

# **OBSERVAÇÃO E GENERALIZAÇÃO: ANÁLISE LINGUÍSTICA DE VERBOS DECLARATIVOS ITALIANOS COM BASE EM CORPORA LINGUÍSTICOS\***

**Nicoletta Calzolari**  
Istituto di Linguistica Computazionale del CNR, Pisa

## **1. Enquadramento Teórico: Lexicografia com Base em Corpora Linguísticos e Semântica Estrutural**

Apresento aqui algumas observações sugeridas pelo trabalho feito sobre a língua italiana, no quadro do projecto DELIS, financiado pelo LRE da UE, sobre «métodos e ferramentas para lexicografia com base em corpora linguísticos.»

Duas hipóteses – uma, teórica, a outra, metodológica – subjazem às presentes observações, e ao trabalho no DELIS, em geral: i) uma abordagem lexicográfica com base em corpora linguísticos, e ii) uma análise semântica baseada num conjunto de estruturas.

i) Assumimos que só uma análise detalhada e cuidadosa de dados provenientes de corpora pode constituir uma base sólida para uma abordagem realista à construção do léxico, seja este representado por um dicionário tradicional ou por um léxico computacional.

ii) Os dados provenientes de corpora não podem, obviamente, ser utilizados de modo simplista. Para se poderem tornar úteis, os dados devem ser analisados à luz de uma dada hipótese teórica, na base da qual se modelará e estruturará o que seria, de outro modo, um conjunto não estruturado de dados. A melhor combinação da abordagem empírica com a abordagem teórica é aquela em que a própria hipótese teórica

surge de e é guiada por sucessivas análises dos dados, e onde, por sua vez, os dados são ciclicamente aperfeiçoados e ajustados à evidência textual.

No âmbito do projecto DELIS utilizámos, como hipótese teórica modeladora, a abordagem da semântica estrutural, definida por Fillmore (cf., por ex., Fillmore e Atkins 1992), e isto, tanto na análise de dados textuais, como no subsequente desenho da entrada lexical. Uma das estratégias descritivas essenciais desta teoria consiste em juntar os níveis descritivos sintáctico e semântico. Trata-se de uma abordagem à análise de corpora e à construção de léxicos que advoga a reutilização dos seus resultados no contexto de aplicações de Processamento de Linguagem Natural (PLN); além disso, no que diz respeito às decisões a tomar quanto à metodologia na análise de corpora, é uma abordagem que permite ao lexicógrafo/linguista i) começar por uma caracterização semântica intuitiva dos chamados «elementos estruturais» semânticos, e só depois procurar a sua realização sintáctica ou, vice versa ii) começar com uma descrição das principais descrições sintácticas e só depois dar a sua interpretação semântica em termos de elementos estruturais.

Ambas as abordagens metodológicas são complementares e podem ser escolhidas de acordo com características particulares da classe semântica das palavras a analisar pelo lexicógrafo (cf. Alonge et al. 1993).

Neste artigo não nos deteremos na relevância de diferentes metodologias de análise de corpora segundo diferentes tipos de classes semânticas de palavras, preferindo fornecer alguns exemplos de resultados linguísticos de uma análise lexical baseada em corpora. Ao optar por esta via, prestar-se-á especial atenção às correlações entre diferentes níveis de descrição linguística. O projecto DELIS concentra-se na correlação entre aspectos semânticos e sintácticos, embora seja evidente que também outros aspectos linguísticos – tais como a morfologia, a morfosintaxe, a co-ocorrência lexical e os dados combinatórios – estão estreitamente correlacionados, devendo essas relações ser captadas no desenho da entrada lexical e, principalmente, na descrição do fenómeno da discriminação de significados. É precisamente a complexidade das interrelações entre todos estes aspectos que torna a desambiguação semântica numa tarefa de execução tão difícil no PLN. Um dos nossos

objectivos é a utilização de corpora textuais como meio de descobrir e revelar o intrincado destas relações; pretendemos, também, utilizar a semântica estrutural para decifrar e desenredar esta complexa situação, transformando-a em peças elementares e de fácil tratamento informático.

