

An HPSG Account of Closest Conjunct Agreement in NP Coordination in Portuguese

Aline Villavicencio, Louisa Sadler, and Doug Arnold
Department of Language and Linguistics
University of Essex
{avill, louisa, doug}@essex.ac.uk

HPSG-05, Lisbon
Aug 24, 2005

Outline

- 1 Aims
- 2 Agreement in HPSG
- 3 Agreement in Coordinate Structures
- 4 Agreement in Portuguese NPs
- 5 Portuguese Coordinate Structures
- 6 Data from a Corpus Study
- 7 HPSG Analysis
- 8 Conclusions/Summary
- 9 Open Questions

1 Aims

- Data about Agreement in Portuguese
 - Coordinate Structures
 - Inside NP
- Results of a corpus study;
- HPSG analysis of data.

2 Agreement in HPSG

Agreement has received considerable attention from the HPSG community, in e.g.

- Pollard and Sag (1994);
- Kathol (2000);
- Wechsler and Zlatic (2003);
- Yatabe (2004);
- Abeillé (2003);
-

A multilayered theory of agreement that allows different agreement relations to hold between different objects simultaneously:

CONCORD:

- relates to the inflected form of nouns;
- includes information about case, number, and gender;
- relevant for agreement between e.g. determiners and nouns.

INDEX:

- relates to semantic characteristics of the noun;
- contains information about number, gender and person;
- relevant for agreement between e.g. subjects and verbs.

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.

Example: in languages like Spanish and French, if a title like ‘Majesty’ refers to a male, it triggers

- masculine agreement on a predicative adjective, but
- feminine agreement on an attributive adjective.

(1) Su Majestad_{*i*} Suprema esta contento. (Éli_{*i*} ...)
His Majesty Supreme_{*fem*} is happy_{*masc*}. (He_{*masc*} ...)

3 Agreement in Coordinate Structures

Similarly intricate patterns of agreement arise with coordinate structures. In particular, where conjuncts differ in some agreement related properties, two strategies are widely attested:

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 only one conjunct — the closest one;

- Corbett (1991), Sadler (1999), Moosally (1999), Yatabe (2004), ...
- languages like Spanish, Arabic, Ndebele, ...

Moosally (1999) proposes a treatment of partial agreement in Ndebele:

- agreement constraints are defined in a multiple inheritance hierarchy;
- the CCA constraint is defined as:

$$\left[\begin{array}{l} \text{SS | LOC | CAT | HEAD | CONT} \quad \left[\text{INDEX} \quad \left[\text{GENDER} \quad \boxed{1} \right] \right] \\ \text{CONJ-DTRS} \quad \left\langle \dots, \left[\dots \mid \text{INDEX} \mid \text{GENDER} \quad \boxed{1} \right] \right\rangle \end{array} \right]$$

Yatabe (2004) formalises CCA as part of a unified treatment which also deals with coordination of unlike categories.

4 Agreement in Portuguese NPs

Portuguese determiners and adjectives straightforwardly agree in gender and number with the noun they scope over:

(2) a parede colorida
the_{*fsg*} wall_{*fsg*} coloured_{*fsg*}

(3) o teto colorido
the_{*msg*} ceiling_{*msg*} coloured_{*msg*}

(4) a/*as parede colorida/*coloridas
the_{*fsg/*fpl*} wall_{*fsg*} coloured_{*fsg/*fpl*}

(5) o/*os teto colorido/*coloridas
the_{*msg/*mpl*} ceiling_{*msg*} coloured_{*msg/*mpl*}

With coordinate structures the picture is more complicated.

5 Portuguese Coordinate Structures

Inside NP, mixed gender coordinate structures can trigger different agreement patterns on different targets.

- Posthead:
 - Resolution (for NUMBER and GENDER);
 - CCA (for NUMBER and GENDER);
 - Resolution for NUMBER, CCA for GENDER
- Prehead (Determiners and Prenominal Adjectives)
 - ?CCA
 - CCA for GENDER, ?Resolution for NUMBER

Resolution for NUMBER and GENDER

- (6) o teto e a parede coloridos
the_{msg} ceiling_{msg} and the_{fsg} wall_{fsg} coloured_{mpl}
the coloured ceiling and wall
- (7) 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.
- Resolution to masculine is the 'normal' pattern.

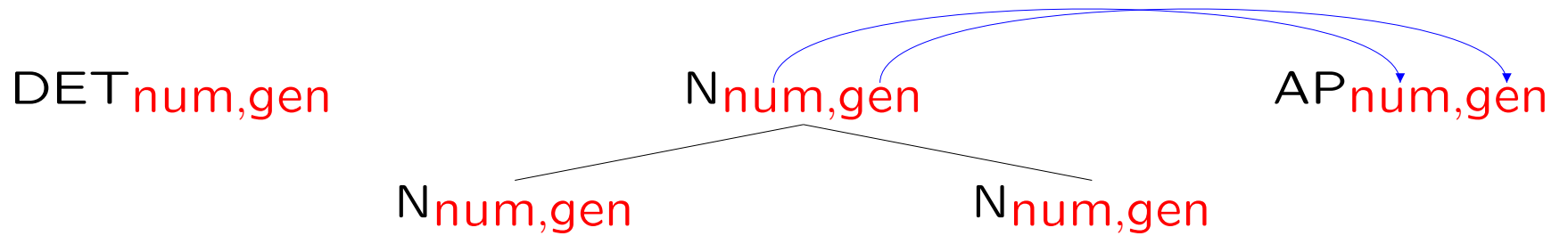


Figure 1: Resolution for Number and Gender

Postnominal CCA

(8) estudos e profissão monástica
studies_{mpl} and profession_{fsg} monastic_{fsg}
monastic studies and profession

(9) no povo e gente hebreia
on the_{msg} people_{msg} and persons_{fsg} hebrew_{fsg}

- CCA is more widespread and common than generally thought.
- There seem to be no cases of ‘furthest’ conjunct agreement.

