

**Cristina Mota**<sup>1</sup>

**Lucília Chacoto**<sup>2</sup>

**Anabela Barreiro**<sup>1</sup>

# INTEGRATING THE LEXICON-GRAMMAR OF PREDICATE NOUNS WITH SUPPORT VERB *FAZER* INTO **PORT4NOOJ**



<sup>1</sup> INESC-ID, Lisbon

<sup>2</sup> Universidade do Algarve & CLUL

# Port4NooJ 3.0

eSPERTo

Smart Paraphrasing System



Port4NooJ



Bilingual  
PT-EN

+

Morpho-  
syntactic  
Relations

Derivational  
Relations

Semantic  
Relations

LG of  
Human  
Intransitive  
Adjectives

SentiLex

Semantico-Syntactic  
(SAL) properties

Support Verb  
Constructions

Stencil  
NER

LG of  
Support  
Verb *Fazer*

# eSPERTo Interface

## eSPERTo - System for Paraphrasing in Editing and Revision of Text

<https://esperto.l2f.inesc-id.pt/esperto/esperto/demo.pl>

### Parameters

Demo mode

Interface idiom English

Resources idiom Portuguese

Dictionary PT-dict

Sample text SAN

Paraphrasing

- Active > Passive
- Passive > Active
- Simple adverb > Compound
- Compound adverb > Simple
- Nominal/adjectival predicate > Verb
- Nominal/adjectival predicate >
- Nominal/adjectival predicate
- Verb > Nominal/adjectival predicate
- Relative construct > Adjective
- Possessive > Relative construct
- Synonyms
- Human intransitive adjective
- Predicate nouns with Vsup fazer

Debug

### Input file or text (click to show/hide)

Choose file:

Insert text in the text box

Ontem, o escritor de Marrocos fez uma aposta com a mulher portuguesa.

### Results (click to show/hide)

Ontem, [ o escritor [ de Marrocos ] [ fez uma aposta ] com a mulher [ portuguesa ] ].

a mulher portuguesa fez uma aposta com o escritor de Marrocos
o escritor de Marrocos e a mulher portuguesa fizeram uma aposta
apostou
Suggest your own paraphrase >

[the writer [of Morocco] [made a bet] with the [Portuguese] woman]

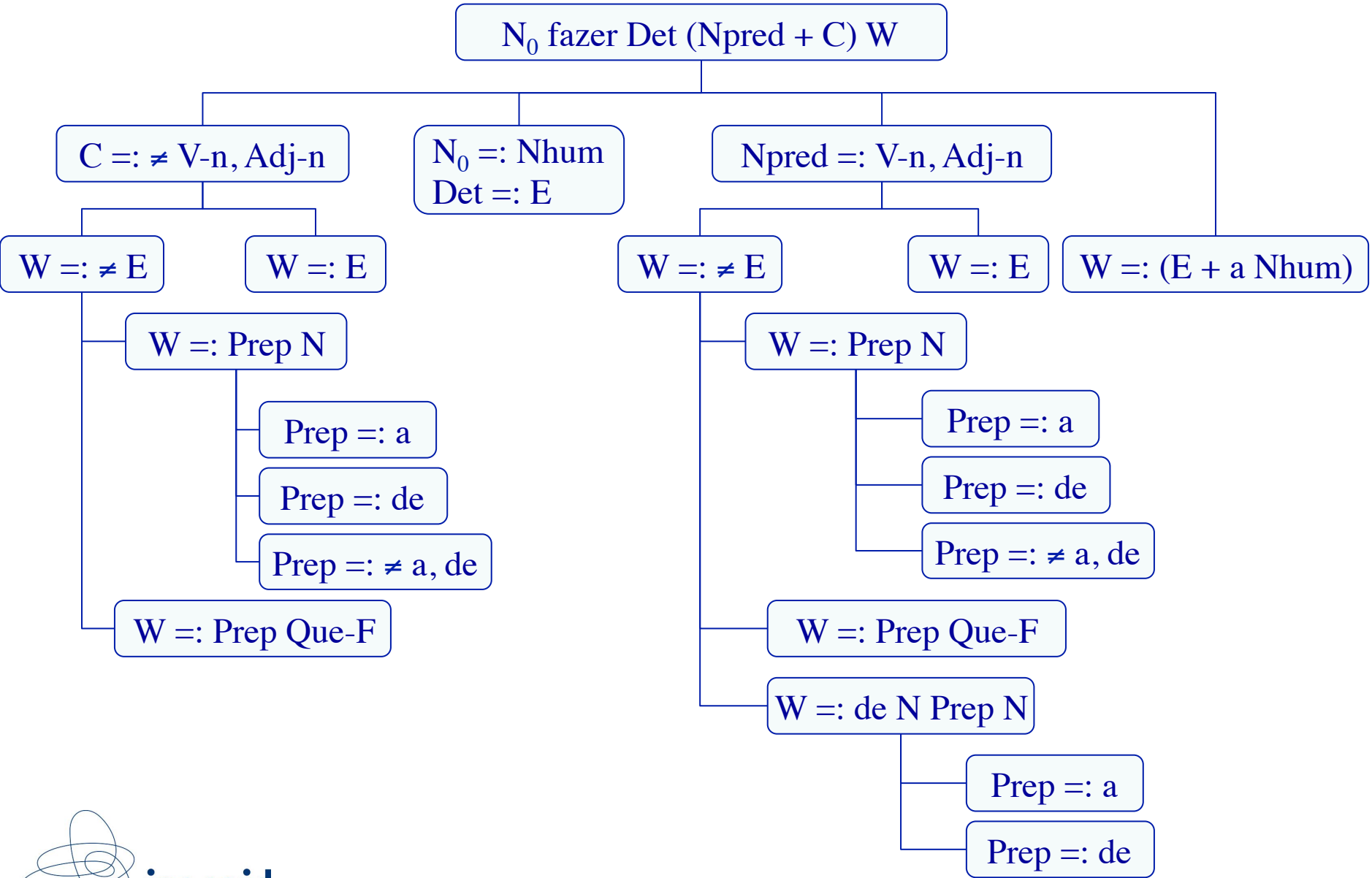
the Portuguese woman made a bet with the writer of Morocco  
**the writer of Morocco and the Portuguese woman made a bet betted**