## **2. Análise de classes semânticas de verbos com base em corpora.**

O primeiro passo dado no projecto DELIS foi coligir, de forma sistemática, e a partir de linguagem real, tal como aquela encontrada em corpora textuais, evidência de regularidades (ou idiossincrasias) no comportamento de algumas classes verbais a diferentes níveis linguísticos (morfosintáctico, sintáctico, semântico, e suas interrelações).

Os resultados desta análise sistemática, e as generalizações possibilitadas pelos dados no quadro das regras teóricas adoptadas como hipótese de trabalho, serviram como ponto de partida para a fase seguinte de desenho e preenchimento das correspondentes entradas lexicais (cf. Monachini et al. 1994).

Utilizou-se o seguinte processo – aqui descrito de forma muito esquemática:

Projectou-se o denominado «Esquema de Codificação DELIS», um guia detalhado para a anotação de corpora (no que diz respeito a informação de tipo morfosintáctico, sintáctico, semântico, combinatório, etc), com exemplos de cada uma das línguas do projecto (francês, inglês, holandês, alemão, italiano), e com descrições das suas peculiaridades quando confrontadas com as recomendações gerais para cada língua, de forma a tomar em consideração requisitos particulares a cada uma delas, especificamente, embora mantendo um esquema de anotação comum para todo o projecto.

O corpus foi automática ou/e manualmente anotado usando-se o Esquema de Codificação DELIS. Os resultados das análises linguísticas do corpus anotado sublinharam, em especial, a importância, para a descrição lexicográfica da entrada lexical, de muitos tipos de correlação entre níveis distintos de análise linguística (morfologia, sintaxe, semântica), e forneceram, também, a base para a subsequente fase de generalização. Foram também avaliados problemas na anotação.

Apontaram-se sistematicamente tipos distintos de regularidades básicas nos dados coligidos que seriam, depois, considerados, mais

precisamente aquando da modelação das entradas lexicais relevantes (cf. Monachini et al., 1994).

Analísaram-se alguns Verbos Declarativos e Perceptivos. Seleccionaram-se os verbos mais gerais de ambos os campos, de modo a conseguir alcançar-se uma representação dos próprios campos. Outros verbos mais específicos foram escolhidos por apresentarem características diferentes dos gerais. Os mesmos verbos foram seleccionados em cada língua, possibilitando assim uma comparação linguística entre as várias línguas.

Produziram-se listas das formas flexionadas para os verbos seleccionados que ocorriam no Corpus (para o Corpus Italiano, cf. Bindi et al. 1991), juntamente com a indicação da sua frequência de ocorrência (esta última constitui tão-só uma sequência de caracteres, pelo que, no caso das palavras homógrafas, este número pode reportar-se a palavras pertencentes a categorias diferentes, por ex., 'promessa' pode significar, quer «prometido» – Particípio Passado Singular do Verbo –, quer «promessa» – Nome Feminino Singular). As formas homógrafas foram marcadas e desambiguadas.

Na análise propriamente dita, apenas se consideraram as ocorrências verbais, tendo sido abandonadas aquelas que participavam de grupos de palavras (mais bem consideradas no léxico como uma só Unidade Lexical) – uma decisão que foi generalizada a todos os grupos DELIS. Os grupos de palavras encontrados no curso desta análise foram simplesmente listados, sem se ter tentado integrá-los em diferentes tipos.

A primeira – e óbvia – observação feita, foi a de que ambas as classes semânticas seleccionadas tinham comportamentos sintácticos e semânticos muito distintos, daí resultando que uma abordagem algo diferente na sua análise lexicográfica e na anotação do corpus seria, talvez, mais apropriada aos dois campos considerados.

