

III Simpósio Doutoral da Linguateca

Análise e Representação de Construções Adjectivais para Processamento Automático de Texto

Objectivos da Tese e Resultados Alcançados

Paula Carvalho
Faculdade de Letras da UL
pcfcarvalho@yahoo.com
Orientadora: Elisabete Ranchhod

Objecto de estudo

Análise e formalização das propriedades sintáctico-semânticas dos **adjectivos-nomes** em Português, tendo em vista o tratamento automático das construções em que os mesmos podem ocorrer.

- Construções humanas
 - O Zé é um (indivíduo + E) (**jovem + impostor + mal-educado**)
 - Os (eleitores + E) **franceses e holandeses** rejeitaram o Tratado da Constituição Europeia
- GN lexicalizados (nomes compostos)
 - O Zé não votou nas últimas (eleições + E) (**legislativas + presidenciais**)
 - O Zé extraiu um (dente + E) (**canino + incisivo + molar + queixal**)

Os Adj-N nas Gramáticas

Lindley Cintra (1986)

Trata-se de um processo geral de formação de palavras - **derivação imprópria** -, o qual explica a possibilidade de uma dada unidade lexical adquirir um novo estatuto linguístico, em função da sua distribuição na frase.

«basta, por exemplo, antepor-se um artigo a qualquer vocábulo da língua para que ele se torne um substantivo» (1986:105)

O Zé é jovem (Adj)	O Zé é católico (Adj)
O Zé é um jovem (N)	O Zé é um católico (N) dedicado
Os jovens (N) não se manifestaram	Os católicos (N) não se manifestaram

Os Adj-N nas Gramáticas

Mateus et al. (2003)

Adjectivos como, por exemplo, **rico, pobre, novo, velho, solteiro, casado, racista, português** e **verde** possuem igualmente valor nominal, tendo em conta que:

- Admitem modificação adjectival
- Designam grupos humanos
- São referenciais (isto é, têm sentido genérico)
- Aparecem geralmente na forma masculina e sempre no plural

Os (ricos + velhos) arrogantes

MAS:

Os estudantes **ricos arrogantes**

O Zé detesta esse velho arrogante

Os Adj-N nos Dicionários Tradicionais

impostor
adjectivo e substantivo masculino

- que ou aquele que usa de impostura; embusteiro; charlatão;
- que ou aquele que se faz passar por quem não é; mentiroso;
- vaidoso; hipócrita;
- caluniador;
- propagandista de falsas doutrinas

jovem
adjectivo 2 géneros

- que tem pouca idade; novo; moço;
- que existe há pouco tempo; recente;
- figurado* que mantém a frescura, a energia e o aspecto característicos de quem tem pouca idade;
- pejorativo* imaturo;
- que se destina ou se adequa a pessoas com pouca idade;
- que é feito ou criado por pessoas com pouca idade;
- (*animal, árvore*) que está em fase de crescimento;

substantivo 2 géneros
pessoa com pouca idade; pessoa que não chegou à idade adulta ou que a atingiu há pouco tempo

© Copyright 2003-2006, Porto Editora

Os Adj-N nos Dicionários Tradicionais

católico
adjectivo

- RELIGIÃO que professa o catolicismo;
- popular, figurado* bom de saúde;
- popular, figurado* bem-disposto;
- universal;

substantivo masculino
o que segue a religião que tem o Papa como chefe

budista
adjectivo 2 géneros
relativo ao budismo;
substantivo 2 géneros
pessoa seguidora do budismo

nazi
adjectivo e substantivo 2 géneros

- POLÍTICA relativo ao nazismo;
- POLÍTICA que defende a ideologia do nacional-socialismo; que segue o nazismo;

substantivo 2 géneros
pessoa partidária do nazismo; nacional-socialista

© Copyright 2003-2006, Porto Editora

Os Adj-N nos Dicionários Tradicionais

asmático

adjectivo e substantivo masculino
que ou aquele que padece de asma

autista ¹

substantivo 2 géneros
pessoa que possui autismo;
adjectivo 2 géneros
relativo a autismo

anorético

adjectivo
1. relativo à anorexia;
2. que sofre de anorexia

psicopata

substantivo 2 géneros
pessoa que padece de psicopatia

© Copyright 2003-2006, Porto Editora

Os Adj-N nos Dicionários Tradicionais

alemão

adjectivo
relativo à Alemanha;
substantivo masculino
1. natural ou habitante da Alemanha;
2. idioma falado na Alemanha, Áustria, Suíça, etc.