interactive application - helps Portuguese language learners in producing and revising their texts

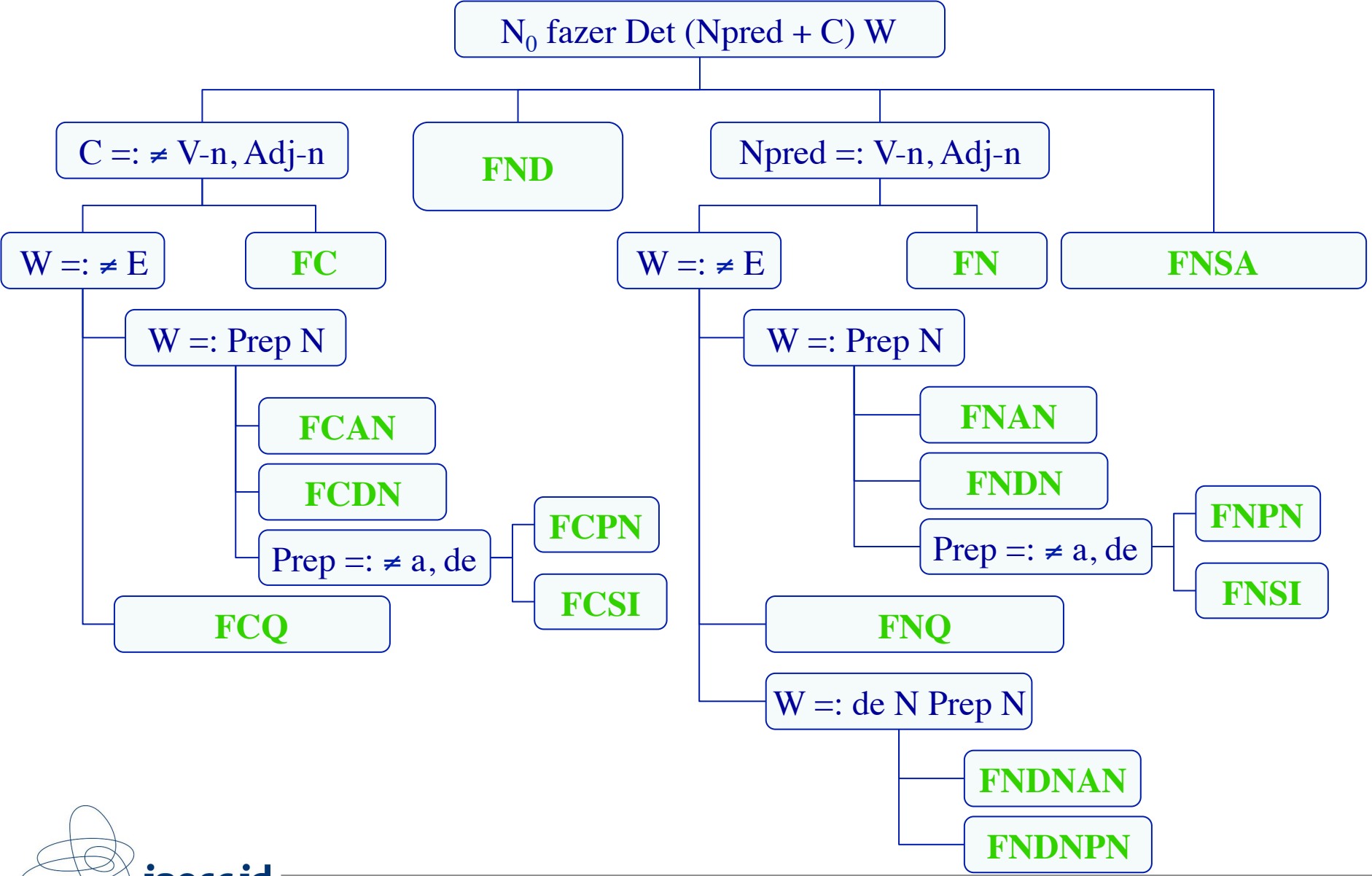
# Predicate nouns with *Vsup fazer* (Chacoto, 2005)

- eSPERTo was enhanced with new paraphrases, derived from **17 Lexicon-Grammar (LG) tables** describing the distributional properties of **3,000 predicate nouns with support verb *fazer***, of which **800 are compound words**:
  - The systematic survey was carried out by assessing seven dictionaries and a corpus of five million words (Part 20 of the online 180 million word corpus CETEMPúblico.11)
  - The predicate nouns in the lexicon are, in general, everyday vocabulary (simple and compound nouns), with the exception of a group of predicate nouns of the sports and medical domains described.
- For now, the 3 tables that formalize medical technical terms were left out for a second stage.

