

COMPUTAÇÃO DE AMBIGUIDADES SINTÁCTICAS EVIDÊNCIAS EM FAVOR DOS MODELOS BASEADOS EM CONHECIMENTO LINGUÍSTICO

Palmira MARRAFA

Departamento de Linguística Geral e Românica e CLG-Grupo de Computação do Conhecimento Léxico-Gramatical (Centro de Linguística) — Faculdade de Letras da Universidade de Lisboa

Alameda da Universidade, 1600-214 Lisboa, Portugal

Mél : Palmira.Marrafa@netcabo.pt

Resumo

Neste artigo discute-se a natureza das especificações linguísticas a integrar num analisador sintático capaz de dar conta de ambiguidades que envolvem estruturas de predicação não frásica e de modificação adjetival. A adequação descritiva do sistema depende crucialmente de uma caracterização fina das propriedades dos itens lexicais, só compatível com uma modelização baseada em conhecimento linguístico. Retira-se, assim, sustentação às abordagens estritamente probabilísticas, a despeito da sua utilidade em grande número de aplicações. A informação lexical é codificada através de estruturas de traços. O formalismo gramatical utilizado é basicamente uma gramática de cláusulas definidas, com pequenas alterações sintáticas.

Abstract

COMPUTATION OF SYNTACTIC AMBIGUITIES: EVIDENCE FOR KNOWLEDGE-BASED MODELING

This paper focuses on the nature of the linguistic specifications for a syntactic parser able to account for ambiguities involving non-sentential predication and adjectival modification. The descriptive adequacy of the system crucially depends on a fine-grained characterization of the properties of lexical items, which is only compatible with a linguistic knowledge-based modelling approach. It becomes evident that strictly probabilistic approaches lack a strong empirical support, despite their usefulness in a wide range of applications. Lexical information is encoded by means of feature structures. The formalism used to model grammatical information is a DCG-like formalism with slight syntactic modifications.

Resumen

COMPUTACIÓN DE AMBIGÜIDADES SINTÁCTICAS : EVIDENCIAS A FAVOR DE LOS MODELOS BASADOS EN EL CONOCIMIENTO LINGUÍSTICO

Este artículo trata de la naturaleza de las especificaciones linguísticas a integrar en un analizador sintático capaz de dar cuenta de ambigüedades provocadas por las estructuras de predicación no frásica y de modificación adjetival. La adecuación descriptiva del sistema depende en gran medida de que disponga de una caracterización precisa de las propiedades de las piezas léxicas, sólo compatible con una modelización basada en el conocimiento linguístico. Se resta así peso a los enfoques estrictamente probabilísticos, a pesar de su

utilidad en un gran número de aplicaciones. La información lexical se codifica mediante estructuras de rasgos. El formalismo gramatical utilizado es básicamente una gramática de cláusulas definidas, con pequeñas alteraciones sintácticas.

Résumé

ANALYSE D'AMBIGUÏTÉS SYNTAXIQUES : ARGUMENTS EN FAVEUR DE MODÈLES FONDÉS SUR LA CONNAISSANCE LINGUISTIQUE

Nous discutons dans cet article la nature des spécifications linguistiques qui doivent intégrer un analyseur syntaxique capable de rendre compte d'ambiguïtés concernant des structures de prédication non phrastique et de modification adjectivale. L'adéquation descriptive du système dépend fondamentalement d'une caractérisation fine des propriétés des mots, compatible seulement avec une modélisation fondée sur la connaissance linguistique. Malgré leur utilité dans un grand nombre d'applications, les résultats présentés mettent en évidence le manque de support empirique des approches strictement probabilistes. L'information lexicale est codée par le biais de structures de traits. Le formalisme grammatical utilisé est fondamentalement une grammaire DCG (grammaire à clauses définies), avec de petites modifications syntactiques.

