

Port4NooJ: an open source, ontology-driven Portuguese linguistic system with applications in machine translation

Anabela Barreiro

barreiro_anabela@hotmail.com

FLUP & CLUP-Linguatca
New York University



Presentation Outline

1. Original Sources
2. Dictionaries
3. Inflectional and Derivational Description
4. Derivation
5. Syntactic-Semantic Ontology
6. Grammars
 - Morphological
 - Syntactic-Semantic
 - Disambiguation
7. Applications
 - Translation
8. Preliminary Quantitative Results
9. Future Goals
10. How to get Port4NooJ
11. Questions?

Original Sources

- OpenLogos English-Portuguese dictionary

<http://logos-os.dfki.de/>

OpenLogos is an open-source derivative of the Logos Machine Translation System

- NooJ linguistic environment

<http://www.nooj4nlp.net/>

Dictionaries

Part of Speech Inflectional Paradigm Syntactic-Semantic Attributes English Transfer

Lemma

mesa, N+FLX=CASA+CO+surf+EN=table
 cair, V+FLX=ATRAIR+INMO+IntoType+EN=fall
 holandês, A+FLX=INGLÊS+AN+lang+EN=Dutch
 actualmente, ADV+FLX=FACILMENTE+TEMP+punc+pres+EN=nowadays
 alguém, PRO+IMPERS+INDEF+EN=somebody
 porque, RELINT+why+EN=why
 e, CONJ+JOIN+EN=and
 durante, PREP+TEMP+EN=during
 cada, DET+IMPERS+INDEF+SG+EN=each
 terceiro+NUM+ord+EN=one third

General dictionary sample representing all PoS, variable and invariable forms

a curto prazo, ADV+TEMP+EN=in the short run
 a favor de, PREP+CAUS+EN=in favor of
 cada um, PRO+INDEF+SG+EN=each one
 de quem, INT+ThatType+EN=whose
 quem quer que seja, REL+WhateverType+EN=whoever
 além disso, CONJ+COOR+EN=besides
 um quarto, NUM+frac+EN=one fourth

Sample of invariable compounds in the general dictionary

HIV, N+FLX=PORTUGAL+AB+state+IMMUN+EN=HIV
 doença maniaco-depressiva, N+FLX=CASA+AB+state+MH+EN=manic-depressive disorder
 doença bipolar, N+FLX=CASA+AB+state+MH+EN=bipolar disorder
 asma, N+FLX=CASA+AB+state+PULM+EN=asthma

Sample of the dictionary of Biomedical Terms

adro da igreja, N+FLX=MENINO+PL+encl+EN=churchyard
 cabo de vassoura, N+FLX=MENINO+COtool+EN=broomstick
 bebida alcoólica, N+FLX=CASA+MA+liqu+EN=alcoholic drink+UNAMB
 bebida alcoólica, N+FLX=CASA+MA+liqu+EN=booze+slang
 cor de laranja, A+NAV+Apred+EN=orange
 sul-americano, A+FLX=ALTO+AN+des+EN=South American
 a curto prazo, ADV+LocTime+TEMP+EN=in the short run
 fora de serviço, ADV+STAT+pnr+EN=out of order
 há muito tempo, ADV+LocTime+TEMP+puncpast+EN=a long time ago
 isto é, CONJ+COOR+EN=i.e.
 já não, CONJ+COOR+EN=no longer
 mesmo assim, CONJ+SUB+EN=even so
 juntamente com, PREP+ASSOC+EN=along with
 à direita de, PREP+Loc+AT+EN=at the right of
 em conformidade com, PREP+ALOG+EN=in congruence with

Sample of the dictionary of Multiword Expressions

Amsterdão, N+PL+city+EN=Amsterdam
 Estados Unidos da América, N+PL+count+EN=United States of America
 África, N+PL+cont+EN=Africa
 Extremo Oriente, N+PL+othprop+EN=Far East
 Mediterrâneo, N+FLX=ANO+PL+water+EN=Mediterranean
 Alpes Peninos, N+FLX=ALPES+PL+othprop+EN=Pennine Alps
 ONU, N+AN+org+EN=UN

Sample of the dictionary of Proper Names

Inflectional and Derivational Description

NooJ Community Edition

File Edit Lab Project Windows Info

Noun Inflectional Paradigm

```
# plural adding -s
# Logos PAT 99
ANO = <E>/m+s + s/m+p

# plural adding -s
# Logos PAT 100
CANSA = <E>/f+s + s/f+p
```