# Predicate nouns with *Vsup fazer* Classification Criteria



# Predicate nouns with *Vsup fazer* Classification Criteria



# Transformations based on noun predicates with *Vsup fazer*

## Symmetrical Nouns

Symmetry	Paraphrases
[N0 V N1]	<i>O Pedro aliou-se (com + a) a Maria</i> <i>Pedro joined / allied himself with / to Maria</i>
[N1 V N0]	<i>A Maria aliou-se (com + a) o Pedro</i> <i>Maria joined / allied himself with / to Pedro</i>
[N0 Vcop Adj N1]	<i>O João é casado com a Ana</i> <i>João is married to Maria</i>
[N1 Vcop Adj N0]	<i>A Ana é casada com o Pedro</i> <i>Ana is married to João</i>
[N0 Vsup N N1]	<i>O Pedro fez um acordo com a Maria</i> <i>Pedro made an agreement with Maria</i>
[N1 Vsup N N0]	<i>A Maria fez um acordo com o Pedro</i> <i>Maria made an agreement with Pedro</i>

## Conversion

Conversion	Paraphrases
[N0 fazer N Prep N1]	<i>O Paulo fez um telefonema à Maria</i> <i>Paulo made a phone call to Maria</i>
[N1 Vsup-conv Prep-conv N0]	<i>A Maria recebeu um telefonema do Paulo</i> <i>Maria received a phone call from Paulo</i>

# Integration of LG of PT Vsup *fazer*

- Major challenges
  - ✧ 63% of the predicate nouns already existed in Port4NooJ
    - ➔ Consolidation of linguistic information in dictionaries and grammars
  - ✧ Verbal and adjectival derivation is only marked with + in the table
    - ➔ Cross information with an additional file
  - ✧ Includes 800 multiword predicate nouns
    - ➔ transcrição integral ⇔ transcrever integralmente
  - ✧ LG complies with Portuguese Ortographic Agreement, but not Port4NooJ
    - ➔ For now, create a non-compliant dictionary of predicate nouns



# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Mostly done automatically with various scripts  
≠ scripts used to integrate the tables of Human Intransitive Adjectives

### Port4NooJ

- Current version (CV)
- Version before removing Npred that derive from verbs (OV)

LG tables

- ✓ Check if noun or AO compliant noun exists in Port4NooJ:
  - ✓ If noun not in CV nor in OV
    - Create new entry
  - ✓ If noun in CV and (not in OV or CV=OV entry)
    - Merge<sub>1</sub> the LG properties with current entry
  - ✓ If noun in OV only or (CV≠OV entry) then
    - Merge<sub>2</sub> the LG properties with old entry
    - Remove nominalization from CV
- ✓ Create FLX and DRV codes and corresponding rules as needed
- ✓ Check for missing FLX and DRV codes

npred\_vsupfazer

# Integration of LG of PT *Vsup fazer*

- From LG tables to NooJ dictionaries
  - Representation of LG table properties

FN																	Exemplo			
N0 =: Nhum	N0 =: N-hum	N0 =: Nplural obrig	Vsup	Det =: E	Det =: UM - Modif	Det =: O	Npred	Vsup =: estar Prep	Vsup =: ter	Vsup =: dar	Vasp =: iniciar	Vasp =: prosseguir	Vasp =: concluir	Vestil =: cometer	V	Adj		de VO inf W	de que F	GN =: N de N0
+	-	-	fazer	+	-	-	espionagem	-	-	-	+	+	+	-	+	+	-	-	-	O homem fazia espionagem
-	+	-	fazer	+	+	+	espuma	-	+	-	-	-	-	-	+	+	-	-	+	O detergente fazia espuma
+	-	-	fazer	+	+	+	estardalhaço	-	-	+	-	-	-	-	+	-	-	-	+	O Zé fez um estardalhaço enorme

+NONnhum

+Vsupter

+Npred  
+Vsup=fazer  
+Table=FN

# Integration of LG of PT Vsup *fazer*

- From LG tables to NooJ dictionaries
  - Representation of LG table properties

FN																	Exemplo			
N0 =: Nhum	N0 =: N-hum	N0 =: Nplural obrig	Vsup	Det =: E	Det =: UM - Modif	Det =: O	Npred	Vsup =: estar Prep	Vsup =: ter	Vsup =: dar	Vasp =: iniciar	Vasp =: prosseguir	Vasp =: concluir	Vestil =: cometer	V	Adj		de VO inf W	de que F	GN =: N de N0
+	-	-	fazer	+	-	-	espionagem	-	-	-	+	+	+	-	+	+	-	-	-	O homem fazia espionagem
-	+	-	fazer	+	+	+	espuma	-	+	-	-	-	-	-	+	+	-	-	+	O detergente fazia espuma
+	-	-	fazer	+	+	+	estardalhaço	-	-	+	-	-	-	-	+	-	-	-	+	O Zé fez um estardalhaço enorme

Espionagem – espiar; ser espião, ia – FN  
 Espuma – espumar; ser espumoso, a – FN  
 Estardalhaço – estardalhaçar – FN

+DRV=N2V2 : FALAR

+DRV=N2A14 : ALTO

- DRV code is determined and formalized automatically by finding the radical between the noun and the verb or adjective that are listed in a separate file  
 espuma(r) => N2V2 = r/V  
 espum(oso) => A2V614= <B1>oso/A
  - FLX code of derived word is determined by consulting Port4NooJ  
 espumar,V+FLX=FALAR+Aux=1+INOPfor46+Subset=363+EN=foam...  
 espumoso,A+FLX=ALTO+NAV+Apred+EN=bubbly...
- If the derived form does not exist, then its code is assigned automatically