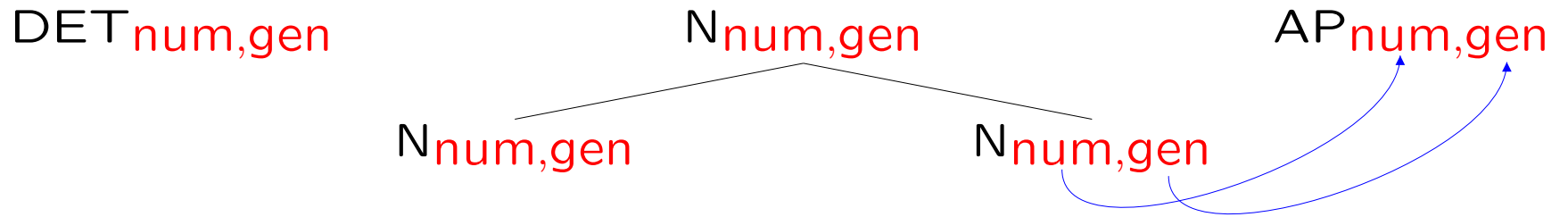


Figure 2: CCA for Number and Gender

Resolution for NUMBER, CCA for GENDER

- (10) 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
- (11) o drama e a loucura vividas
the_{msg} drama_{msg} and the_{fsg} madness_{fsg} lived/felt_{fpl}
the drama and the madness felt
- (12) ...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.

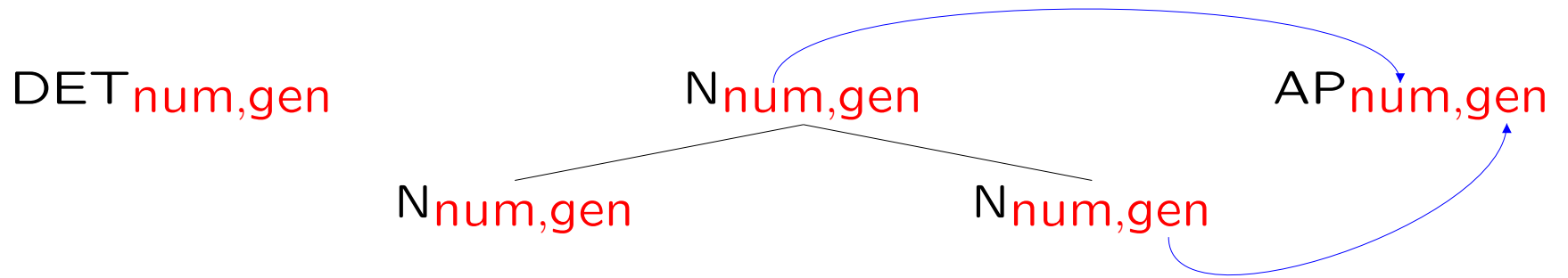


Figure 3: CCA for Gender, Resolved Number

Prenominal CCA

- (13) suas próprias reações ou julgamentos
his_{fpl} own_{fpl} reactions_{fpl} or judgements_{mpl}
his own reactions or judgements
- (14) diversas secções ou subgrupos
diverse_{fpl} sections_{fpl} or subgroups_{mpl}
various sectors or subgroups
- (15) a correcta gestão e preservação
the_{fsg} correct_{fsg} management_{fsg} and conservation_{fsg}
- No resolution for GENDER (i.e. CCA for GENDER).
 - Possibly CCA for NUMBER.

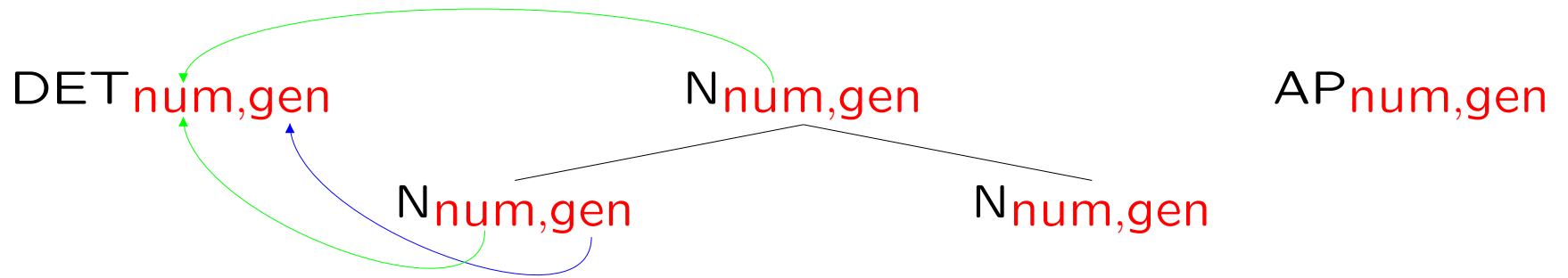


Figure 4: Prenominal CCA for Gender

Notice that different forms of agreement can be triggered on the prenominal and postnominal modifiers at the same time.

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

(16) Esta canção anima os corações e mentes brasileiras.
This song animate the_{mpl} hearts_{mpl} and minds_{fpl} Brazilian_{fpl}

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

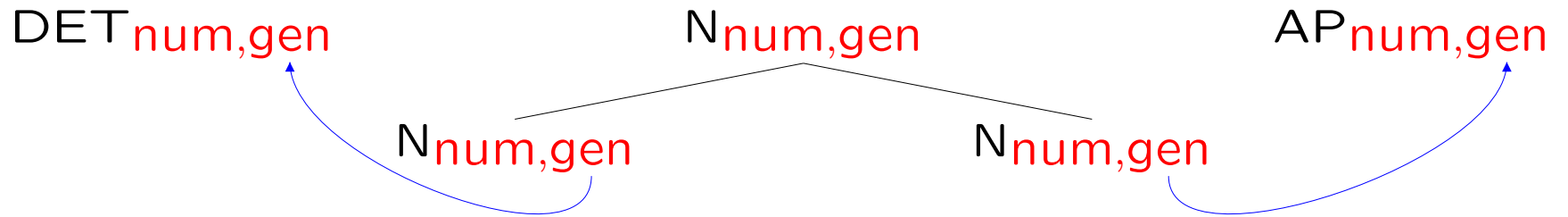
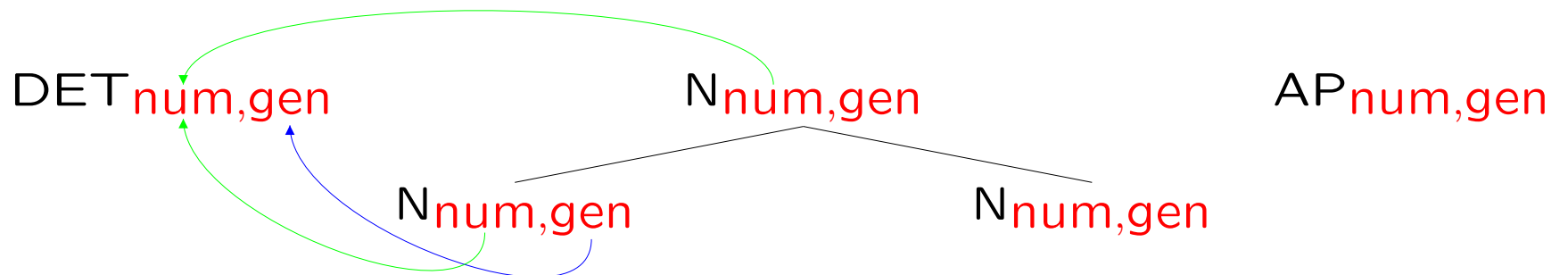