Seguidamente delinear-se-ão alguns exemplos do tipo de resultados conseguidos com esta análise, sublinhando-se uma classe semântica em particular – a dos verbos Declarativos.

### **3. Algumas observações sobre verbos Declarativos.**

A partir da análise das ocorrências de todas as formas flexionadas de verbos declarativos, conseguimos captar:

1) Dados por vezes – ou frequentemente – negligenciados pela lexicografia tradicional, por ex., i) significados do verbo estreitamente ligados a contextos particulares ou a flexões morfológicas específicas, ii) um grande número de expressões idiomáticas e combinatórias, iii) todos os padrões sintácticos aceites por um verbo, e as suas respectivas frequências, iv) preferências dadas por um verbo a determinadas construções sintácticas, itens lexicais, etc., v) evidência de pistas sintácticas, semânticas e morfológicas que parecem ser cruciais no que diz respeito à desambiguação (semi-)automática do significado e na caracterização não-ambígua de um dado uso do verbo.

2) Dados passíveis de serem tomados em consideração para futura automatização da extracção de conhecimentos (por ex., padrões sintácticos, argumentos verbais, etc.) a partir de corpora textuais.

De facto, a análise detalhada de ocorrências provenientes do corpus destinava-se a sublinhar as possíveis configurações sintácticas permitidas pelo verbo, bem como a sua correlação – quando existente – com os significados do verbo e com os elementos estruturais. A anotação e extracção destes dados foi feita manualmente no projecto DELIS, embora com isso se pretendesse, também, coligir evidências úteis para a sua futura automatização.

Na análise de, por ex., *dire* (dizer) e *promettere* (prometer) sobressaíram imediatamente dois factos:

1) estes verbos não são globalmente complexos, de um ponto de vista semântico, isto é, não apresentaram um grau de polissemia muito elevado (sem considerar a sua presença em grupos de palavras ou expressões idiomáticas, tratados como entradas lexicais separadas),

2) tais verbos ocorreram, no entanto, numa grande variedade de padrões subcategorizadores, embora apenas alguns desses padrões (em número muito restrito) apresentassem univocamente uma diferença de significado. A maioria dos padrões pode ser considerada como manifestações sintácticas possíveis e diferentes do mesmo significado básico, talvez dentro de um conjunto parcialmente diferente de Elementos Estruturais circundantes.

O tipo básico de informação que podemos extrair do Corpus, no que diz respeito a esta classe de verbos, é, portanto, de natureza sintáctica, em especial informação relacionada com a estrutura oracional,

com as relações gramaticais e com a regência. Este tipo de informação sintáctica foi extraído e anotado manualmente, embora um tal tipo de anotação se possa fazer – pelo menos parcialmente – de modo automático.

Como já foi referido, o principal tipo de variação com que nos confrontámos na análise de ocorrências do corpus de verbos declarativos relacionava-se com o seu enquadramento sintáctico. Pareceu-nos, portanto, apropriado, anotar as ocorrências do corpus de acordo com tipos distintos de enquadramento sintáctico. Como resultado desta anotação sintáctica surgiu uma lista de padrões frásicos sintácticos, lista essa que continha todos os padrões que, de facto, tinham ocorrido no subconjunto analisado do Corpus. Esta lista não pretende ser exaustiva, e será actualizada à medida que forem analisadas outras ocorrências verbais. Está a ser usada como base para uma anotação sintáctica e uniforme, e – o que é mais importante – pode ser utilizada por uma ferramenta que se destine à procura, no Corpus, de padrões sintácticos de superfície relevantes.

Um campo que contenha uma etiqueta correspondente a um elemento desta lista é, portanto, uma característica importante do Esquema de Anotação DELIS em geral. Por exemplo, para formar a palavra *disse* (disse) no contexto:

CE (Contexto à Esquerda) – *Il contadino mi*  
(O camponês ...-me)

PC (Palavra-Chave) – *disse*  
(disse)

CD (Contexto à Direita) – *di aver fatto la guerra del 1716...*  
(ter lutado na guerra de 1716)

poderíamos associar um código de Padrão Sintáctico, tal como:

«OI+diInf(rs)» (obj. ind. + diInf. (regência/controlo do Suj.)).