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ① Noun not in Port4NooJ (old or current):

- ✓ Create new entry:
  - ✓ FLX code is assigned automatically given the ending of the word
  - ✓ Entries are checked for missing FLX codes and reviewed by a linguist
  - ✓ All other properties come from LG table
- ✓ Add entry to new standalone dictionary `npred_vsupfazer.dic`

`batota, N+FLX=CASA+Npred+Vsup=fazer+Table=FN+N0Nhum+DetE+DetUMModif  
+Npred+DRV=N2V2:FALAR+DRV=N2A6:ALTO+GNNdeN0`

`rodagem, N+FLX=ANO+Npred+Vsup=fazer+Table=FNAN+N0Nhum+DetO+Npred+Prepa  
+N1Nnhum+VsuestarPrep+Vaspiniciar+Vaspprosseguir+Vasppreparar  
+DRV=N2V27:FALAR`

`rodagem, N+FLX=ANO+Npred+Vsup=fazer+Table=FNDNh1+N0Nhum+DetUMModif  
+DetO+Npred+Prepde+ConstConversa+DRV=N2V27:FALAR`

`rodagem, N+FLX=ANO+Npred+Vsup=fazer+Table=FNDNh1+N0Nhum+DetUMModif  
+DetO+Npred+Prepde+Vaspiniciar+Vaspprosseguir+Vasppreparar  
+DRV=N2V27:FALAR`

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ② Noun exists both in current and old Port4NooJ

#### A. If entries are the same do Merge 1:

- ✓ Blindly add additional properties as specified by the LG tables to current entries
- ✓ Add merged entries to `npred_vsupfazer.dic`

`curva,N+FLX=ANO+Set=56+Subset=280+EN=curve` ←

`curva,N+FLX=ANO+Set=56+Subset=280+EN=curving` ←

`curva,N+FLX=ANO+PresPart+Set=56+Subset=280+EN=curving` ←

```
+Npred
+Vsup=fazer
+Table=FN
+N0Nhum
+N0Nnhum
+DetUMModif
+DetO
+Npred
+Vsupdar
+Vaspiniciar
+Vaspprosequir
+Vasconcluir
+DRV=N2V2:FALAR
```

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ② Noun exists both in current and old Port4NooJ

#### A. If entries are the same do Merge 1:

- ✓ Blindly add additional properties as specified by the LG tables to current entries
- ✓ Add merged entries to `npred_vsupfazer.dic`

```
curva,N+FLX=ANO+Set=56+Subset=280+EN=curve+Npred+Vsup=fazer+Table=FN  
+NONhum+NONnhum+DetUMModif+DetO+Npred+Vsupdar+Vaspiniciar  
+Vaspprossequir+Vaspcocluir+DRV=N2V2:FALAR
```

```
curva,N+FLX=ANO+Set=56+Subset=280+EN=curving+Npred+Vsup=fazer+Table=FN  
+NONhum+NONnhum+DetUMModif+DetO+Npred+Vsupdar+Vaspiniciar  
+Vaspprossequir+Vaspcocluir+DRV=N2V2:FALAR
```

```
curva,N+FLX=ANO+PresPart+Set=56+Subset=280+EN=curving+Npred+Vsup=fazer  
+Table=FN+NONhum+NONnhum+DetUMModif+DetO+Npred+Vsupdar+Vaspiniciar  
+Vaspprossequir+Vaspcocluir+DRV=N2V2:FALAR
```

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ② Noun exists both in current and old Port4NooJ

B. If entries are not the same do **Merge 2** with old entries as shown in case 3:

- ✓ Remove previous Npred related properties
- ✓ Blindly add additional properties as specified by the LG tables to current entries
- ✓ Add merged entries to **npred\_vsupfazer.dic**
- ✓ Remove nominalization from CV

Entries in CV:

`cruzamento, N+FLX=ANO+AB+mot+EN=crossover`

`cruzamento, N+FLX=ANO+CO+recp+EN=frog`

`cruzamento, N+FLX=ANO+PL+nagcom+EN=crossings`

Entries in OV:

`cruzamento, N+FLX=ANO+PresPart+Npred+Nom+Set=68+Subset=551+EN=intersecting  
+VRB=cruzar`

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ② Noun exists both in current and old Port4NooJ

B. If entries are not the same do **Merge 2** with old entries as shown in case 3:

- ✓ Remove previous Npred related properties
- ✓ Blindly add additional properties as specified by the LG tables to current entries
- ✓ Add merged entries to **npred\_vsupfazer.dic**
- ✓ Remove nominalization from CV

Entries in CV:

`cruzamento, N+FLX=ANO+AB+mot+EN=crossover`

`cruzamento, N+FLX=ANO+CO+recp+EN=frog`

`cruzamento, N+FLX=ANO+PL+nagcom+EN=crossings`

Entries in OV:

`cruzamento, N+FLX=ANO+PresPart+Npred+Nom+Set=68+Subset=551+EN=intersecting  
+VRB=cruzar`



# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ② Noun exists both in current and old Port4NooJ