Riassunto

RISOLUZIONE AUTOMATICA D'AMBIGUITÀ SINTATTICHE : PROVE PER I MODELLI A BASE DI CONOSCENZE LINGUISTICHE

In questo articolo si discute della natura delle specificazioni linguistiche di un parser sintattico capace di rendere conto delle ambiguità relative alla predicazione non-frastica e alla modificazione aggettivale. L'adeguatezza descrittiva del sistema dipende fondamentalmente da una fine caratterizzazione delle proprietà grammaticali dei lemmi, che è paragonabile solamente ad una modellizzazione basata sulla conoscenza linguistica. Risulta evidente che gli approcci strettamente probabilistici, malgrado la loro utilità in un numero di applicazioni, manchino di un adeguato supporto empirico. L'informazione lessicale è codificata grazie a strutture di tratti. Il formalismo usato per modellizzare le informazioni grammaticali è un formalismo simile a quello della DCG (grammatica a clausole definite), con qualche piccola modificazione sintattica.

1. Introdução

Para a história da modelização formal e computacional das línguas naturais concorrem desde o início, duas perspectivas distintas: uma orientada para o conhecimento linguístico, outra orientada para os dados.

Nesse contexto, surgem diferentes paradigmas de investigação, diversificam-se os objectivos e vão variando as ponderações atribuídas à adequação descritiva, à plausibilidade psicológica e à eficiência dos sistemas, requisitos nem sempre fáceis de conciliar.

Na última década ganham grande popularidade os modelos probabilísticos e os métodos empíricos, em geral, com aplicação em grande número de subdomínios do processamento da fala e da língua escrita.

Simultaneamente, vão-se aduzindo evidências em desfavor da perspectiva racionalista sobre a linguagem. Procura-se desmontar os argumentos baseados na

ausência de evidências negativas e na “pobreza dos estímulos” no que respeita aos contextos de aquisição das línguas naturais, comumente usados para sustentar o carácter inato da faculdade da linguagem. Recupera-se a ideia de Horning (1969) de que as gramáticas probabilísticas são adquiríveis a partir de evidências positivas unicamente, que encontra suporte em experiências recentes na área da Psicolinguística¹. Põe-se em causa a discrepância dos conceitos de gramatical vs. agramatical, defendendo-se antes que os juízos de gramaticalidade se situam num *continuum* e que são estatisticamente preditíveis². Aponta-se a inoperância dos chamados modelos categoriais e põe-se em causa a universalidade de grande parte dos princípios gramaticais. Nas abordagens mais radicais assume-se que toda a cognição humana é probabilística. E, logo, a linguagem, como parte integrante da cognição, também o é³.

Não cabe nos objectivos deste artigo discutir profundamente os argumentos em que se alicerça esta perspectiva. Vale, contudo, chamar a atenção para o facto de os modelos de lógica de traços não se confrontarem com o problema do monolitismo categorial e de os modelos que admitem a hierarquização e a conflitualidade entre princípios encontrarem explicação para a existência de diferentes níveis de gramaticalidade e para o facto de algumas propriedades serem observáveis em algumas línguas e não em outras, sem que seja posta em causa a universalidade das restrições. Assim sendo, afiguram-se infundadas as objecções maiores aos modelos simbólicos no seu conjunto. E naturalmente que perde peso a tese de que os juízos de gramaticalidade e os universais linguísticos só encontram justificação num quadro de predição estatística. Haverá, pois, que encontrar outro tipo de evidências para sustentar que a faculdade da linguagem não é mais do que um sistema probabilístico.

Como quer que seja, aqui apenas se pretende demonstrar a necessidade de um sistema de análise automática, capaz de dar conta de ambiguidades que envolvem estruturas de predicação não frásica e estruturas de modificação adjectival, integrar especificações relativas a um conjunto não trivial de restrições sintáctico-semânticas, dificilmente dissociáveis de uma gramática da competência.