açoriano

adjectivo
1. pertencente ou relativo aos Açores;
2. pertencente ou relativo aos habitantes dos Açores;
substantivo masculino
natural ou habitante dos Açores

raiano

adjectivo e substantivo masculino
que ou o que nasceu ou mora na arraia (fronteira)

© Copyright 2003-2006, Porto Editora

Os Adj-N nos Dicionários Tradicionais

A classificação gramatical dos adjectivos-nomes nos dicionários tradicionais revela-se, de um modo geral, assistemática e incoerente.

- Em alguns casos, os **adjectivos-nomes** estão representados:
 - em **entradas diferentes**, onde se prevê, de forma independente, o estado gramatical (**N** e **Adj**) que os mesmos podem assumir;
 - numa só **entrada**, que alberga a dupla classificação de **N** e **Adj**.
- Noutros casos, porém, unidades lexicais com características idênticas encontram-se dicionarizadas apenas com uma daquelas informações gramaticais.

Os Adj-N nos Dicionários Electrónicos

Nos dicionários electrónicos, em regra mais sistemáticos e coerentes do que os tradicionais, os **Adj-N** são frequentemente tratados como duas entradas, que se distinguem apenas pela informação gramatical: **N** e **A(dj)**.

açoriano, A001+Pna+z1

açoriano, N001+z1

alemão, A025+Pna+z1

alemão, N025+z1

alemão, N291+z1

asmático, A001

asmático, N001+z1

budista, A101

budista, N101+z1

católico, A001+z1

católico, N001+z1

impostor, A005+Pd

impostor, N005+z1

jovem, A110.dh015.dt015+Pde

jovem, N110+z1

nazi, A101+Pd

nazi, A101+Rel

nazi, N101+z1

psicopata, A101

psicopata, N101+z1

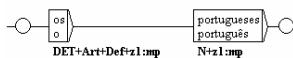
raiano, A001

raiano, N001+z1

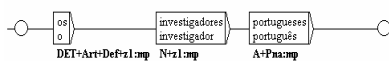
LABEL-LEX-sw ©, <http://label.ist.utl.pt>

O processamento dos Adj-N

Os **portugueses** estão assustados com o desemprego e acham que a economia irá de mal a pior.



Os investigadores **portugueses** quase não os escrevem e as crianças vêem-nos mais do que os lêem.



FSTs resultante da aplicação das gramáticas de resolução de ambiguidades ao texto, previamente etiquetado pelos recursos lexicais

LabEL/Intex

O processamento dos Adj-N

P1395-12 Os **portugueses** estão assustados com o desemprego e acham que a economia irá de mal a pior.

```
A1
STA:fol
S:np=
DN:pron-undef("o" <artd> DET M P) Os
=H:n("português" <Hnat> M P) portugueses
X:par
=CJT:x
==P:v-fin("estar" <predco> <fmc> <mv> PR 3P IND VFIN) estão
==Cs:icl==P:v-ppc("assustar" <mv> M P) assustados (...)
```

CP2343-4 Os **investigadores portugueses** quase não os escrevem e as crianças vêem-nos mais do que os lêem.

```
A1
STA:par
CJT:fol
=S:np
==DN:pron-undef("o" <artd> DET M P) Os
==H:n("investigador" <Hprof> M P) investigadores
==DN:adj("português" <nat> <np-close> M P) portugueses
=IA:advp (...)
```

FlorestaVirgem_CP_3.0, E. Bick

Argumentos para a análise dos Adj-N como Adj

- O processo de lexicalização pressupõe, entre outras características, a aquisição de um novo **significado**. Porém, o valor básico dos *Adj-N* parece manter-se inalterado, qualquer que seja o contexto em que ocorram.