Verb Inflectional Paradigm

```
# regular ending in -ar (1st conjugation), like falar
FALAR = <E>/INF +
<E>/INFI+1+s + es/INFI+2+s + <E>/INFI+3+s + mos/INFI+1+p + des/INFI+2+p + em/INFI+3+p +
<B>do/PP+m+s + <B>dos/PP+m+p + <B>da/PP+f+s + <B>das/PP+f+p +
<B>ndo/G +
<B2> (o/PR+1+s + as/PR+2+s + a/PR+3+s + amos/PR+1+p + ais/PR+2+p + am/PR+3+p) +
<B2> (ei/PS+1+s + aste/PS+2+s + ou/PS+3+s + ámos/PS+1+p + astes/PS+2+p + aram/PS+3+p) +
<B> (va/PI+1+s + vas/PI+2+s + va/PI+3+s + <B>ávamos/PI+1+p + <B>áveis/PI+2+p + vam/PI+3+p) +
<B2> (ara/PMP+1+s + aras/PMP+2+s + ara/PMP+3+s + áramos/PMP+1+p + áreis/PMP+2+p + aram/PMP+3+p) +
ei/F+1+s + ás/F+2+s + á/F+3+s + emos/F+1+p + eis/F+2+p + ão/F+3+p +
ia/C+1+s + ias/C+2+s + ia/C+3+s + íamos/C+1+p + íeis/C+2+p + iam/C+3+p +
<B2> (e/PRS+1+s + es/PRS+2+s + e/PRS+3+s + emos/PRS+1+p + eis/PRS+2+p + em/PRS+3+p) +
<B2> (asse/PIS+1+s + asses/PIS+2+s + asse/PIS+3+s + ássemos/PIS+1+p + ásseis/PIS+2+p + assem/PIS+3+p) +
<B2> (ar/FS+1+s + ares/FS+2+s + ar/FS+3+s + armos/FS+1+p + ardes/FS+2+p + arem/FS+3+p) +
<B2> (a/IMP+2+s + e/IMP+2+s + emos/IMP+1+p + ai/IMP+2+p + em/IMP+3+p) ;
```

Adjective Inflectional Paradigm

```
# ending in -o, like alto
# Logos PAT 123
ALTO = <E>/m+s + s/m+p +
<B>a/f+s + <B>as/f+p +
<B>inho/Dim+m+s + <B>inhos/Dim+m+p +
<B>inha/Dim+f+s + <B>inhas/Dim+f+p +
<B>ão/Aum+m+s + <B>ões/Aum+m+p +
<B>ona/Aum+f+s + <B>onas/Aum+f+p +
<B>íssimo/Sup+m+s + <B>íssimos/Sup+m+p +
<B>íssima/Sup+f+s + <B>íssimas/Sup+f+p ;
```

Adverb Inflectional Paradigm

```
# ending in -velmente; like amigavelmente
# Logos PAT 175
AMIGAVELMENTE = <E>/Reg +
<B8>bilíssimamente/Sup ;
```

Determiner Inflectional Paradigm

```
# definite article, like o
# Logos PAT 151
O = <E>/m+s +
<B>a/f+s +
s/m+p +
<B>as/f+p ;
```

Pronoun Inflectional Paradigm

```
# demonstrative pronoun, like este/esse/aquele
# Logos PAT 155
ESSE = <E>/m+s + s/m+p +
<B>a/f+s + <B>as/f+p ;

# indefinite pronoun, like um/algum
# Logos PAT 156
```

Interrogative Pronoun Inflectional Paradigm

```
# interrogative pronoun, like qual
# Logos PAT 251
QUAL = <E>/m+s + <B>is/m+p +
<E>/f+s + <B>is/f+p ;
```

Nominalization Derivational Paradigm

```
# 3são:
# decidir > decisão
# compreender > compreensão
# FLX=CANÇÃO
NDRV01 =
<B3>são/N+Nom;
```

Derivation

Verb entries:

- Identification of derivational paradigms for **nominalizations** (annotation *NDRV*) and **predicate adjectives** (annotation *ADRV*)

adaptar,V+FLX=FALAR+Aux=1+INOP57+Subset=132+EN=adapt+VSUP=fazer+DRV=NDRV00:CANÇÃO

azedar,V+FLX=LIMPAR+Aux=1+OBJTRundif98+Subset=740+EN=sour+VCOP=estar+DRV=ADRV00:ALTO

- Link to the derived noun's support verbs and to the adjective's copula verbs (annotation *VSUP* and annotation *VCOP*)

adaptar,V+FLX=FALAR+Aux=1+INOP57+Subset=132+EN=adapt+VSUP=fazer+DRV=NDRV00:CANÇÃO

azedar,V+FLX=LIMPAR+Aux=1+OBJTRundif98+Subset=740+EN=sour+VCOP=estar+DRV=ADRV00:ALTO