Na secção 2. define-se a problemática e discutem-se os factos empíricos que suportam a perspectiva defendida. Em 3. descreve-se a arquitectura geral e o funcionamento de um analisador capaz de descrever adequadamente as estruturas em causa. Finalmente, em 4. conclui-se em favor da modelização baseada em conhecimento, em conformidade com as evidências.

2. Problemática

Um dos problemas maiores que se põem à modelização formal e computacional das línguas naturais respeita à representação e à resolução de ambiguidades, sejam de natureza lexical, sejam de natureza sintáctica.

¹ Em Saffran *et al.* (1996), por exemplo, descreve-se uma experiência com crianças de 8 meses cujos resultados são interpretados como evidências que favorecem a tese de mecanismos estatísticos inatos de aquisição da linguagem em detrimento do conhecimento inato.

² A este respeito, destacam-se Keller (2000), Manning (2003) e Bod *et al.* (2003).

³ Cf. Bod *et al.* (2003) e Manning e Schütze (1999), entre outros.

Dado o papel de relevo que a semântica lexical computacional tem vindo a ganhar no desenvolvimento de recursos linguísticos, a investigação recente tem incidido proeminentemente sobre o tratamento da ambiguidade lexical.

As abordagens estatísticas têm desenvolvido estratégias de desambiguação que envolvem aprendizagem automática de preferências lexicais e estruturais a partir de *corpora*, sendo bem sucedidas fundamentalmente no que respeita à homonímia, ou polissemia contrastiva.

Sem eludir os avanços que os métodos quantitativos têm permitido neste domínio, há, contudo, que apontar as dificuldades que se põem ao tratamento da polissemia lógica, em que se revela inoperante o pressuposto simplista, quase sempre assumido, de que a polissemia respeita a formas portadoras de um conjunto finito de sentidos discretos, sendo que, como é sabido, a polissemia está longe de poder ser tratada como um fenómeno homogéneo, em particular porque os sentidos nem sempre são dissociáveis de forma inequívoca⁴.

Os casos de ambiguidade que aqui se discutem são de natureza sintáctica e envolvem fortes restrições de ordem semântica. Vejamos um exemplo:

- (1) a. O João tomou o champanhe gelado.
b. Foi o champanhe gelado que o João tomou.
c. Foi gelado que o João tomou o champanhe.

A ambiguidade de (1)a. resulta do facto de *gelado* poder ser interpretado como modificador de *o champanhe*, integrando, por conseguinte, o sintagma nominal objecto directo, como se dá conta em (1)b., ou como predicado secundário⁵ de *o champanhe*, sendo, neste caso, excluído do objecto directo, como se evidencia em (1)c. Enquanto predicado secundário, *gelado* descreve um estado de *o champanhe*, que se verifica no intervalo de tempo definido pelo predicado primário. O sintagma *o champanhe* constitui-se, assim, simultaneamente como objecto directo do predicado primário (*tornou*)⁶ e como sujeito de *gelado*, predicado secundário.

Compare-se agora (1) com (2), abaixo:

- (2) a. O João tomou o champanhe francês.
b. Foi o champanhe francês que o João tomou.
c. *Foi francês que o João tomou o champanhe.

⁴ Sobre esta matéria ver, entre outros: Ravin e Leacock (2000), para um confronto de abordagens; Marrafa (2002), sobre as dificuldades de decisão para evitar sobrediferenciação e subdiferenciação de sentidos na construção da WordNet.PT; e Kruengkrai *et al.* (2004), para uma recente e relativamente bem sucedida abordagem, que, contudo, deixa para o futuro casos de aglutinação de sentidos como em *o banco que me emprestou o dinheiro ardeu*.

⁵ Por predicado secundário entende-se um predicado cujo sujeito é também marcado tematicamente por um predicado primário e que, em consequência, não é de ocorrência obrigatória, visto não ter qualquer impacto na gramaticalidade da frase.

⁶ O termo *predicado* é aqui utilizado por abreviação de predicado lexical, no sentido de Marrafa (1993), que distingue entre predicado lexical — um núcleo lexical que define uma estrutura eventiva (à *la* Pustejovsky, 1991) — e predicado sintáctico — a projecção máxima do predicado lexical.