O Zé é (idiota + racista + hipocondríaco)
O Zé é **um** (idiota + racista + hipocondríaco)

- Em Português, a **aposição** de **nomes** não é produtiva, a não ser que os mesmos se encontrem no contexto de uma estrutura lexicalizada (nome composto). No entanto, os *Adj-N* podem ser precedidos de um nome classificador apropriado.

O Zé é um (**indivíduo** + E) (idiota + racista + hipocondríaco)
Os (**indivíduos** + E) (idiotas + racistas + hipocondríacos) não falaram

Argumentos para a análise dos Adj-N como Adj

- A **coordenação** pressupõe um enorme paralelismo sintáctico-semântico entre os termos coordenados; por exemplo, de um modo geral, não é possível coordenar palavras pertencentes a categorias gramaticais diferentes. No que respeita aos *Adj-N*, observa-se que eles apenas podem ser coordenados com expressões adjectivais.

O Zé é **jovem** (e + mas) **dedicado**
?Os **jovens** (e + mas) **dedicados** conseguem sempre o que querem

*O Zé é **jovem** (e + mas) **médico**
*Os **jovens** (e + mas) **médicos**, conseguem sempre o que querem

*O Zé é **médico** (e + mas) **dedicado**
*Os **médicos** (e + mas) **dedicados** conseguem sempre o que querem

Argumentos para a análise dos Adj-N como Adj

- Certos adjectivos podem sofrer **quantificação** por um advérbio quantificador ou por um morfema de grau. No que respeita aos *Adj-N*, observa-se que eles podem construir-se com os morfemas de grau **o** (**mais + menos**), constituindo aquilo a que as gramáticas designam de *superlativo relativo de superioridade e inferioridade*, respectivamente. Já os nomes não podem, pela sua própria natureza, integrar tais construções.

O Zé é (**racista + jovem**)
O Zé é o (mais + menos) (**racista + jovem**) de todos
Os (**racistas + jovens**) ainda não se manifestaram
Os (mais + menos) (**racistas + jovens**) ainda não se manifestaram
*O Zé é o (mais + menos) (**médico + professor**) de todos
*Os (mais + menos) (**médicos + professores**) ainda não se manifestaram

Argumentos para a análise dos Adj-N como N

- Quando ocupam um contexto tipicamente nominal, os *Adj-N* parecem deixar de exibir certas propriedades que apresentariam quando se encontram noutra situação. Por exemplo, a quantificação pelo advérbio **muito** ou pelo morfema- **íssimo** nem sempre se mostra natural:

O Zé é **muito** (**idoso + gordo**)
*O Zé é um **muito** (**idoso + gordo**)
O Zé é **gordíssimo**
*O Zé é um **gordíssimo**
MAS
Os **muito idosos** não puderam comparecer
Os **muito gordos** têm dificuldade em encontrar roupa gira
?Os **gordíssimos** têm dificuldade em encontrar roupa gira

Argumentos para a análise dos Adj-N como N

- Os *Adj-N*, quando precedidos de determinante, podem ter à sua direita um modificador (*Adj, Relativa, ...*), como se de verdadeiros nomes se tratassem. Essa modificação não pode, contudo, ser exercida sobre o adjectivo, se o mesmo se encontrar em posição pós-verbo auxiliar.

O Zé é um **jovem** (**insolente + que tem cerca de 30 anos**)
*O Zé é **jovem** **insolente** (**insolente + que tem cerca de 30 anos**)
Os **jovens** (**insolentes + que têm cerca de 30 anos**) ...