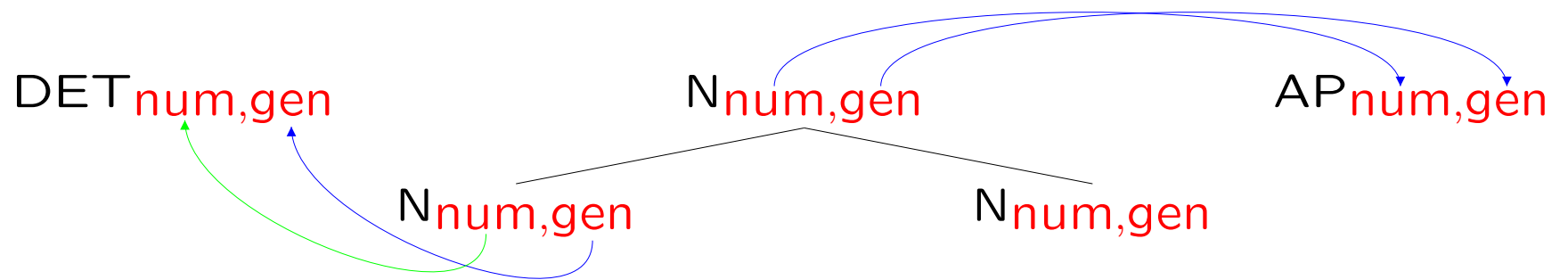
Figure 5: Prenominal *and* Postnominal CCA (for Gender)

Summary

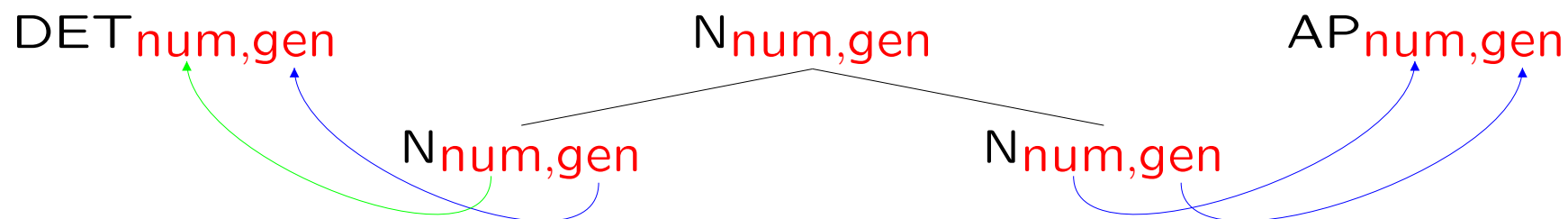
Prehead CCA for Gender (?Number):



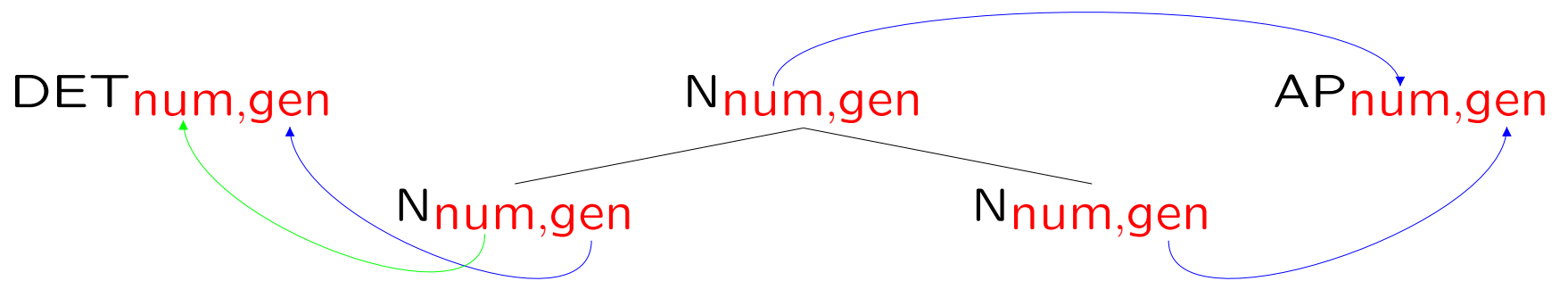
Resolved Number and Gender:



CCA for Number and Gender:



CCA for Gender, Resolved Number:



We have to take into account information about the conjuncts at *both* ends of the coordinate structure.

We cannot manage with just resolved features, or just CC features.

This HPSG analysis below will give access to the agreement features of the first and last conjuncts of coordinate structures.

6 Data from a Corpus Study

A Corpus Study was carried out to estimate the approximate frequency of different agreement strategies in coordinate NPs modified by postnominal adjectives.

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 distinction
 - * 9,915 masculine, and
 - * 9,811 feminine
- Only exact matches.

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

- All cases show number resolution (the postnominal adjective was plural, even when both NPs are singular).
- These are all cases where the adjective scopes over both nouns.