B. If entries are not the same do **Merge 2** with old entries as shown in case 3:

- ✓ Remove previous Npred related properties
- ✓ Blindly add additional properties as specified by the LG tables to current entries
- ✓ Add merged entries to **npred\_vsupfazer.dic**
- ✓ Remove nominalization from CV

```
+Npred  
+Vsup=fazer  
+Table=FNPN  
+NONhum  
+DetE  
+DetUMModif  
+DetO  
+Prepara  
+N1Nhum  
+DRV=N2V16:FALAR  
+GNNdeN0PrepN1
```

#### Entries in CV:

cruzamento, N+FLX=ANO+AB+mot+EN=crossover

cruzamento, N+FLX=ANO+CO+recp+EN=frog

cruzamento, N+FLX=ANO+PL+nagcom+EN=crossings

#### Entries in OV:

cruzamento, N+FLX=ANO+PresPart+Set=68+Subset=551+EN=intersecting



# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ② Noun exists both in current and old Port4NooJ

B. If entries are not the same do **Merge 2** with old entries as shown in case 3:

- ✓ Remove previous Npred related properties
- ✓ Blindly add additional properties as specified by the LG tables to current entries
- ✓ Add merged entries to **npred\_vsupfazer.dic**
- ✓ Remove nominalization from CV

Entries in CV:

`cruzamento, N+FLX=ANO+AB+mot+EN=crossover`

`cruzamento, N+FLX=ANO+CO+recp+EN=frog`

`cruzamento, N+FLX=ANO+PL+nagcom+EN=crossings`

Entries in OV:

`cruzamento, N+FLX=ANO+PresPart+Set=68+Subset=551+EN=intersecting+Npred  
+Vsup=fazer+Table=FNPN+N0Nhum+DetE+DetUMModif+DetO+Npred+Prepara  
+N1Nhum+DRV=N2V16:FALAR+GNNdeN0PrepN1`

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ③ Noun exists only in old Port4NooJ

- ✓ Do Merge 2 with old entries as shown in Case 2-B:
  - ✓ Remove previous Npred related properties
  - ✓ Blindly add additional properties as specified by the LG tables to current entries
  - ✓ Add merged entries to [npred\\_vsupfazer.dic](#)
  - ✓ Remove nominalization from CV

protesto, N+FLX=ANO+AB+strvb+Npred+Nom+EN=outcry+VRB=protestar

protesto, N+FLX=ANO+PNT+Npred+Nom+Set=32+Subset=248+EN=protest  
+VRB=protestar

# Integration of LG of PT Vsup<sub>fazer</sub>

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ③ Noun exists only in old Port4NooJ

- ✓ Do Merge 2 with old entries as shown in Case 2-B:
  - ✓ Remove previous Npred related properties
  - ✓ Blindly add additional properties as specified by the LG tables to current entries
  - ✓ Add merged entries to `npred_vsupfazer.dic`
  - ✓ Remove nominalization from CV

`protesto, N+FLX=ANO+AB+strvb+Npred+Nom+EN=outcry+VRB=protestar`

`protesto, N+FLX=ANO+PNT+Npred+Nom+Set=32+Subset=248+EN=protest  
+VRB=protestar`

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ③ Noun exists only in old Port4NooJ

- ✓ Do Merge 2 with old entries as shown in Case 2-B:
  - ✓ Remove previous Npred related properties
  - ✓ Blindly add additional properties as specified by the LG tables to current entries
  - ✓ Add merged entries to `npred_vsupfazer.dic`
  - ✓ Remove nominalization from CV

protesto, N+FLX=ANO+AB+strvb+EN=outcry

protesto, N+FLX=ANO+PNT+Set=32+Subset=248+EN=protest

```
+Npred
+Vsup=fazer
+Table=FNPN
+N0Nhum
+DetUMModif
+DetO
+Npred
+Prepcontra
+N1Nhum
+N1abstr
+Vaspiniciar
+Vaspprosseguir
+Vasppconcluir
+DRV=N2V8:FALAR
+GNNdeN0PrepN1
```

# Integration of LG of PT Vsup *fazer*

## – From LG tables to NooJ dictionaries

- Integration with eSPERTo dictionary entries

### ③ Noun exists only in old Port4NooJ

- ✓ Do Merge 2 with old entries as shown in Case 2-B:
  - ✓ Remove previous Npred related properties
  - ✓ Blindly add additional properties as specified by the LG tables to current entries
  - ✓ Add merged entries to `npred_vsupfazer.dic`
  - ✓ Remove nominalization from CV