Este valor é seleccionado pelo anotador a partir de uma lista pré-definida, podendo, também, ser conferido por um programa, após análise frásica.

#### 4. Correlação entre significado e morfologia.

Num número apreciável de formas flexionadas, e por vezes em ligação com determinados padrões sintácticos, o verbo *dire* adquire um significado muito específico, que não se encontra em outras flexões. Num contexto multilinguístico, estas formas flexionadas requerem traduções muito específicas que, além do mais, nem sempre incluem um verbo declarativo na língua de chegada.

É o que se passa com a forma *diresti* (2ª pess. sing. Pres. Condicional) (tu dirias), que se encontra, em muitas ocorrências, nas duas seguintes construções:

*Cosa\che ne diresti + diInf. (regência Suj.)?*

*Cosa\che ne diresti + PP(di)?*

com o significado de «ti piacerebbe?» (gostarias?); por exemplo em:

*Cosa ne diresti di andare a mangiare qualcosa?*

(lit.<sup>1</sup>: O que dirias a ir comer qualquer coisa?)

*Che ne diresti di una fetta de focaccia?*

(lit.: O que dirias a uma fatia de pizza?)

O mesmo significado deveria, também, encontrar-se na 3ª pess. sing do Pres. Condicional, *direbbe* (ele diria), quando este é utilizado como forma de cortesia da 3ª pessoa, o que não ocorreu no nosso Corpus, embora seja desejável utilizar um corpus de fala para confirmar este princípio.

A forma *dici* (2ª pess. sing. do Pres. do Ind.), quando ocorrendo no mesmo padrão sintáctico, tem um significado totalmente diferente, mais especificamente, «*cosa ne pensi*» (o que pensas disso?):

*Che ne dici di questa storia?*

(lit.: O que dizes desta história?)

Outro exemplo de significado idiossincrático, ligado a uma flexão particular, é dado por *direi* (1ª pess. sing. do Pres. Condicional)

(eu diria), que, em muitas ocorrências, é usado numa construção parentética, normalmente, mas nem sempre, antes de um adjectivo. Nestes casos, a palavra comporta-se, tanto a nível sintáctico como a nível semântico, como um advérbio, adquirindo o significado da locução adverbial «*per così dire*» (por assim dizer).

Seguem-se alguns exemplos:

*lo sguardo era sincero, direi amichevole*  
(lit.: o seu olhar era sincero, diria amigável)

*in modo direi scorretto*  
(lit.: de maneira, diria, incorrecta)

*non lo fa trascinare, ma direi arrancare*  
(lit.: não o fazia arrastar-se, mas diria andar penosamente)

*tre o quattro veicoli, direi camion o autobus*  
(lit.: três ou quatro veículos, diria camiões ou autocarros)

Esta mesma construção regista-se, também, na 1ª pess. do pl. do Fut. do Ind. *diremo* (diremos), como em:

*l'iniziativa diremo reaganiana di Bush*  
(lit.: a iniciativa, diremos reaganiana, de Bush)

O facto – frequentemente registado – de certos comportamentos semânticos ou/e sintácticos estarem ligados a um número muito restrito de palavras conduz à necessidade, num léxico computacional, de permitir o acesso às componentes Sintácticas e Semânticas para além do nível do Lema, às próprias formas flexionadas, ou, melhor ainda, à componente Morfológica do léxico.