	Frequency	NP1	NP2	Adj	
(a)	0	f	m	f	(resolve to f)
(b)	489	f	m	m	(ncc/resolve to m)
(c)	460	m	f	f	(ncc)
(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	(ncc/resolve to m)
(c)	90	m	f	f	(ncc)
(d)	1737	m	f	m	(resolve to m)
total	1964				

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

- (d)-cases are unambiguously resolution for gender;
- (c)-cases are unambiguously CCA for gender;
- (b)-cases might be CCA (but could also be resolution to masculine);

- Even on the narrowest interpretation, CCA for gender is widespread:

<u>(c)/total</u>			
460/3266	≈	14%	(plural conjuncts)
90/1964	≈	45%	(singular conjuncts)
550/5230	≈	10%	(all cases)

- Notice the ratio of unambiguous cases ($c/(c+d)$):

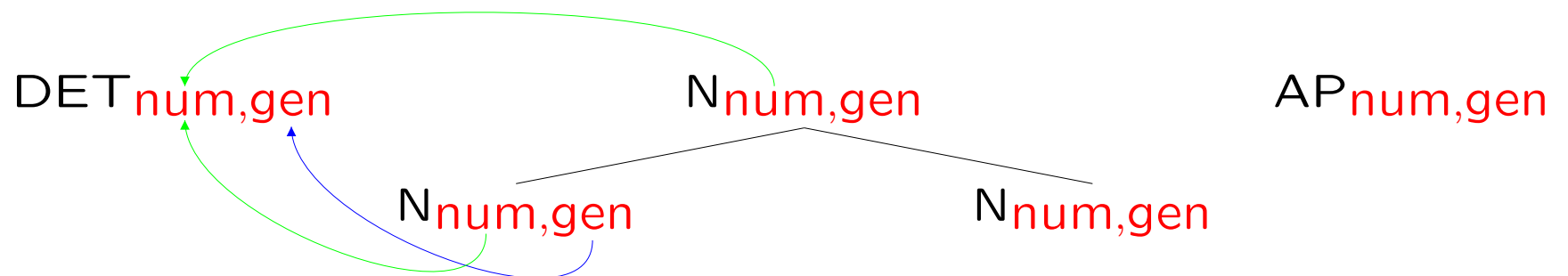
$$\begin{array}{l} \hline (c)/(c+d) \\ \hline 460/(460+2317) = 460/2777 \approx 0.17 \text{ (plural conjuncts)} \\ 90/(90+1737) = 90/1827 \approx 0.05 \text{ (singular conjuncts)} \\ 550/(550+4054) = 550/4604 \approx 0.12 \text{ (all cases)} \\ \hline \end{array}$$

If these data are representative, the odds on speakers using CCA are better than 1 in 10.

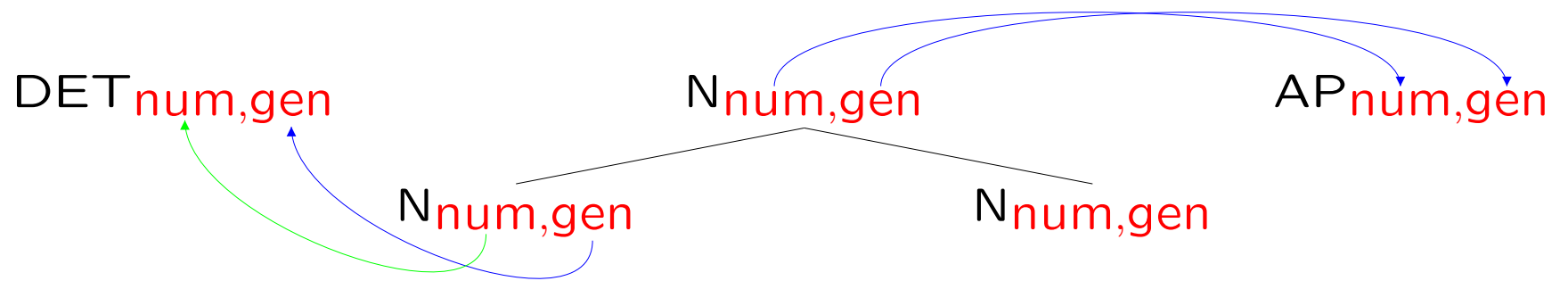
7 HPSG Analysis

Summary

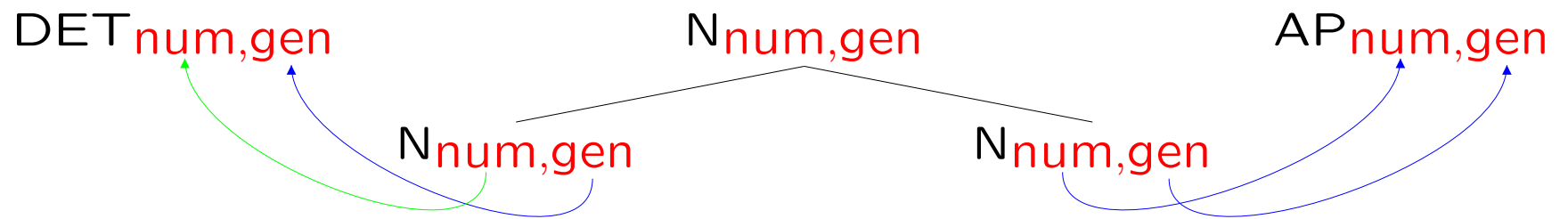
Prehead CCA for Gender (?Number):



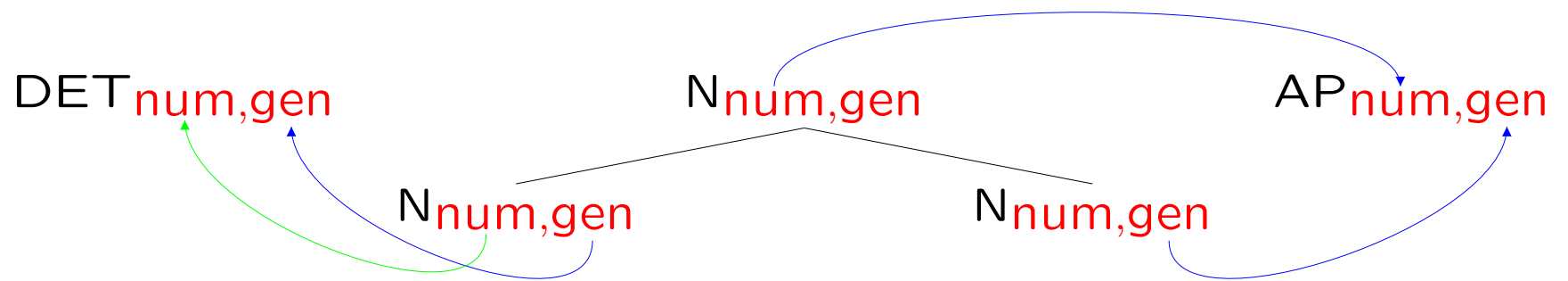
Resolved Number and Gender:



CCA for Number and Gender:



CCA for Gender, Resolved Number:



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:

$$(17) \textit{noun} \wedge \textit{lexical} \rightarrow \left[\begin{array}{l} \text{SS} \mid \text{LOC} \mid \text{CAT} \mid \text{HEAD} \\ \text{LAGR} \\ \text{RAGR} \\ \text{CONCORD} \end{array} \begin{array}{l} \boxed{1} \\ \boxed{1} \\ \boxed{1} \end{array} \right]$$

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.