Derivation

Adjective entries:

- Identification of derivational paradigms for **adverbializations** (annotation *AVDRV*)

rápido,A+FLX=RÁPIDO+AV+state+EN=rapid+**DRV=AVDRV05:RAPIDAMENTE**

veloz,A+FLX=FELIZ+AV+state+EN=rapid+**DRV=AVDRV00:FELIZMENTE**

Autonomous predicate nouns, adjectives and adverbs are lexicalized:

- Identification of autonomous predicate nouns (annotation *Npred*), adjectives (*Apred*) and adverbs (annotation *ADV* + SynSem)

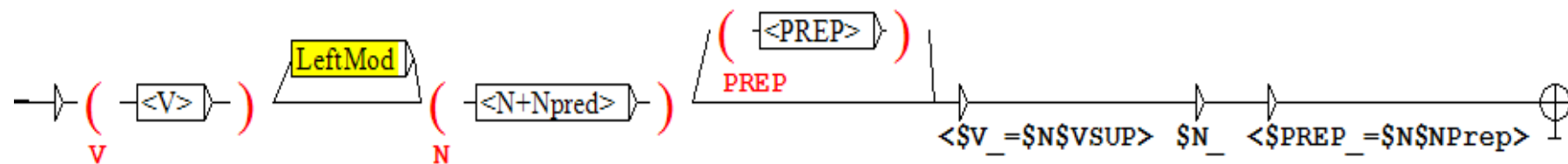
curso,N+FLX=ANO+**Npred**+IN+inst+EN=course+VSUP=tirar+VRB=estudar+NPrep=de+Det=um

adiposo,A+FLX=ALTO+NAV+**Apred**+EN=adipose+VSUP=ser

árabe,A+FLX=ÁRABE+NAV+**Apred**+EN=arabic +VSUP=ser

agora,**ADV**+Time+TEMPpuncpres+EN=now

Derivation



gosto de ver o comboio a	fazer corridas /correr	à velocidade máxima ao long
o de cheque especial para	fazer doações /doar	às entidades que escolher. A
cores e, quando é preciso ir	fazer filmagens/filmar	fora do estúdio, às vezes fic
se queria trocar de pares e	fazer um jogo /jogar	ao melhor de três sets , mas
o dra deu-me um papel para	fazer uma lista de/listar	todas as coisas boas que ex
res foram à caracterização	fazer uns retoques/retocar	, outros estão a descansar n

Syntactic-Semantic Ontology

- Representation abstract language
- Hierarchical taxonomy (sets, supersets and (sometimes) subsets)
- Based on Logos SAL ontology
- Integrated in the dictionary
- It represents both meaning (semantics), and structure (syntax)
- Over 1,000 categories

Syntactic-Semantic Ontology

Sets and **Subsets** of the **CONCRETE Noun Superset**
Click on **CONCRETE Superset**, **sets** and **subsets** for explanations

Noun Supersets

concrete
mass
animate
place
information
abstract
process (intr)
process (tr)
measure
time
aspective

functionals

- receptacles
- bearing surfaces
- links/bridges
- thresholds, focal points, barriers
- conduits
- fasteners
- devices, tools
- cloth thing
- structural elements
- concretizations of verbals
- concretizations of mass nouns
- undifferentiated functionals
- product/brand names

* * *

agentives

- software
- vehicles
- meters
- machines/systems
- communication agents
- concrete chemical agents
- undifferentiated agentives

* * *

natural things

- minute flora
- plants
- trees
- trees/wood
- miscellaneous natural things

* * *

other concrete sets*

- impulses/lights
- blemishes/marks
- edibles (non-mass)
- edibles/color
- classifiers
- amorphous
- atomistic
- undifferentiated concrete things