## 5. Propriedades sintácticas e correlação entre Semântica e Sintaxe

A primeira observação registada a nível da sintaxe foi a de que nem todos os padrões sintácticos possíveis ocorriam, de facto, com todas as formas flexionadas. O leque de variação verificou-se ser muito



lato – tal como se esperava – devido, em parte, à diferença nas frequências das formas. As formas mais frequentes apresentavam, como é evidente, um maior número de padrões. Por exemplo, a forma *dire* (Infinitivo), a mais frequente, apresenta quase todos os padrões teoricamente aceitáveis. A partir da análise de outras formas verificou-se que a discrepância entre o que é teoricamente aceitável pela introspecção (isto é, de forma geral, o mesmo conjunto de padrões que ocorrem com *dire*) e o que é, de facto, encontrado no uso se torna bastante importante. Pensamos que este tipo de informação, que não pertence ao tipo absoluto (sim/não), mas que é antes de natureza preferencial, ou seja, em que se encara a frequência do uso como um contínuo, é de utilidade prática num Léxico Computacional, e deveria ser registada com todas as formas flexionadas distintas.

Este tipo de anotação (que deveria, em princípio, ser registada para diferentes tipos de texto, já que também as diferenças a nível de tipos de texto podem introduzir distinções substanciais a este nível) apenas se tornará possível, a larga escala, quando se encontrarem disponíveis ferramentas automáticas de anotação sintáctica de Corpora, o que facilitará a extracção da frequência relevante dos dados.

Apesar de existir uma grande variação entre padrões sintácticos potencialmente possíveis e padrões sintácticos utilizados na realidade, não observámos muita variação, a nível das palavras, no que diz respeito a agrupamentos de elementos estruturais, ou seja, variações nos padrões semânticos, bem como no tipo de elementos estruturais associados com as formas flexionadas do verbo *dire*.

É particularmente importante, para o desenvolvimento de futuras ferramentas de anotação automática que avancem para além da sintaxe de superfície, i) estabelecer as correlações relevantes entre os diferentes elementos sintácticos de superfície e os elementos estruturais a nível semântico, e ii) observar a regularidade destas correlações.

Mantêm-se sempre as seguintes correlações:

- o Sujeito corresponde ao Emissor
- o Objecto Indirecto corresponde ao Receptor
- o Objecto Directo corresponde à Mensagem (de diferentes tipos)
- di-Infinitivo corresponde à Mensagem
- che-oracção corresponde à Mensagem
- qu-oracção corresponde à Mensagem

- citação directa corresponde à Mensagem
- PP(su) corresponde ao Tema
- PP(di) corresponde ao Tema

Apesar destas regularidades, a conhecida ausência de uma cartografia unívoca entre a sintaxe de superfície e o significado ainda se mantém, o que constitui um importante obstáculo à exequibilidade do reconhecimento e extracção automáticos de padrões sintácticos e argumentos verbais semânticos de forma não-ambígua. Observamos, por exemplo, que os dois principais significados do verbo italiano *dire*, ou seja, i) «dizer alguma coisa, de forma a torná-la conhecida», e ii) «dizer qualquer coisa de forma a pedir que qualquer coisa seja feita», são dados exactamente através do mesmo tipo de superfície de padrão sintáctico. Os dois significados podem ser exemplificados com as seguintes frases:

*Mark é sincero quando dice a Rachel di amarla.*  
(lit.: Mark é sincero quando diz a Raquel de amá-la)

*Lei dice a me di non alzare la voce?*  
(lit.: você diz-me de não levantar a voz?)

O di-Infinitivo pode ser interpretado como um complemento Mensagem-tipo em ambos os casos, mas a força de elocução é diferente, ‘declarativa’ na primeira frase e ‘imperativa’ na segunda.

A primeira mensagem-tipo está, de facto, em variação com uma che-oracção, ou seja, podíamos igualmente ter:

*Mark é sincero quando dice a Rachel che la ama.*  
Mark é sincero quando diz a Rachel que a ama.

Notamos, obviamente, que o sujeito do Infinitivo na primeira frase é o sujeito da oracção principal (regência de sujeito), enquanto que na segunda frase ele passa a ser o objecto indirecto (regência de objecto indirecto). Esta informação deve, portanto, ser anotada num corpus, se queremos distinguir entre dois tipos de significado principais, apesar de não vermos maneira de conseguir isto automaticamente através do uso de um analisador puramente sintáctico.