LAGR, RAGR

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

(18)

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

CONCORD (Resolution)

Different principles operate for

- NUMBER, and
- GENDER

NUMBER

Number resolution is just semantics:

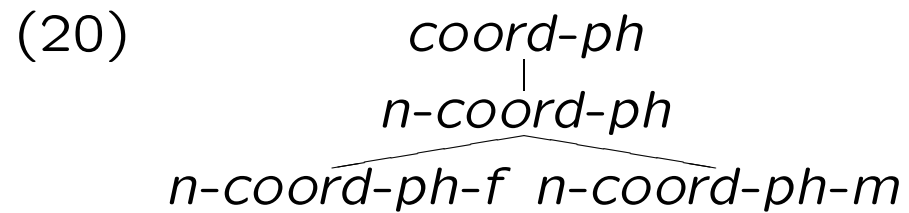
- INDEX | NUM is plural whenever an NP denotes a plurality;
- CONCORD just reflects this

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

$$(19) \quad \text{head} \left[\begin{array}{l} \text{CONTENT | INDEX | NUM} \\ \text{CONCORD | NUM} \end{array} \right]$$

GENDER

Resolution for GENDER is more complex.



n-coord-ph →

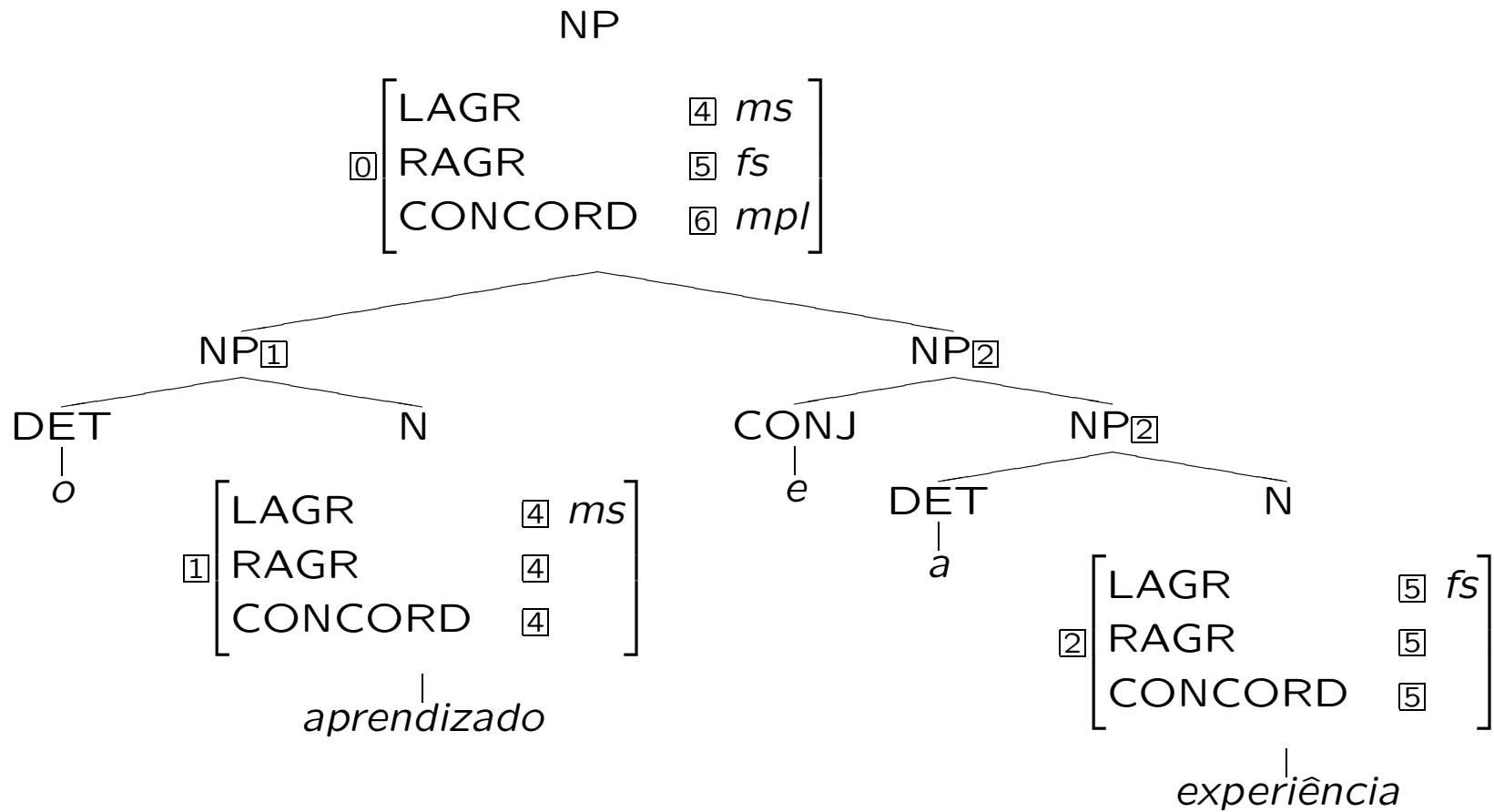
$$\begin{array}{c} ncpf \left[\begin{array}{l} SS | L | CAT | HD | CONCORD | GEND \quad \boxed{3} \\ CONJ-DTRS \quad \left\langle \left[\begin{array}{l} SS | L | HD | CONCORD | GEND \quad \boxed{3} \text{ fem} \end{array} \right]^* \right\rangle \end{array} \right] \\ \vee \\ ncpm \left[\begin{array}{l}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

- 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

(21) ... o aprendizado e a experiência (vivas) ...
 ... the_{msg} learning_{msg} and the_{fsg} experience_{fsg} (lived_{fpl}) ...