* * *

*With one exception, these sets have no subsets

Syntactic-Semantic Ontology

Category	Mnemonic	Examples in English	Examples in Portuguese
agentives	CO+undagt	See subsets	See subsets
software	CO+soft	<i>routine</i>	<i>rotina, ficheiro</i>
concrete chemical agents	CO+chem	<i>catalyst, warhead</i>	<i>ácido sulfúrico</i>
machines/systems	CO+mach	<i>battery, camera</i>	<i>máquina fotográfica</i>
vehicles	CO+vehic	<i>truck, ship</i>	<i>automóvel</i>
meters	CO+meter	<i>clock, gauge</i>	<i>manómetro</i>
communication agents	CO+comm	<i>radio, radar</i>	<i>rádio</i>
functionals	CO+undfunc	<i>trinket, ornament</i>	<i>ornamento</i>
devices/tools	CO+tool	<i>pliers</i>	<i>alicate</i>
fasteners	CO+fast	<i>nail, tendon</i>	<i>prego</i>
bearing surfaces	CO+surf	<i>table, shelf</i>	<i>mesa</i>
receptacles	CO+recp	<i>bottle, barrel</i>	<i>garrafa</i>
conduits	CO+cond	<i>chute, artery</i>	<i>artéria</i>
thresholds/focal points/barriers	CO+barr	<i>wall, door</i>	<i>porta</i>
links/bridges	CO+link	<i>circuit, nerve</i>	<i>circuito</i>
cloth things	CO+cloth	<i>shirt, blanket</i>	<i>camisola</i>
structural elements	CO+struc	<i>spar, bone</i>	<i>osso</i>
concretizations of verbals	CO+verb	<i>threading</i>	
concretizations of mass nouns	CO+mass	<i>acid lining</i>	
product/brand names	CO+brand	<i>Windows NT</i>	<i>Windows NT</i>
natural things	CO+nat	See subsets	See subsets
minute flora	CO+flora	<i>algae, spore</i>	<i>alga</i>
plants	CO+plant	<i>rose, weed</i>	<i>erva</i>
trees	CO+tree	<i>apple, willow</i>	<i>macieira</i>
trees/wood	CO+trwd	<i>oak, maple</i>	<i>carvalho</i>
misc. natural things	CO+mnat	<i>pebble, iceberg</i>	<i>iceberg</i>
edibles (non-mass)	CO+ednm	<i>pork chop</i>	<i>costoleta</i>
edibles/color	CO+edcol	<i>orange, cherry</i>	<i>laranja</i>
impulses/lights	CO+ight	<i>lamp, beam</i>	<i>lâmpada</i>
blemishes/marks	CO+blem	<i>scratch, freckle</i>	<i>sarda</i>
classifiers	CO+class	<i>element</i>	<i>elemento</i>
amorphous	CO+amor	<i>breeze, tide</i>	<i>brisa</i>
atomistic	CO+atom	<i>electron, atom</i>	<i>átomo</i>

Categories of
CONCRETE nouns

Syntactic-Semantic Ontology

ME - MEASURE Noun Sets and Subsets		
Sets and Subsets	Mnemonics (= SynSem)	Examples
abstract concepts measured by unit	ME+abs	<i>humidity, length</i>
discrete measurable concepts	ME+dis	<i>sum, increment</i>
units of measure	ME+unit	See subsets
units of weight	ME+unit+wt	<i>ounce, pound</i>
units of velocity	ME+unit+vel	<i>mph, megahertz</i>
units of volume measure	ME+unit+vol	<i>gallon, liter</i>
units of temperature	ME+unit+temp	<i>degrees celsius</i>
units of energy/force	ME+unit+ener	<i>watt, horsepower</i>
measurement systems	ME+unit+sys	<i>fahrenheit, kelvin</i>
units of duration	ME+unit+dur	<i>hour, minute, year</i>
specialized units of measure	ME+unit+spec	<i>oersted, ohm, phon</i>
units of money/value	ME+unit+value	<i>dollar, euro, forint</i>
units of linear/area measure	ME+unit+lin	<i>inch, yard, mile</i>
general undifferentiated measure	ME+undif	<i>degree, gross, share</i>

