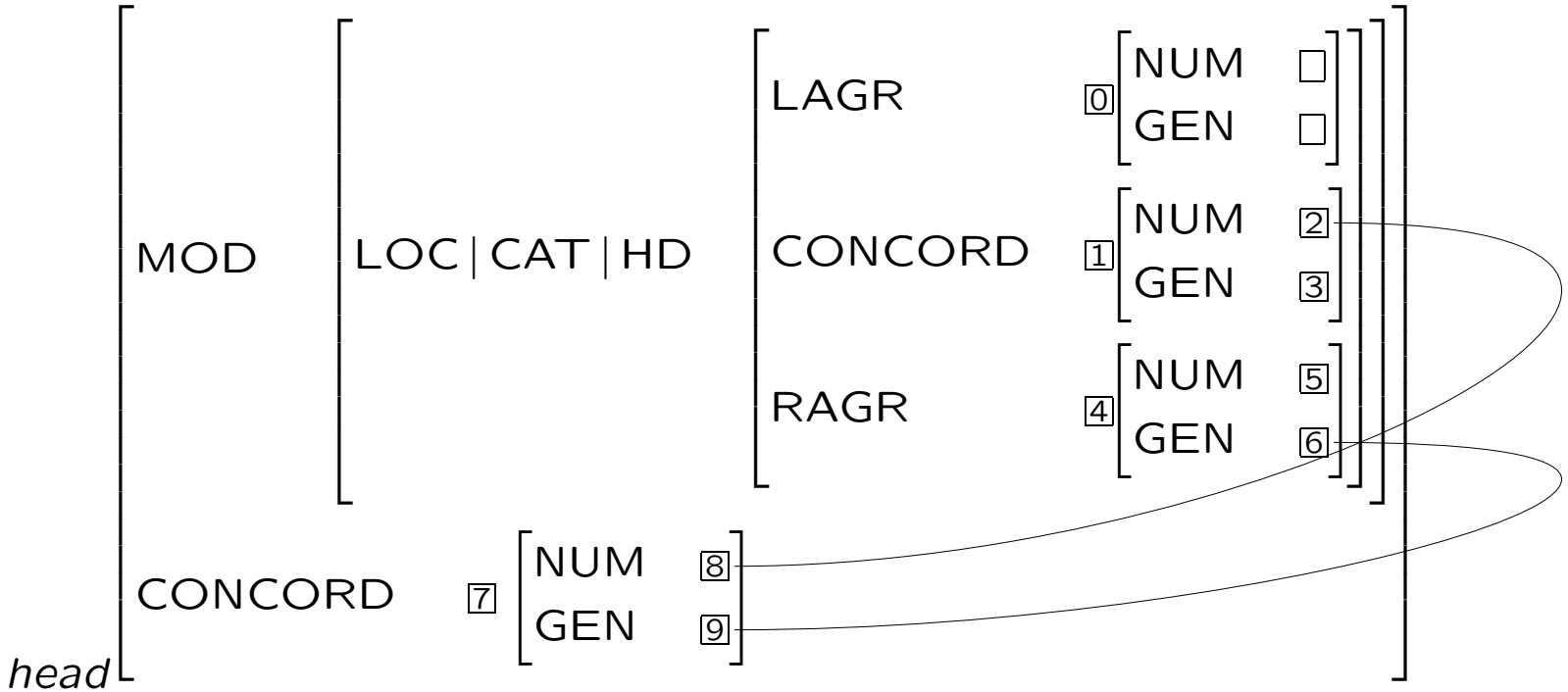
(22) o homem e a mulher **modernos**
the$_{msg}$ man$_{msg}$ and the$_{fsg}$ woman$_{fsg}$ modern$_{mpl}$
the modern man and woman

2. $\boxed{7} = \boxed{4}$ (CCA with last conjunct, NUM and GEN):