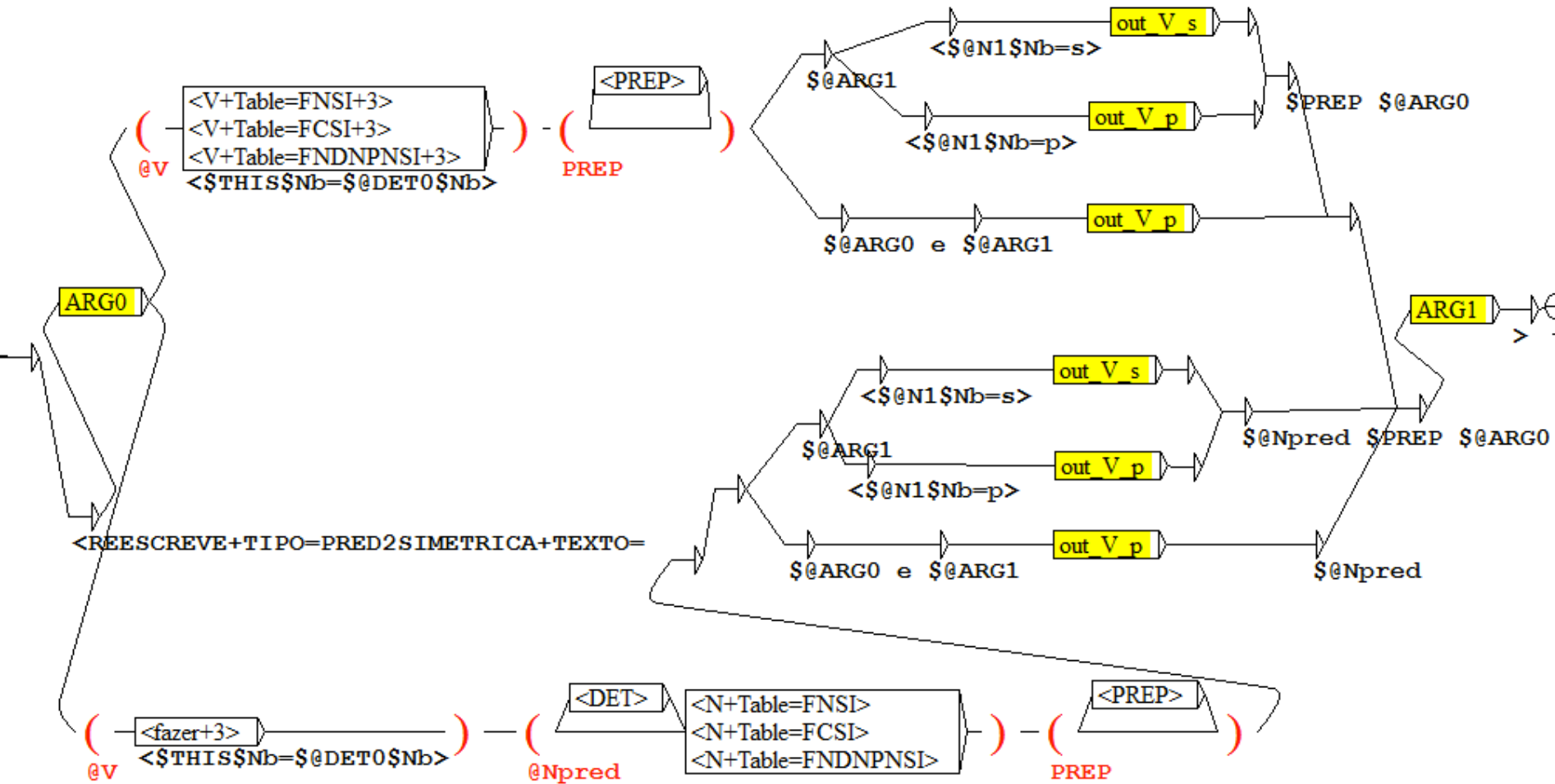
```
protesto, N+FLX=ANO+AB+strvb+EN=outcry+Npred+Vsup=fazer+Table=FNPN  
+N0Nhum+DetUMModif+DetO+Npred+Prepcontra+N1Nhum+N1abstr  
+Vaspiniciar+Vaspprosseguir+Vaspconcluir+DRV=N2V8:FALAR  
+GNNdeN0PrepN1
```

```
protesto, N+FLX=ANO+PNT+Set=32+Subset=248+EN=protest+Npred  
+Vsup=fazer+Table=FNPN+N0Nhum+DetUMModif+DetO+Npred+Prepcontra  
+N1Nhum+N1abstr+Vaspiniciar+Vaspprosseguir+Vaspconcluir  
+DRV=N2V8:FALAR+GNNdeN0PrepN1
```