Proposta de Análise

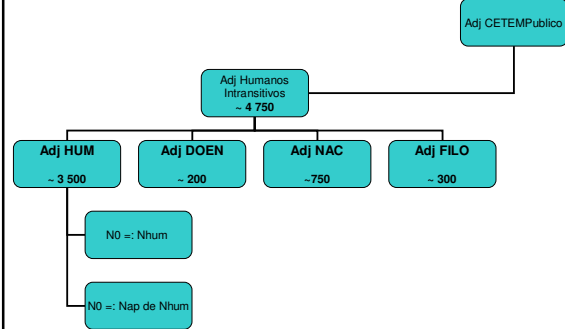
- Os *Adj-N* (**Humanos**) devem ser analisados como adjectivos.
- Em causa está um processo sintáctico produtivo, que se observa em várias línguas e afecta várias classes de palavras. Tal processo explica a redução lexical, em certas construções, de uma dada expressão (**nome apropriado**).
- Os nomes apropriados são passíveis de serem reconstituídos mediante a análise das propriedades linguísticas das expressões que os podem superficialmente representar e do contexto em que estão integradas.
- Factores de ordem sintáctico-semântica podem ajudar a perceber as razões pelas quais certos *Adj* podem ocupar a posição de nome reduzido e outros não.

Quadro Teórico-Metodológico

Léxico-Gramática (M. Gross, 1975, 1981), inspirado na Gramática Transformacional Harrisiana (Z. Harris, 1964, 1976, 1978)

- A unidade mínima de análise é uma frase elementar, que compreende um predicado e os seus argumentos (a atribuição de um valor sintáctico e semântico a uma dada expressão lexical depende da estrutura sintáctica em que se encontra integrada).
- É possível estabelecer relações entre duas ou mais frases elementares equivalentes por meio de um conjunto de transformações não-orientadas.
- As diferentes frases elementares de uma dada língua e o conjunto de transformações que sobre elas podem operar podem ser adequadamente descritas em tabelas (ou matrizes) sintácticas.
- As tabelas têm a vantagem de ser legíveis para o Humano e compactas para o computador.

Classes sintácticas de *Adj Hum* estudadas



Análise dos *Adj-N* Humanos como *Adj*

Na perspectiva da Gramática Transformacional:

- Os GN humanos podem ser entendidos como o resultado da concatenação de uma frase classificadora, que integra um nome classificador humano, com o predicado adjectival, mediante um processo de formação de adjunto adnominal.

O Zé é um indivíduo # Esse indivíduo é (idiota + racista)
[Rel] O Zé é um indivíduo que é (idiota + racista)
[Red Rel] O Zé um indivíduo (idiota + racista)
[Red Nap] O Zé é um (idiota + racista)

- A noção de **Nome Apropriado** (Nap), introduzida por Harris (1968, 1976) pode explicar a redução da cabeça destes GN.

Adjectivos Hum

- De um modo geral, os *Adj Hum* que se constroem com **ser** são os que mais naturalmente ocupam a posição de nome reduzido no GN:

O Zé é (indisciplinado + irresponsável + desonesto)
O Zé é um (indivíduo + E) (indisciplinado + irresponsável + desonesto)
Os (indivíduos + E) (indisciplinados + irresponsáveis + desonestos) ...

O Zé está (cansado + contente + triste)
*O Zé está um (indivíduo + E) (cansado + contente + triste)
?Os (cansados + contentes + tristes) ...

O Zé (é + está) (lindo + gordo + infeliz)
O Zé é um (indivíduo + E) (?lindo + gordo + infeliz)
Os (?lindos + gordos + infelizes) ...

Adjectivos Hum

- Contudo, nem todos os *Adj Hum* que se constroem com **ser** podem ocupar a posição de nome reduzido no GN.

O Zé é (disciplinado + responsável + honesto)
O Zé é um (indivíduo + *E) (disciplinado + responsável + honesto)
Os (indivíduos + ?E) (disciplinados + responsáveis + honestos) ...

- Num considerável número de casos, essa posição é preenchida por adjectivos que expressam **valor negativo** ou **depreciativo**.
- Ainda assim, é possível encontrar certos adjectivos *Hum* que não apresentam tais características na posição de cabeça do GN:

O Zé é (beijoqueiro + querido + sortudo)
O Zé é um (indivíduo + E) (beijoqueiro + querido + sortudo)
Os (indivíduos + E) (beijoqueiros + queridos + sortudos) ...

Adjectivos de Doença

- Os *Adj Doen* constroem-se com os auxiliares **ser** e/ou **estar**, de acordo com as propriedades semânticas e aspectuais que veiculam.