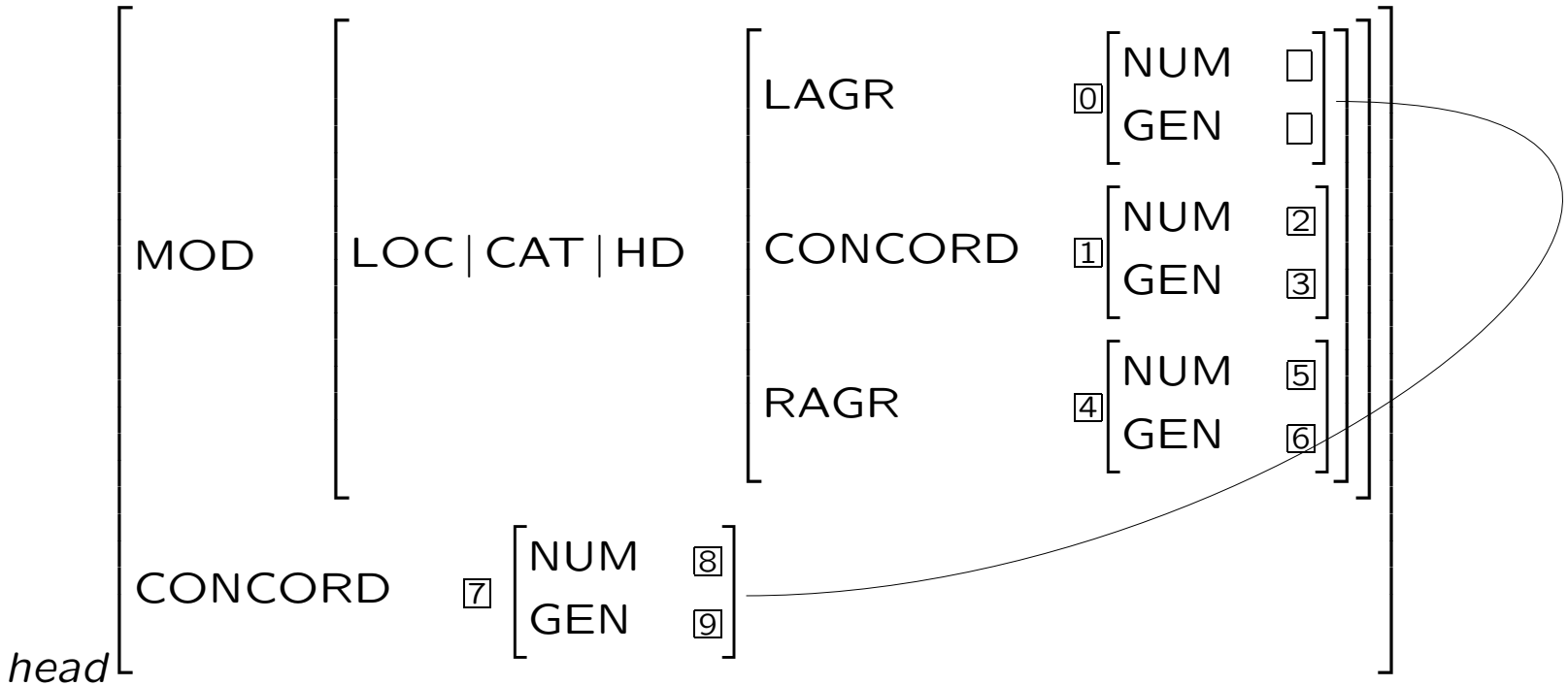
(23) estudos e profissão **monástica**
studies$_{msg}$ and profession$_{fsg}$ monastic$_{fsg}$
monastic studies and profession

3. $\boxed{8} = \boxed{2}$ (resolved number)
   $\boxed{9} = \boxed{6}$ (CCA with last conjunct for gender):



(24) todo o constrangimento e a dor **sofridas**
all$_{msg}$ the$_{msg}$ embarrassment$_{msg}$ and the$_{fsg}$ pain$_{fsg}$ suffered$_{fpl}$

4. $\boxed{7} = \boxed{0}$ (CCA with first conjunct, GEN (?and NUM)):



(25) suas **próprias** reações ou julgamentos
his$_{fpl}$ own$_{fpl}$ reactions$_{fpl}$ or judgements$_{mpl}$
his own reactions or judgements

Outline (4)

4 Methodological Issues

- Synergy of theoretical and corpus based study;
- Necessity of *interpreted* data:
  - scope of modification;
  - 'boolean' interpretation of CSs.
- Value of corpus data:
  - importance of CCA had been underestimated
  - previously unobserved examples of 'bi-directional CCA'
  - previously unobserved examples of 'mixed strategy'
- Limitations of corpus data:
  - almost 5% of examples show 'mixed strategy' . . . but some speakers reject them. So. . .
  - lack of examples relating to CCA for number prenominally:

(26) a correcta gestão e preservação
the$_{fsg}$ correct$_{fsg}$ management$_{fsg}$ and conservation$_{fsg}$
'the correct management and conservation'

Outline (5)

5 Conclusions/Summary

Summary

- A detailed description of some aspects of the phenomena;
- Novel data: 'bidirectional CCA', mixtures of Resolution and CCA;
- An extensive empirical study, showing that CCA is more frequent than is generally thought;
- An analysis whereby coordinate structures make available:
  - agreement properties from left conjunct;
  - agreement properties from right conjunct;
  - resolved agreement properties.
- Formalization: two new features, with percolation principles.

Open Questions

- What factors influence choice of CCA? (e.g. animates seem to favour resolution)
- Prenominal CCA for number?
- Other analyses (especially with respect to linearity effects)
- Syntactic persistence of features outside NP:

(27)

o	travestismo	e	a	copulação	ritual	são	realizadas		
the$_{msg}$	transvestism$_{msg}$	and	the$_{fsg}$	copulation$_{fsg}$	ritual	be$_{pl}$	realized$_{fpl}$		
para	expressar	...							
to	express	...							

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

http://privatewww.essex.ac.uk/~louisa/agr/NPagreement.html

References

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.

McCloskey, James. 1986. Inflection and Conjunction in Modern Irish. *Natural Language and Linguistic Theory* 4(2), 245–282.

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 Publications.

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.

Villavicencio, Aline, Sadler, Louisa and Arnold, Doug. 2005. An HPSG Account of Closest Conjunct Agreement in NP Coordination in Portuguese. In Stephan Müller (ed.), *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Lisbon*, Stanford, CA: CSLI Publications.

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.