# Integration of LG of PT Vsup *fazer*

## - From LG to NooJ grammars

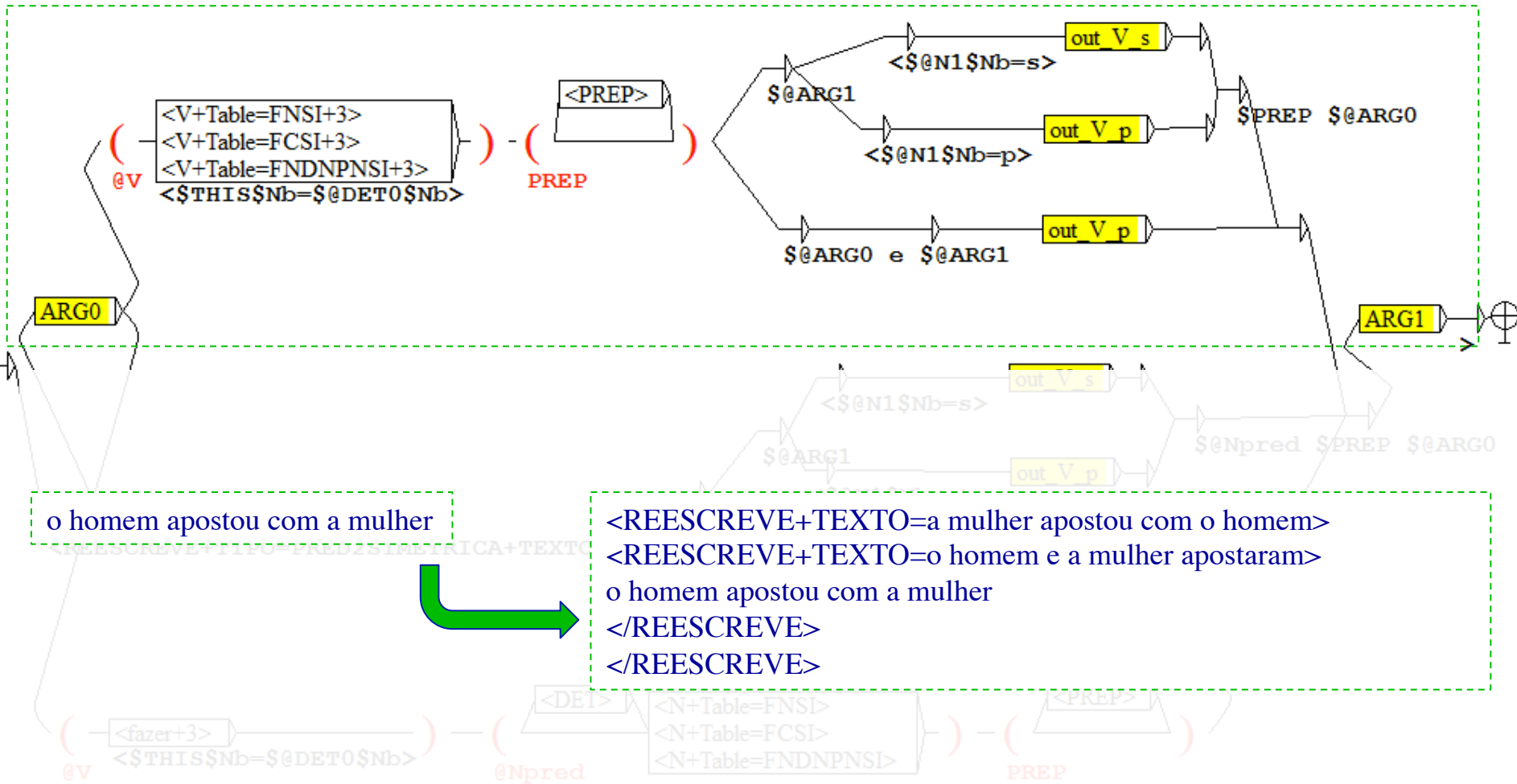
- New grammars to paraphrase constructions based on specific LG properties, such as paraphrase of symmetric predicates



# Integration of LG of PT Vsup *fazer*

## - From LG to NooJ grammars

- New grammars to paraphrase constructions based on specific LG properties, such as paraphrase of symmetric predicates





# Integration of LG of PT Vsup *fazer*

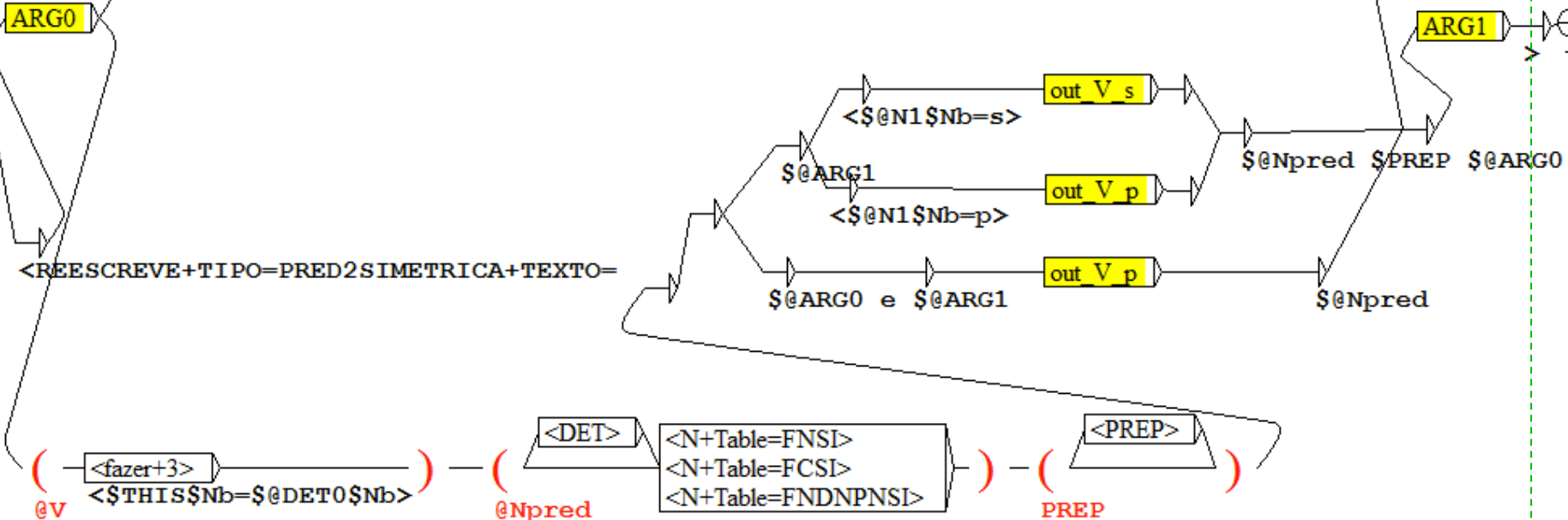
## - From LG to NooJ grammars

- New grammars to paraphrase constructions based on specific LG properties, such as paraphrase of symmetric predicates

o homem fez uma aposta com a mulher

<REESCREVE+TEXTO=a mulher fez uma aposta com o homem>  
 <REESCREVE+TEXTO=o homem e a mulher fizeram uma aposta>  
 o homem apostou com a mulher  
 </REESCREVE>  
 </REESCREVE>