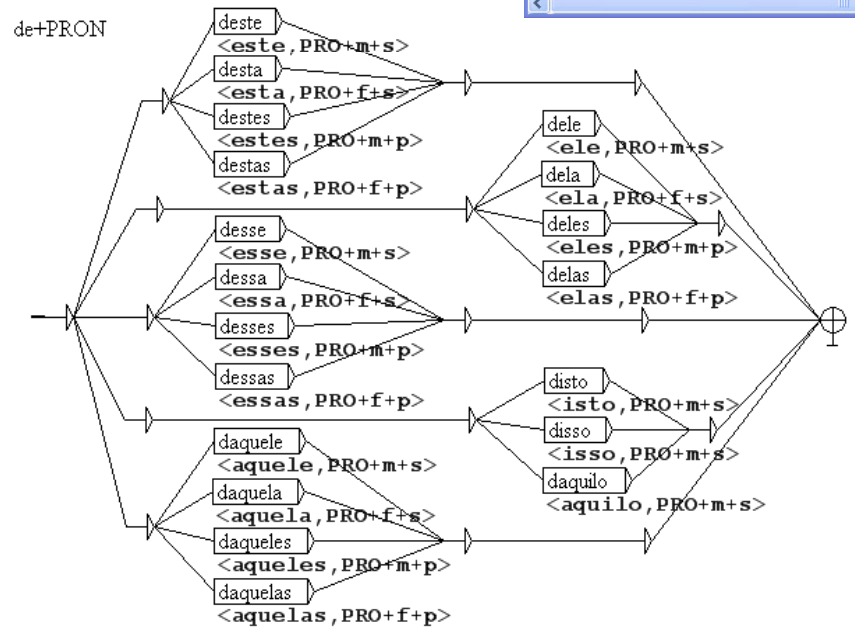
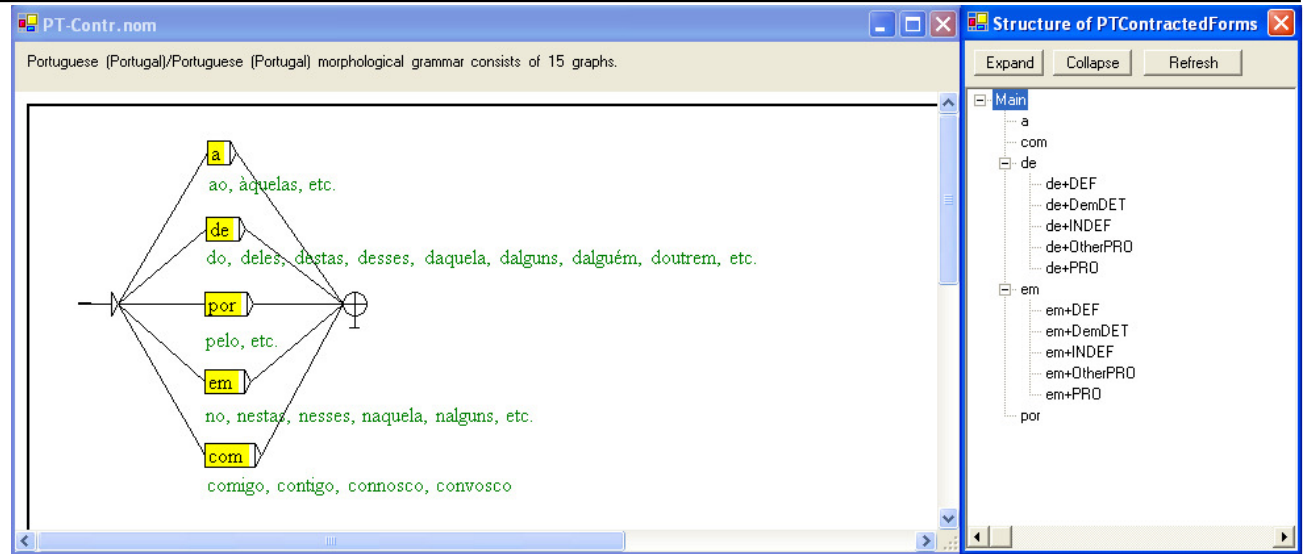
Categories of
MEASURE nouns

Grammars

- Morphological
- Syntactic-Semantic
- Disambiguation
- Translation

Morphological Grammar

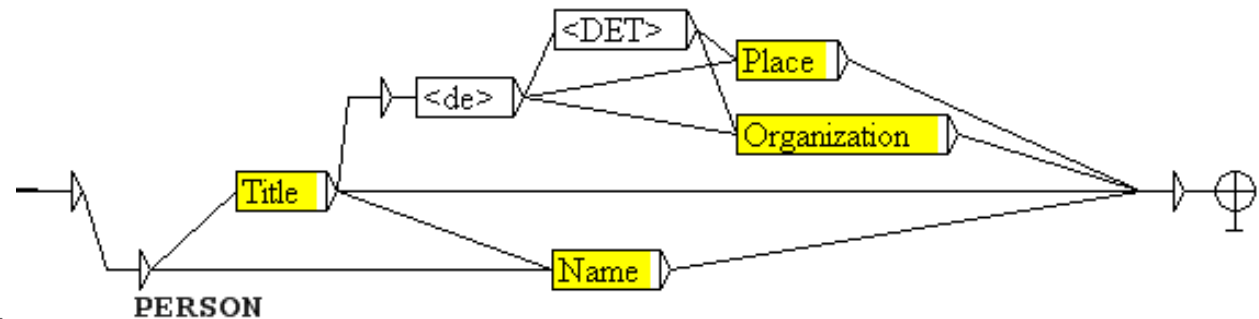
Morphological grammar for Portuguese contracted forms