Como a agramaticalidade de (2)c. demonstra, a interpretação correspondente à predicação secundária não está disponível. Ou seja, contrariamente a (1)a., (2)a. não é uma frase ambígua.

O contraste resulta, naturalmente, das propriedades semânticas das expressões adjectivais envolvidas. Em concreto, e usando a dicotomia de Carlson (1977), *gelado* denota uma propriedade accidental, ou transitória, enquanto *francês* denota uma propriedade individual.

Os dados evidenciam, pois, que a predicação secundária só é compatível com a expressão de propriedades accidentais, restrição que se verifica também quando o predicado secundário tem como sujeito não o objecto directo, mas o sujeito da predicação primária, como se ilustra:

- (3) a. O João saiu contente.
- b. *O João saiu alto.

O analisador terá, evidentemente, de ser dotado de conhecimento que lhe permita reconhecer esta restrição. De contrário, sobregerará representações.

Como terá de “saber” que um mesmo adjectivo poderá ser predicado secundário com determinados nomes, mas não com determinados outros. Comparem-se os seguintes dados:

- (4) a. O João leu o livro triste.
 - b. Foi tristei que o Joãoi leu o livro.
 - c. Foi o livro triste que o João leu.
 - d. *Foi o livroi que o João leu tristei.
- (5) a. O João encontrou a amiga triste.
 - b. Foi tristei que o Joãoi encontrou a amiga.
 - c. Foi a amiga triste que o João encontrou.
 - d. Foi a amigai que o João encontrou tristei.

Tanto (4)a. como (5)a. são frases ambíguas, dado que *triste* pode ser interpretado como predicado ou como modificador, como a posição de contraste nos restantes exemplos torna claro. Contudo, (5)a. é mais ambígua do que (4)a. — passe a sugestiva expressão Orwellana. Ou seja, em (5)a. *triste* pode ser predicado secundário do sujeito, modificador integrado no objecto directo e predicado secundário do objecto directo, enquanto em (4)a. esta última interpretação é excluída (cf. o contraste de gramaticalidade entre (4)d. e (5)d.). Tal diferença decorre do facto de *triste* ser um item ambíguo entre propriedade accidental e propriedade individual (cf. *O João está triste. vs O João é triste*). A ocorrência como modificador de nomes que denotam entidades desprovidas de emoções é possível para interpretações causativas (no caso, a expressão *livro triste* é interpretada como equivalente a *livro que causa tristeza*). Qualquer outra interpretação é, naturalmente, excluída.

Mas a questão é ainda mais complexa. Não só as propriedades não são individuais, ou accidentais, em absoluto, como a restrição apontada acima parece não ter de se verificar em todos os contextos. Vejam-se os exemplos seguintes:

- (6) a. O João chegou magro.
- b. *O João falou magro.

(7) a. O João comprou a bicicleta amarela (não a comprou azul).

b. *O João lavou a bicicleta amarela (não a lavou azul).⁷

Como se verifica, a predicação secundária é restringida também pela natureza semântica do predicado primário. Como evidenciado em Marrafa (1993)⁸, tal restrição justifica-se pelo facto de a predicação secundária ser “parasita” da predicação primária, no sentido em que o predicado secundário expressa um estado que afecta um dos participantes da predicação primária e se verifica em todos os subeventos que integram a estrutura léxico-conceitual definida pelo predicado primário⁹. Noutros termos, o predicado secundário e o predicado primário atribuem propriedades à mesma entidade para o mesmo intervalo de tempo, como abaixo se esquematiza (ver figura 1) para (8), um exemplo simples, em que o evento denotado pelo verbo é um processo¹⁰ que envolve um único participante.

(8) O João dançou descalço.

$$\left[\begin{array}{c} J \\ e_1(x)^{pk} \dots e_i(x)^{pk} \dots (e_n(x)^{pk})_m \\ \\ e(x)^{pm} \end{array} \right]$$

Figura 1 — Relação evento secundário – evento primário.