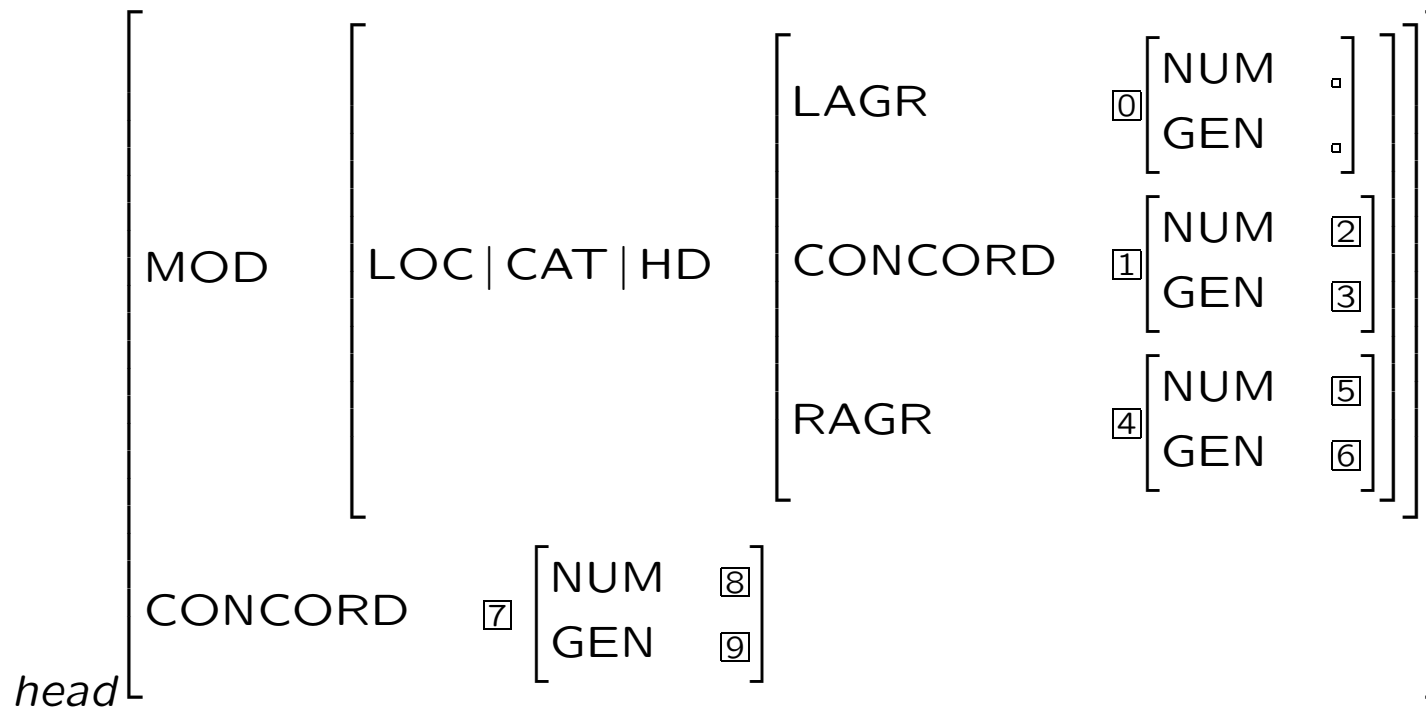
(22)



Agreement inside NP

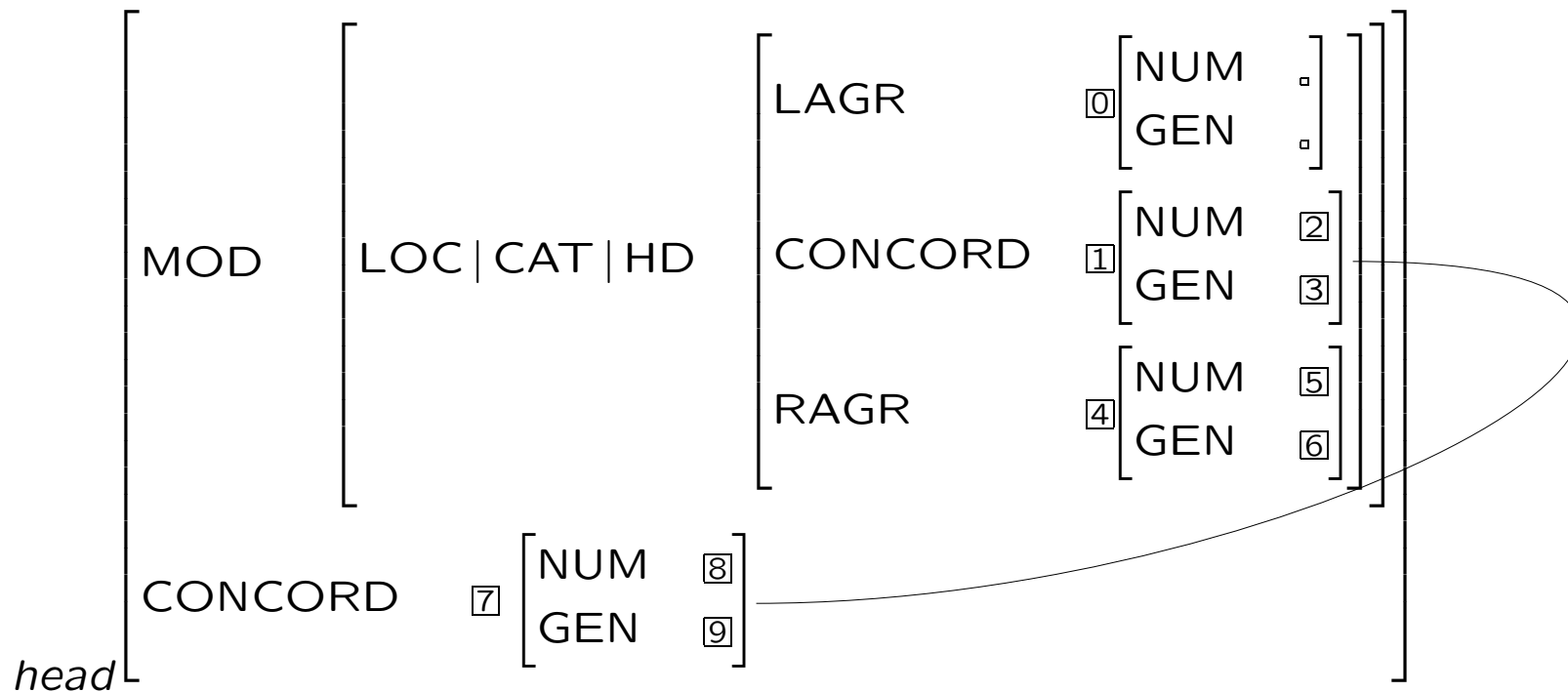
- Posthead:
 - CCA for NUMBER and GENDER; or
 - Resolution for NUMBER and GENDER; or
 - Resolution for NUMBER, CCA for GENDER.
- Prehead (Determiners and Prenominal Adjectives)
 - CCA (at least for GENDER).