Graph for contracted forms resulting from preposition *de* with demonstrative and personal pronouns

Syntactic-Semantic Grammar

Graph for recognition and annotation of *PERSON*-type named entities



certo ponto, o Dom Quixote/ PERSON+Title+Name
 o. O que eu fiz na Dona Branca/ PERSON+Title+Name
 A Dona Branca/ PERSON+Title+Name
 esus, logo o das Donas/ PERSON+Title
 - S. Frei Gil e o Doutor Fausto/ PERSON+Title+Name
 antes e depois do Doutor Fausto/ PERSON+Title+Name
 amei já o nosso Doutor Fausto/ PERSON+Title+Name
 vê que o nosso doutor/ PERSON+Title
 impressão era a Duquesa de Abrantes/ PERSON+Title+Place
 que lhe deram. - A Duquesa de Abrantes/ PERSON+Title+Place
 adíssimo bobo de el-rei de Dinamarca/ PERSON+Title+Place
 e e o Condestável, el-rei D. Fernando/ PERSON+Title+Name
 meiro-ministro de el-rei D. José/ PERSON+Title+Name
 seu augusto amo el-rei de Dinamarca/ PERSON+Title+Place
 s promessas - como el-rei de Prússia/ PERSON+Title+Place
 o.- De como S. M. El-Rei de Dinamarca/ PERSON+Title+Place
 a, em que domina el-rei Sancho/ PERSON+Title+Name
 ssível. Desenganado el-rei/ PERSON+Title
 ho, beijou a mão à el-rei/ PERSON+Title
 nosso Carlos; e o enfermeiro/ PERSON+Title
 emada elegância do enfermeiro/ PERSON+Title
 - O hospital. - O enfermeiro/ PERSON+Title
 parte da nossa Espanha/+Place
 rase didática das Espanhas/+Place

três, Frei Soeiro, Frei Lopo e S. Frei Gil - faz quatro
 , depois o de S. Domingos, célebre pelo jazigo do no
 . - De como o A. foi ao túmulo do santo bruzo e o achou va
 . Mas sem Homero ou Goethe é que se não chega à rep
 : e é com efeito. Não lhe falta senão o seu Goethe.
 de bivaque, o soldado que lhe chamou maluco ao pensad
 . Mas em meia hora de conversação, de trato, descobriam-
 . - Chega- se enfim ao Vale de Santarém.
 , o que alguns anos depois ressuscitou em Sterne com tâ
 e a Rainha D. Leonor, Camões desterrado aqui, Frei Lu
 ? Por onde está Ixião e Tântalo, por onde demora Sísifo
 . Por pouco mais que se generalize o princípio,
 prometeu uma constituição; e não faltou ainda, porque,
 tinha menos juízo do que Yorick, seu bobo. - Doutrina deste
 . Depois há
 de que um poder sobre-humano não permitia que ele s
 , e dai tomou um dia o caminho de Santarém, chegou àquel
 que o velava, uma bela mulher de estatura não acima de
 que o velava. O quarto era com efei
 . - Georgina.
 é, geologicamente falando, já tão África, tão pouco
 chamou romance em endechas. Eu, adotando para ele,

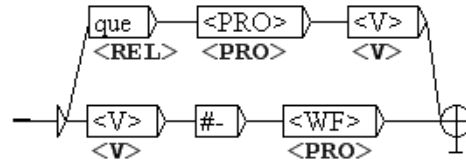
Concordance with annotation for *PERSON*-type named entities

Disambiguation Grammar

Ambiguity of the word "a"

1. a noun (*N*), masculine, singular, alphanumeric (*alnum*), corresponding to the EN word *a*, that is, the letter of the alphabet;
2. a preposition (*PREP*), locative (*Loc*), corresponding to the EN preposition *at*;
3. a pronoun (*PRO*), personal (*PERS*), direct object (*OD*), third person, singular, feminine, corresponding to the EN pronoun *her*;
4. a determiner (*DET*), definite article (*DEFART*), feminine singular and whose canonical form is *o*, corresponding to the EN article *the*.