O Zé é (asmático + autista + canceroso)
O Zé está (constipado + febril + bexigoso)
O Zé (é + está) (anorético + artrítico + cego)

- Em geral, **ser** só se combina com predicados *Adj Doen* que remetem para doenças inatas ou incuráveis; **estar**, pelo contrário, aparece em construções cujas propriedades atribuídas ao sujeito pelo *Adj Doen* são de natureza transitória ou temporária.

Adjectivos de Doença

- A propriedade de quantificação e de variação do *Adj Doen* em grau observa-se preferencialmente em construções adjectivais com **estar**:

O *Zé* é **muito** (?asmático + *autista + *canceroso)
O *Zé* está **muito** (constipado + febril + bexigoso)

- A quantificação adjectival parece ser igualmente possível nos casos em que o *Adj Doen* está associado à manifestação de uma doença do foro psicológico ou comportamental ou quando as propriedades por ele transmitidas puderem ser quantificadas ou graduáveis numa determinada escala, remetendo para o estado de gravidade ou de evolução da doença.

O *Zé* é **muito** (esclerosado + hipocondríaco + neurótico)
O *Zé* é **muito** (diabético + estrábico + miope)

Adjectivos de Doença

- A presença dos *Adj Doen* em contexto pós-*Dind* depende frequentemente da presença de um modificador na construção.

O *Zé* é (asmático + artrítico + hipertenso)
O *Zé* é um (indivíduo + ?E) (asmático + artrítico + hipertenso)
O *Zé* é um (asmático + artrítico + hipertenso) corajoso

- Porém, nos casos em que os *Adj Doen* remetem para uma doença do foro psicológico, a presença obrigatória desse modificador deixa de se colocar.

O *Zé* é (anorético + esclerosado + hipocondríaco)
O *Zé* é um (indivíduo + E) (anorético + esclerosado + hipocondríaco)
O *Zé* é um (anorético + esclerosado + hipocondríaco) infeliz

Adj Nac e Adj Filo

- Contrariamente a alguns *Adj Hum* e *Adj Doen*, os *Adj Nac* e os *Adj Filo* não admitem:

- Construção com o verbo auxiliar **estar**;
- Quantificação nem variação em grau;

O *Zé* (é + *está) (açoriano + africano + alemão + arraiano)
O *Zé* (é + *está) (católico + comunista + cubista + darwinista)

*O *Zé* é (muito + pouco) (açoriano + africano + alemão + arraiano)
O *Zé* é **muito** (?católico + ?comunista + *cubista + *darwinista)

Adj Nac e Adj Filo

- A ocorrência dos *Adj Nac* em contexto pós-*Dind* implica a presença obrigatória de um modificador à sua direita.
- A maioria dos *Adj Filo* parece apresentar o mesmo comportamento.

O *Zé* é um (indivíduo + ?*E) (açoriano + africano + alemão + arraiano)

O *Zé* é um (indivíduo + E) (açoriano + africano + alemão + arraiano) simpático

O *Zé* é um (indivíduo + ?*E) (católico + comunista + cubista + darwinista)

O *Zé* é um (indivíduo + E) (católico + comunista + cubista + darwinista) convicto