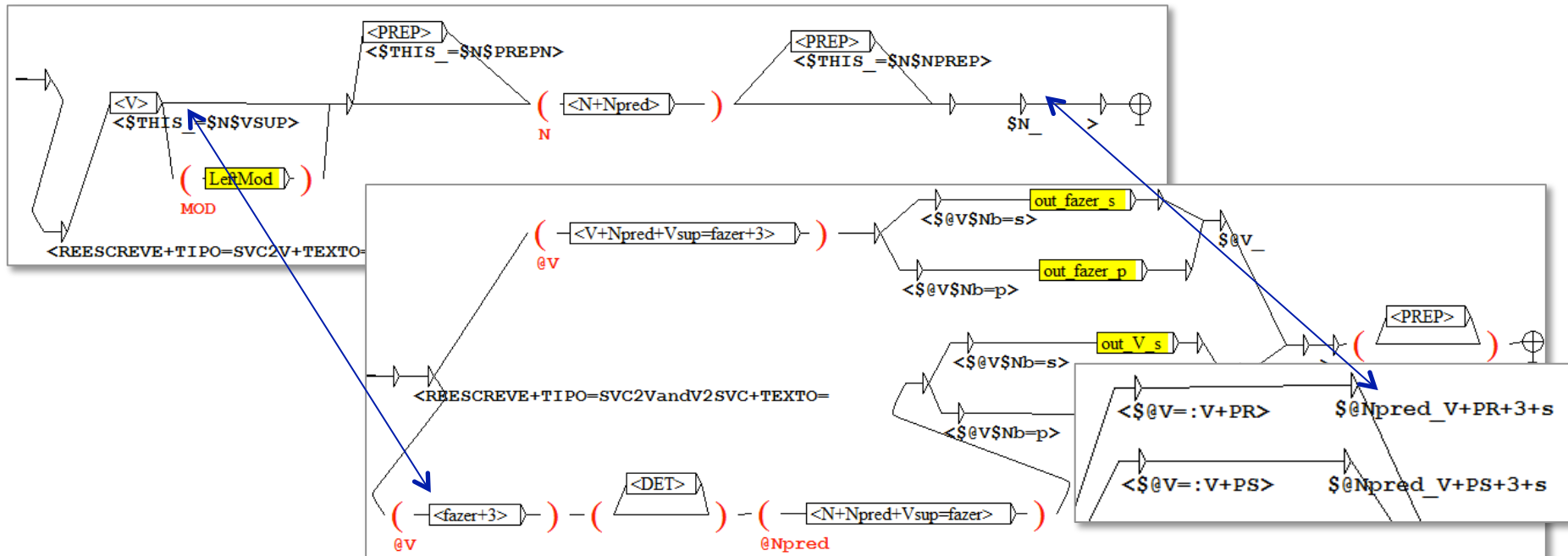
\$@ARG0 e @\$ARG1



# Integration of LG of PT Vsup *fazer*

## - From LG to NooJ grammars

- Updating the Port4NooJ grammars that already describe constructions involving support verbs with properties from these tables, and extending those grammars to take into account greater paraphrastic knowledge



- Extending grammars that paraphrase active and passive sentences so that they will also paraphrase equivalent active and passive noun groups

# Preliminary Results

- 5,236 predicate nouns with *Vsup fazer* (1,610 different noun lemmas)
- 332 new derivational paradigms
- Example grammars for the syntactic parser
- Most nouns already existed in Port4NooJ (63%)

→ 5% increase in Port4NooJ nominal entries and 17.5% increase in predicate nouns

Table	Example	In Port4NooJ	New	% In
FNDN-hl	O barco fez a <b>abordagem</b> do cais	367	169	68%
FNDNa	O advogado fez a <b>alegação</b> de insanidade mental	208	95	69%
FN	A instituição fez uma <b>angariação de fundos</b>	112	127	47%
FNAN	O Zé fez um <b>aceno</b> ao Tó	121	82	60%
FNP	O Zé fez um <b>acrescento</b> na prova tipográfica	127	46	73%
FNDNh	A Ana faz o <b>acolhimento</b> dos convidados	98	42	70%
FNDNAN	O Zé fez uma <b>adaptação</b> do romance ao cinema	60	16	79%
FND	O Zé faz <b>ciclismo</b>	17	56	23%
FNDNPN	O Tó fez o <b>câmbio</b> das pesetas em liras	52	13	80%
FCSI	O Zé fez um <b>acordo</b> com a Ana	19	40	32%
FNSI	O Zé fez uma <b>aliança</b> com o Tó	28	14	67%
FNDNPNSI	O padre fez o <b>casamento</b> da Ana com o Zé	21	4	84%
FCQ	O Tó fez a <b>fineza</b> de convidar a Ana	5	5	50%
FNQ	A Maria faz <b>tenção</b> de ter filhos	4	1	80%
Total		1239	710	63%

# Next Steps

- **Complete** the integration of the LG of predicate nouns with *Vsup fazer*
  - By creating all grammars to process the constructions formalized in LG
- Revise and evaluate the new resources
- Integrate and adapt additional LG grammars:
  - Constructions with *Vsup ser de* (Baptista, 2005)

**Thank you!**

شكرا!



**TECH  
NOLOGY  
FROM  
SEED**

[cmota@ist.utl.pt](mailto:cmota@ist.utl.pt) | [lchacoto@ualg.pt](mailto:lchacoto@ualg.pt) | [anabela.barreiro@inesc-id.pt](mailto:anabela.barreiro@inesc-id.pt)