Graph to annotate and disambiguate pronouns after verb and before verb after relative pronoun *que*



Concordance showing annotation of sequences <V> <PRO> and <REL> <PRO> <V>

utor prestou-se a
g-la ele mesmo,
rngu-a, aditou-a,
to, e como resolveu
e um bom charuto.
cias de um amigo,
por isso mesmo vou
das populações, e
ita reação antes de
ossos charutos, e
u cigarro de papel,
ribatejano.
raros pintos por
e forçado, assim
i a nossa gente
, e as ricas terras
- A força é
estão em terreno
nha - A força é
m homem do campo

dirigi-la/<V><PRO>
corrigiu-a/<V><PRO>
alterou-a/<V><PRO>
imortalizar-se/<V><PRO>
Travam-se/<V><PRO>
decidem-me/<V><PRO>
pronunciei-me/<V><PRO>
que se chama/<REL><PRO><V>
completar-se/<V><PRO>
deixe-mos/<V><PRO>
que me vai/<REL><PRO><V>
Acenderam-se/<V><PRO>
que se manifesta/<REL><PRO><V>
que o viu/<REL><PRO><V>
que o sachou/<REL><PRO><V>
que lhes levam/<REL><PRO><V>
que se fala/<REL><PRO><V>
que se lhe conzinha/<REL><PRO><V>
que se fala/<REL><PRO><V>
que se deita/<REL><PRO><V>
Declararam-se/<V><PRO>
mostra-se/<V><PRO>
que ele esteja/<REL><PRO><V>
que se mete/<REL><PRO><V>
que se explica/<REL><PRO><V>
personalizar-se/<V><PRO>
simbolizar-se/<V><PRO>
representar-se/<V><PRO>
que se esqueça/<REL><PRO><V>
Parece-me/<V><PRO>

civilização: e
rao leitor, para
moda germânica,
em tudo e com
, e que pode bem
em personalizar-se,
utopias, pode bem
porque receio muito

ele mesmo, corrigiu-a, aditou-a, alterou-a em muitas
, aditou-a, alterou-a em muitas partes, e a ilustrou em
muitas partes, e a ilustrou com as notas mais in
escrevendo estas suas viagens. Parte para Santarê
de razões os ilhaves e os Bordas-d'Água: os da calç
as tonleiras de um jornal, que por mezenquitei quis
São 17 deste mês de julho, ano
a si mesma por excelência a Sociedade, os seus passeio
... No entretã
os precintos aristocráticos da ré, à proa, que é país d
emprestar lume. - Dou-lho eu, senhor
os charutos, e atentamos mais devagar na companhia
o sempre clamoroso e sempre vazio entusiasmo das multít
) Isto é um fidalgo como se quer. Nunca o viu numa ferr
plantou, e o fez o que é, e fez terra das areias da charn
as enchentes. Mas nos , pé no barco, pé na terra, tã
- tornou o campino para estabelecer a questão em t
- A força é que se fala um homem do campo que se deita
um homem do campo que se deita ali à cernelha de um toi
ali à cernelha de um touro que uma companhia inteira de
típicas, simbólicas e míticas estas viagens. Faz o A.
como ela é dirigida pelo cavaleiro da Mancha, D.
prevendo, não cuide que são quaisquer dessas rabis
hoje em tudo e com que se explica tudo... quanto se nã
tudo... quanto se não sabe explicar.
, simbolizar-se pelo famoso mito do cavaleiro da mancha, D.
pelo famoso mito do cavaleiro da mancha, D. Quixote;
pela rotunda e anafada presença do nosso amigo velho,
Somos chegados ao triste desembarc
estar mais deserto e suio, mais abandonado e em ruín

temura, beijou-a muitas vezes, e tomando-lhe o novelo
das mãos num instante desembaraçou o fio e lho tronou a
entregar.
A velha sorria com aquele sorriso satisfeito que
exprime os tranqüilos gozos de alma, e que parecia dizer:
- Como eu sou feliz ainda, apesar de velha e de
cegal Bendito sejas meu Deus.
Esta última frase, esta bênção de um coração
agradecido que espira suavemente para o céu como sobe do
altar o fumo do incenso consagrado, esta última frase
transbordou-lhe e saiu articulada dos lábios.
- Bendito seja Deus minha filha minha Ioaquina