O evento global realiza-se no intervalo de tempo J e é aqui representado pelos subeventos e_1, e_i, e_n , que se realizam nos correspondentes subintervalos t_1, t_i, t_n e envolvem o participante x . O predicado secundário é representado pelo evento atómico e , que envolve igualmente x como participante. Admitindo que um estado se caracteriza por um número n de propriedades, consideramos aqui, por simplificação, $n=1$. Assim, a cada um dos estados correspondentes a e_1, e_i, e_n está associada a propriedade pk . Ao estado denotado pelo predicado secundário, representado por e , associa-se a propriedade pm , que é predicada à mesma

⁷ Não está em causa a boa-formação da frase para a interpretação em que *amarela* é modificador.

⁸ No que respeita à questão linguística que equaciona, este artigo é inspirado em grande medida em Marrafa (1993), trabalho em que o tema da predicação secundária é tratado aprofundadamente.

⁹ Para uma definição formal cf. Marrafa (op. cit.: 184).

¹⁰ No que respeita à classificação da estrutura dos eventos, segue-se, no essencial, a proposta de Pustejovsky (1991) e trabalhos subsequentes. A saber: (i) estado: evento atómico, não ramificado, não avaliado em relação a qualquer outro (e.g. *saber, amar*); (ii) processo: estrutura de ramificação múltipla, constituída por uma sequência de eventos idênticos, (e.g. *correr, cantar*); transição: estrutura binária, que envolve um evento avaliado relativamente a outro (e.g. *construir, rasgar*).

entidade, no mesmo intervalo de tempo, J , e logo em $e1$, ei , en . Ou seja, pm tem de ser compatível não apenas com x , mas igualmente com pk .

Torna-se, por conseguinte, evidente que a adequação de qualquer sistema de análise automática para as construções em causa depende da inclusão de especificações que captem não apenas as propriedades dos possíveis predicados secundários, mas também as propriedades dos subeventos e dos participantes que os “acolhem”.

3. Modelização

Como se evidenciou na secção anterior, um sistema de análise automática que descreva adequadamente o tipo de expressões em causa envolve a codificação de restrições que implicam uma caracterização fina da semântica dos itens lexicais. Por outro lado, tem de incluir especificações sintácticas que permitam gerar a representação adequada a cada interpretação.

Nesta secção descreve-se, em traços gerais, o funcionamento de um sistema que segue, no que respeita às linhas básicas, o proposto em Marrafa (1993), e que integra:

- (i) dois módulos relativos à descrição linguística: um Léxico, que contém as especificações lexicais, e uma Gramática que define as configurações sintácticas e especifica os predicados de controlo, utilizados para garantir a observância de restrições especificadas no Léxico, bem como de restrições sintácticas gerais;
- (ii) dois instrumentos relativos ao tratamento automático: um *parser*, com uma estratégia de resolução ascendente, implementado directamente em PROLOG, e um mecanismo de unificação de traços, que opera como um filtro que impede a formação de constituintes que envolvam informação conflituosa.

Face aos objectivos deste artigo, a descrição centra-se nos módulos relativos à descrição linguística, limitando-se a referência aos instrumentos computacionais propriamente ditos ao estritamente necessário.

3.1. Léxico

Como atrás se deixou claro, as especificações lexicais assumem um papel de grande proeminência. É no Léxico que se codifica toda a informação relativa aos contextos sintácticos e semânticos em que os diversos itens podem ocorrer.

A despeito da complexidade que, como se demonstrou, a determinação da informação relevante envolve, a codificação em si é relativamente simples. As propriedades são representadas sob a forma de estruturas de traços, sendo estas constituídas por traços —, pares *atributo:valor*, mais precisamente. Vejam-se, a título ilustrativo, dois exemplos muito simplificados¹¹:

(9) det (... (o), [num: sing, gen: masc, ...]) --> [o].