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



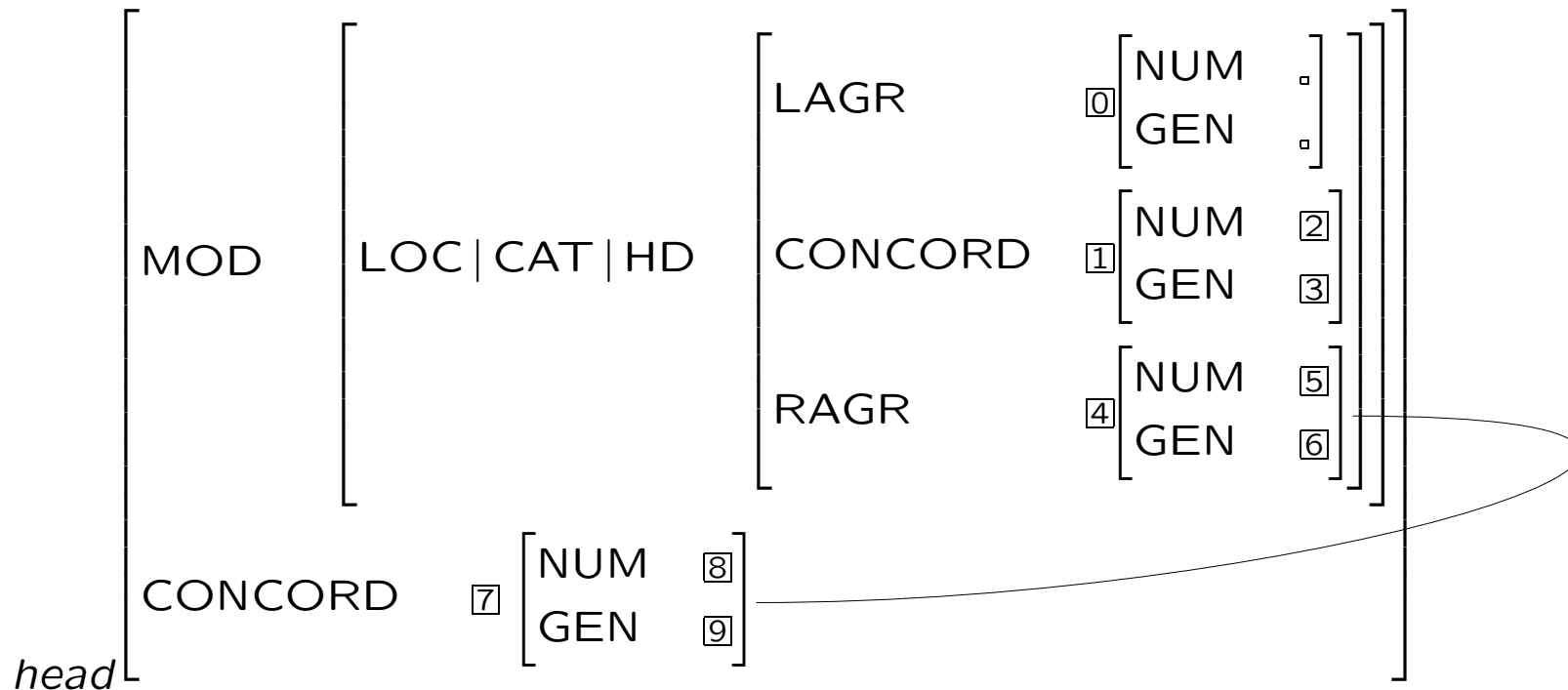
1. $\boxed{7} = \boxed{1}$ (resolution for NUM and GEN)
2. $\boxed{7} = \boxed{4}$ (CCA with last conjunct, NUM and GEN)
3. $\boxed{8} = \boxed{2}$ (resolved number)
4. $\boxed{9} = \boxed{6}$ (CCA with last conjunct for gender)
4. $\boxed{7} = \boxed{0}$ (CCA with first conjunct, GEN (?and NUM))

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



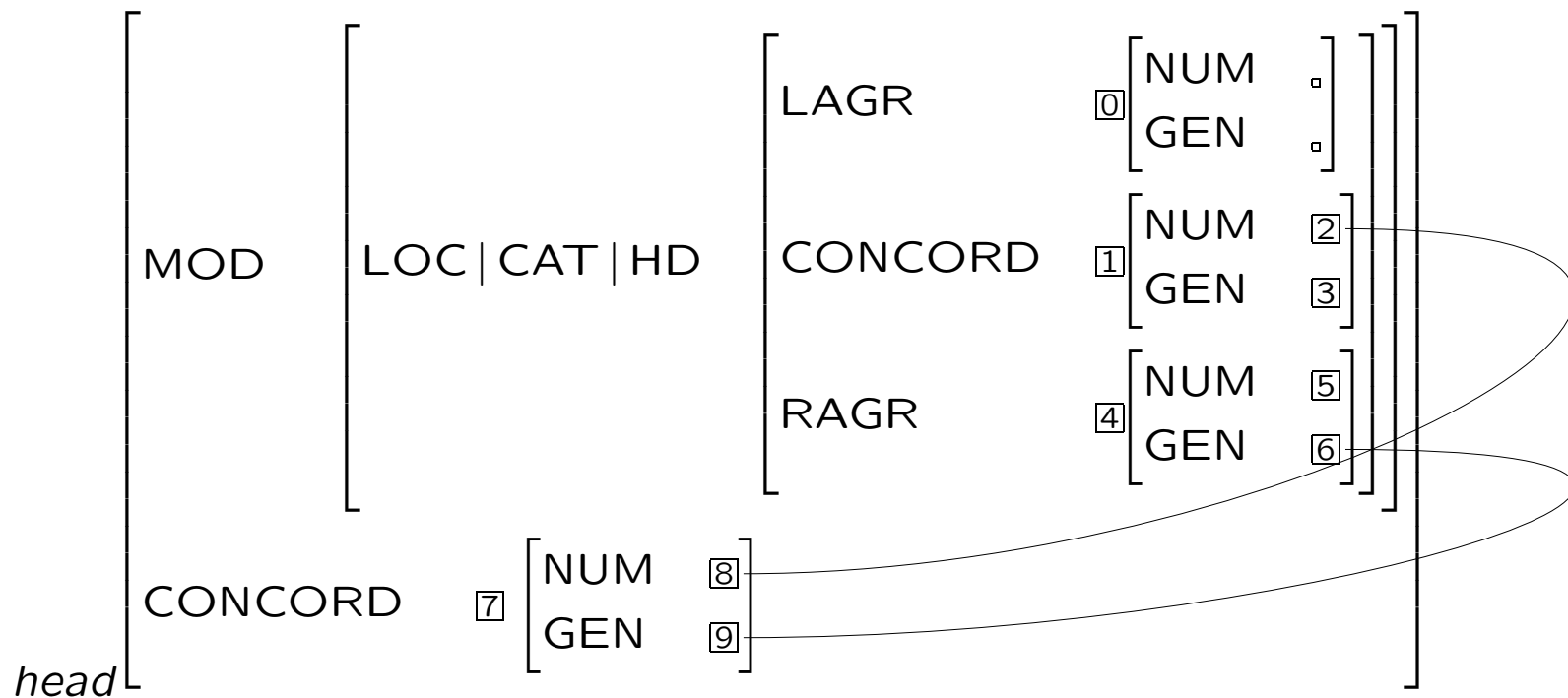
(23) o homem e a mulher **modernos**
 the_{msg} man_{msg} and the_{fsg} woman_{fsg} modern_{mpl}
 the modern man and woman