Ambiguous a

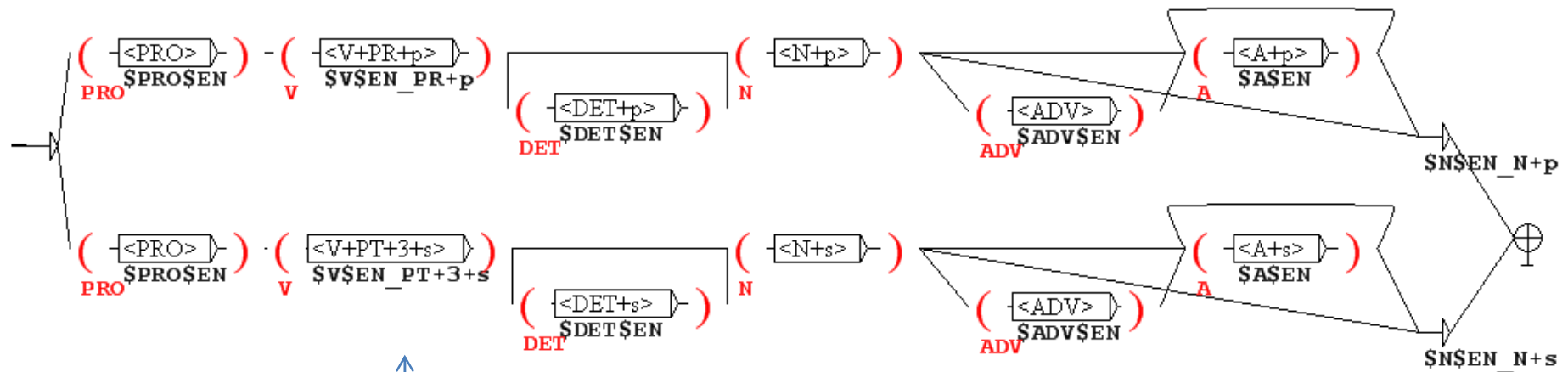
Disambiguated a

Word	Annotation
beijar, V+Aux=1+OBHUM68+Subset=570+EN=kiss+PS+3+s	a,N+UNalnum+EN=a+m+s
beijar, V+Aux=1+RECTRinclude-type+Subset=747+EN=oscultate+PS+3+s	a,PREP+Loc+ATat+EN=at
beijar, V+Aux=1+Reflex=4+INOPfor46+Subset=570+EN=kiss+PS+3+s	a,PRO+PERS+OD+3+s+f+EN=her
beijar, V+Aux=1+Reflex=4+INOPfor46+Subset=747+EN=oscultate+PS+3+s	o,DET+DEFART+the+EN=the+f+s
beijar, V+Aux=1+Reflex=4+INOPmisc29+Subset=270+EN=smooch+PS+3+s	

Applications

- Information Extraction
- Summarization
- Paraphrasing
- Pre-Editing of Texts
- Controlled Writing
- **Machine Translation**

Translation Grammar



Graph to translate simple sentences

Translation and bilingual paraphrasing of simple sentences

Concordance for Text PTSimpleSentences-3pl.not

Clear Concordance characters before, and after. Display: Inputs Outputs

Text	Before	Seq.	After
	Elas visitam cidades.	Eles são homens honestos/they are honest men	. Elas são mulheres simpáticas. Eles
	Elas visitam cidades.	Eles são homens honestos/they are straightforward men	. Elas são mulheres simpáticas. Eles
	comida. Eles vendem sapatos.	Elas compram casas/they purchase houses	. Elas sonham acordados. Eles sonham
	comida. Eles vendem sapatos.	Elas compram casas/they purchase homes	. Elas sonham acordados. Eles sonham
	comida. Eles vendem sapatos.	Elas compram casas/they buy houses	. Elas sonham acordados. Eles sonham
	comida. Eles vendem sapatos.	Elas compram casas/they buy homes	. Elas sonham acordados. Eles sonham

Query 6/29

Preliminary Quantitative Results

Dictionaries	Nouns	Verbs	Adjectives	Adverbs	Other PoS	Total (Lemmas)
General Dictionary	30,000	11,000	2,800	4,700	11,500	60,000
Named Entity Dictionary	6,000	-	-	-	-	6,000 *
Multiword Expression Dictionary	20,000	10,000	5,000	5,000	-	40,000 *
Total	-	-	-	-	-	106,000

Grammars	---	---	---	---	---	Total
Portuguese-English SVC Grammars	---	2,000	---	---	---	2,000 *
Other Portuguese-English Multiword Expression Grammars	---	---	---	---	---	5,000
Total	---	---	---	---	---	7,000

* Not available as open source, but will be available in the annotation system

Future Goals

- Extend and Enhance dictionaries
- Enhance 6,000 existing Portuguese-English local grammars and convert over 40,000 for other language pairs
- Extend Bilingual Dictionary to multilingual with inclusion of French, Italian, Spanish and German targets (approx. 60,000 entries each language)
- Develop controlled language writing tools
- Develop Portuguese to English machine translation

How to Get Port4NooJ

Publicly available at:

<http://www.nooj4nlp.net>

<http://www.linguateca.pt/Repositorio/Port4Nooj/>

Questions?

Thank you for your attention!

Anabela