¹¹ Deixam-se de lado todos os aspectos formais não relevantes.

(10) n (... (livro), [num: sing, gen: masc, sem: nhum, ...])
 --> [livro].

Nestas entradas lexicais abreviadas especifica-se informação relativa a número (num) e género (gen) em ambas, sendo que (10) inclui ainda um traço semântico (sem: nhum).

Embora (10) inclua informação que não consta em (9), tal não impede a constituição do sintagma *o livro*, dado que os atributos comuns têm os mesmos valores.

Desta forma, é possível proceder à necessária caracterização sintáctica e semântica dos itens lexicais e concertar a informação expressa para um dado item com a dos restantes.

Cabe notar que a dificuldade em encontrar soluções com vista à adequação descritiva do sistema não depende em nada das estruturas de modelização utilizadas, dado que os traços constituem um conjunto aberto e são arbitrários quanto à natureza, ao número e à ordem por que são especificados.

3.2. Gramática

O formalismo adoptado para definição das configurações sintácticas é basicamente uma Gramática de Cláusulas Definidas, com pequenas alterações sintácticas que permitem a sua interpretação por um analisador ascendente¹². Em (11) e (12) apresentam-se os formatos das regras e dos símbolos gramaticais, respectivamente, e em (13) e (14) os predicados de controlo:

(11) símbolo gramatical --->
 [qualquer sequência de símbolos gramaticais e de predicados de controlo].

(12) <nome do símbolo> (<árvore sintáctica>, <estrutura de traços>).

(13) extrair(T,A,V).

(14) unificar(A,B,C).

Os predicados de controlo fazem apelo às estruturas de traços especificadas nas entradas lexicais. Assim, *extrair* garante que uma estrutura de traços *T* contém o atributo *A*, com o valor *V*. Por seu turno, *unificar* estabelece que *A* e *B* unificam em *C*, sendo que a operação de unificação — definida no mecanismo de unificação — obriga a que atributos comuns assumam os mesmos valores. Desta forma, assegura-se que os diversos constituintes não incluem subconstituintes sintáctica ou semanticamente incompatíveis. Vejam-se as seguintes regras, meramente ilustrativas¹³:

(15) sn(sn(Det,N),T) --->
 [det(Det,T1),
 n(N,T2),
 unificar(T1,T2,T)].

¹² A estratégia de resolução ascendente evita problemas de recursividade à esquerda. Na gramática das construções em questão tais problemas surgiriam em consequência das várias estruturas de adjunção que a sua sintaxe envolve.

¹³ Por simplificação, utilizam-se apenas projecções máximas e respectivos núcleos.

(16) $sv(sv(V, SN), T3) \dashrightarrow$
 $[v(V, T3),$
 $sn(SN, T4), \text{extrair}(T3, \text{scat}, sn)]$.

Como se pode observar, o lado esquerdo das regras é constituído pelo símbolo gramatical — *sn* num caso, *sv* no outro —, pela respectiva árvore sintáctica — $sn(Det, N)$, em (15), $sv(V, SN)$, em (16) — e pela estrutura de traços que corresponde a cada símbolo — T , no primeiro caso, $T3$ no segundo. O lado direito das regras integra os símbolos gramaticais dos subconstituintes com as correspondentes árvores sintáticas e estruturas de traços. (15) inclui ainda o predicado *unificar* e (16) o predicado *extrair*.

No que respeita ao funcionamento dos predicados, *unificar*, em (15), obriga a que a estrutura de traços do determinante, representada por $T1$, e a estrutura de traços do nome, representada por $T2$, sejam compatíveis, excluindo, assim, que expressões como *as livros* ou *a cadeira desesperada* sejam reconhecidas como sintagmas nominais sintáctica e semanticamente legitimados. Quanto a *extrair*, em (16), estabelece como condição necessária para que a regra seja satisfeita que a estrutura de traços do verbo, representada por $T3$ ¹⁴, contenha o par *scat: sn*, ou seja, que o verbo subcategorize para *SN*. Desta forma, uma expressão como *telefonei esta manhã*, por exemplo, não será analisada como um *SV*, dado que a estrutura de traços associada ao verbo não incluirá tal informação.