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



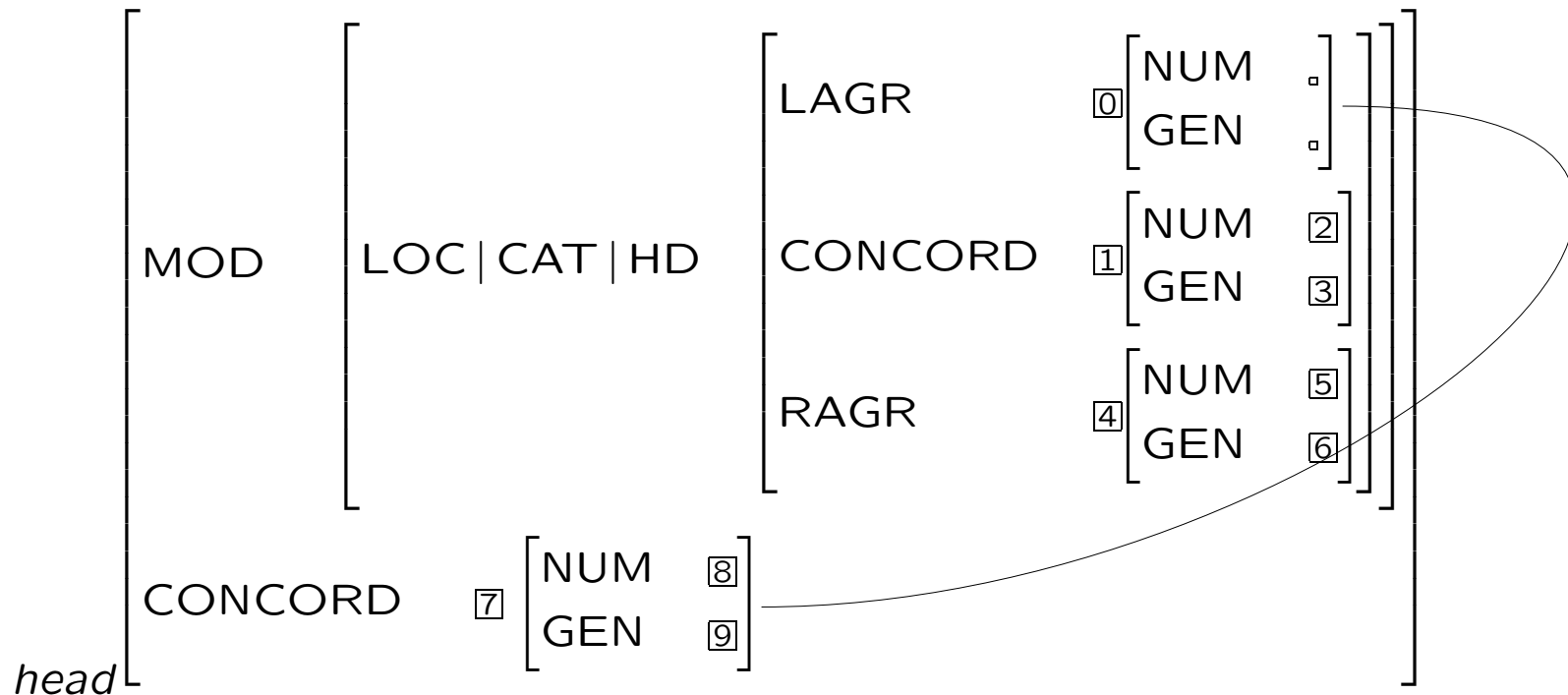
(24) estudos e profissão **monástica**
 studies_{msg} and profession_{fsg} monastic_{fsg}
 monastic studies and profession

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



(25) todo o constrangimento e a dor **sofridas**
 all_{msg} the_{msg} embarrassment_{msg} and the_{fsg} pain_{fsg} suffered_{fpl}

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



(26) suas **próprias** reações ou julgamentos
 his *fpl* own *fpl* reactions *fpl* or judgements *mpl*
 his own reactions or judgements

8 Conclusions/Summary


- On going work;
- Investigation of agreement processes involving NP/noun coordinations in Portuguese;
- Part of a larger project on agreement processes;

- A detailed description of some aspects of the phenomena;
- Novel data: 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.
- HPSG formalization:
 - Two new features, with percolation principles;
 - No defaults (!)

9 Open Questions

- What factors influence choice of CCA?
 - animacy (animates favour resolution)
- Semantic factors
 - esp in relation to NUMBER for DET.
- Other analyses
 - esp. linearity effects
- Syntactic persistence of features outside NP:

(27) nos projectos e realizações que têm vindo a ser concretizadas
in projects_{mpl} and realizations_{fpl} that have come to be made-concrete_{fpl}



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

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