Formalização das Propriedades Linguísticas

NO	Vaux	Class	Quant	Adjectivo	Reest GN			Exemplo	Npred
					Nhum	Adj Prep	Npc		
Nhum	Nhum	Npc de Nhum	estar	UM Indivíduo	UM	UM+Modif	Adv Quant	Morf Grau	
+	-	-	+	anorético	-	-	-	O <i>Zé</i> (é + está) anorético	anorexia
+	-	-	+	artrítico	-	-	-	O <i>Zé</i> (é + está) artrítico	artrite, artrismo
+	-	-	+	asmático	-	-	-	O <i>Zé</i> é asmático	asma
+	-	-	+	autista	-	-	-	O <i>Zé</i> é autista	autismo
+	-	-	+	astigmático	+	+	+	O <i>Zé</i> é astigmático	astigmatismo
+	-	-	+	bexigoso	-	-	-	O <i>Zé</i> está bexigoso	bexigas
+	-	-	+	bibliomano	-	-	-	O <i>Zé</i> é bibliomano	bibliomania
+	-	-	+	bolhoso	+	+	+	O <i>Zé</i> está bolhoso	bolhas
+	-	-	+	borbulhento	+	+	+	O <i>Zé</i> está borbulhento	borbulhas
+	-	-	+	borbulhoso	+	+	+	O <i>Zé</i> está borbulhoso	borbulhas
+	-	-	+	caloso	+	+	+	O <i>Zé</i> está caloso	calos, calosidade
+	-	-	+	canceroso	+	+	+	O <i>Zé</i> (é + está) canceroso	câncer
+	-	-	+	caquéctico	-	-	-	O <i>Zé</i> (é + está) caquéctico	caquexia
+	-	-	+	catiléptico	-	-	-	O <i>Zé</i> (é + está) catiléptico	catilépsia
+	-	-	+	cego	-	-	-	O <i>Zé</i> (é + está) cego	cegueira
+	-	-	+	clausrofóbico	-	-	-	O <i>Zé</i> é clausrofóbico	clausrofobia
+	-	-	+	cleptomaniaco	-	-	-	O <i>Zé</i> é cleptomaniaco	cleptomania
+	-	-	+	cleptômico	-	-	-	O <i>Zé</i> é cleptômico	cleptomania
+	-	-	+	colérico	-	-	-	O <i>Zé</i> está colérico	colera
+	-	-	+	constipado	-	-	-	O <i>Zé</i> está constipado	constipação

Formalização das Propriedades Linguísticas

Nloc	Nhum	Nhum	Npc de Nhum	estar	Class	Quant	Adjectivo	Exemplo	Nloc
+	+	+	+	+	-	-	açoriano	O <i>Zé</i> é açoriano	Açores
+	+	+	+	+	-	-	aflegio	O <i>Zé</i> é aflegio	Alegria
+	+	+	+	+	-	-	africano	O <i>Zé</i> é africano	África
+	+	+	+	+	-	-	aguiarense	O <i>Zé</i> é aguiarense	Aguiar da Beira
+	+	+	+	+	-	-	albanês	O <i>Zé</i> é albanês	Albânia
+	+	+	+	+	-	-	albano	O <i>Zé</i> é albano	Albânia, Aba Longa
+	+	+	+	+	-	-	albacarense	O <i>Zé</i> é albacarense	Castelo Branco
+	+	+	+	+	-	-	alcacerense	O <i>Zé</i> é alcacerense	Alcácer do Sal
+	+	+	+	+	-	-	alcantarense	O <i>Zé</i> é alcantarense	Alcântara
+	+	+	+	+	-	-	alcofocense	O <i>Zé</i> é Alcofocense	Alcofoba
+	+	+	+	+	-	-	alegalense	O <i>Zé</i> é alegalense	Alegria
+	+	+	+	+	-	-	alemão	O <i>Zé</i> é alemão	Alemanha
+	+	+	+	+	-	-	alenquense	O <i>Zé</i> é alenquense	Alenquer
+	+	+	+	+	-	-	alentijano	O <i>Zé</i> é alentijano	Alentejo
+	+	+	+	+	-	-	alexandrino	O <i>Zé</i> é alexandrino	Alexandria
+	+	+	+	+	-	-	alfamita	O <i>Zé</i> é alfamita	Alfama
+	+	+	+	+	-	-	algarvio	O <i>Zé</i> é algarvio	Algarve
+	+	+	+	+	-	-	alhândrense	O <i>Zé</i> é alhândrense	Alhândria
+	+	+	+	+	-	-	alcanino	O <i>Zé</i> é alcanino	Alcante
+	+	+	+	+	-	-	almadense	O <i>Zé</i> é almadense	Almada
+	+	+	+	+	-	-	almodovarense	O <i>Zé</i> é almodovarense	Almodovar
+	+	+	+	+	-	-	alpino	O <i>Zé</i> é alpino	Alpe
+	+	+	+	+	-	-	arraiano	O <i>Zé</i> é arraiano	Arraia