Importa sublinhar que o sucesso da análise depende crucialmente da informação codificada nas entradas lexicais.

4. Conclusão

Qualquer que seja a abordagem que se adopte no tratamento automático de ambiguidades como as discutidas neste artigo terá de se confrontar com a necessidade de identificação de um conjunto de intrincadas restrições de difícil, se não impossível, predição puramente estatística.

As evidências aduzidas são, pois, favoráveis à modelização baseada em conhecimento, como a que se apresenta. E não se proporciona, por ora, o contraste com propostas probabilísticas, dado que os fenómenos em causa não têm estado no centro dos interesses da investigação nesse âmbito.

Referências bibliográficas

- Bod R., Hay J., Jannedy S. (eds.) (2003). *Probabilistic Linguistics*. Cambridge: The MIT Press.
- Carlson G. (1977). *Reference to Kinds in English*. Dissertação de PhD. Universidade de Massachusetts.
- Horning J. (1969). *A study of grammatical inference*. Dissertação de PhD. Universidade de Stanford.
- Keller F. (2000). *Gradience in grammar*. Dissertação de PhD. Universidade de Edimburgo.

¹⁴ Faz-se notar que ao sintagma verbal corresponde a mesma estrutura de traços, pelo facto de o sintagma verbal só herdar informação do seu núcleo.

- Kruengkrai C., Charoenporn T., Sornlertlamvanich V., Isahara H. (2004). Enriching a Thai lexical database with selectional preferences. *Proceedings of the IV International Conference on Language Resources and Evaluation*. Vol. VI, 2201-2204.
- Manning C. D., Schütze H. (1999). *Foundations of statistical natural language processing*. The MIT Press.
- Manning C. (2003). Probabilistic syntax. In Bod R., Hay J., Jannedey S. (eds.), *Probabilistic linguistics*. Cambridge: The MIT Press. 289-341.
- Marrafa P. (1993). Predicação secundária e predicados complexos em português: Análise e modelização. Dissertação de Doutoramento. Universidade de Lisboa.
- Marrafa P. (2002). Portuguese WordNet: General architecture and internal semantic relations. *DELTA*. Vol. 18, número especial, 131-146.
- Pustejovsky (1991). The syntax of event structure. *Cognition*. 41, 47-81.
- Ravin Y., Leacock C. (eds.) (2000). *Polysemy: Theoretical and computational approaches*. Oxford University Press.
- Saffran J., Newport E., Aslin R. (1996). Statistical learning by 8-month-old infants. *Science*. 274, 1926-1928.

Autora



Palmira Marrafa é Professora na Faculdade de Letras da Universidade de Lisboa, Portugal, onde se doutorou em Linguística (com uma tese em Linguística Computacional) e onde lecciona Linguística Geral e Linguística Computacional. Coordenou, pela Faculdade de Letras, a Licenciatura em Engenharia da Linguagem e do Conhecimento, de 1994 a 1999. De 1999 a 2003 coordenou o Mestrado na área de especialização em Processamento e Tecnologias das Línguas Naturais e coordena o actual

Mestrado na área de especialização em Linguística Computacional, na mesma Faculdade. Criou e coordena o CLG-Grupo de Computação do Conhecimento Léxico-Gramatical, do Centro de Linguística da Universidade de Lisboa, que desenvolve vários projectos na área da modelização e das tecnologias das línguas, de que se destaca a WordNet do Português (WordNet.PT), no quadro da EuroWordNet. Tem várias colaborações internacionais, em particular com universidades no Brasil, em Espanha, em França e nos Estados Unidos da América. Foi membro do *board* da *Global WordNet Association* de 2000 a 2004.