## 6. Evidência proveniente de corpora vs. Introspecção como base para a descrição lexicográfica: real vs. potencial

Um dos aspectos mais interessantes no uso de um corpus para levar a cabo uma tarefa lexicográfica é o imediato confronto com a impossibilidade, baseada na evidência, de utilizar qualquer tipo de descrição que repouse numa fronteira estanque entre o que é admissível e o que o não é. É evidente que, no uso real da língua, um grande número de propriedades aparece como um contínuo, e não como propriedades do tipo sim/não. De facto, esta é uma das características principais encontradas no uso real da língua. O mesmo se verifica no caso das chamadas «regras», em que encontramos, nos exemplos extraídos de corpora, mais uma «tendência» para a regra, do que uma regra precisa. A maioria dos tipos de informação não devem ser tratados como correspondendo a uma restrição absoluta, cuja violação gerará uma frase totalmente inaceitável, mas antes como preferências que tornam uma dada frase mais ou menos aceitável num dado contexto, sem afectar a sua gramaticalidade.

Isto coloca um problema a nível da representação linguística, onde este tipo de informação preferencial tem de ser adaptada, o que pode não ser fácil e que não é, certamente, simples, para um formalismo baseado em restrições. Infelizmente, apesar da existência de algumas propostas nessa direcção, os formalismos baseados na unificação (ou em restrições), tal como hoje os conhecemos, não captam facilmente esta distinção, por outras palavras, ou se ignoram as preferências, ou estas são tratadas como restrições absolutas, ou, ainda, se utilizam mecanismos ad-hoc para lidar com elas.

Além disso, a evidência do uso real contrasta frequentemente com o que se esperaria encontrar se o julgamento fosse apenas baseado na introspecção. Um claro exemplo desta situação é dado pelo comportamento de dois verbos declarativos italianos, *chiedere* e *domandare*, ambos possíveis traduções do verbo inglês «to ask», e que introduzem, contudo, um problema de selecção lexical no que diz respeito à tradução.

Estes verbos apresentam fortes semelhanças. Intuitivamente, os falantes nativos consideram-nos sinónimos, exibindo comportamentos semelhantes, mesmo a nível da sua polissemia. São, também, descritos de modo similar em dicionários (onde um é frequentemente utilizado

para explicar e definir o outro) e gramáticas tradicionais. No entanto, surpreendentemente, comportam-se de modo muito distinto no Corpus, oferecendo resultados inesperados.

Enquanto que, em termos teóricos, ambos admitem exactamente o mesmo tipo de padrões sintácticos de superfície e nuances semânticas de significado relacionadas entre si, na realidade, exibem um uso completamente distinto no que diz respeito a um dado número de padrões sintácticos/semânticos.

Enquanto que tanto *domandare* como *chiedere* podem transmitir quer o significado interrogativo, quer o imperativo, parecendo ambos bastante naturais para o falante, a análise de corpus revela que *domandare* é quase sempre utilizado no sentido interrogativo, enquanto que *chiedere* o é no imperativo. Quando a mensagem imperativa é expressa através de uma *che-oracão*, o verbo utilizado é **sempre** *chiedere*, e isto ao longo de todo o corpus, embora, é claro, a mesma construção com *domandare* não possa ser considerada como agramatical de um ponto de vista teórico. Este exemplo é surpreendente, mas muitos outros casos de evidência inesperada – para a introspecção – se podem encontrar.

Mais uma vez, um léxico (computacional ou tradicional) deve representar fielmente estes factos e estas divergências de uso em relação ao que é potencialmente aceitável. Devem adoptar-se duas regras:

1) A primeira regra consiste em não julgar o que é descrito no léxico a partir da intuição do falante nativo, já que isto conduz a uma descrição de uma «língua teórica», e não à descrição da língua tal como ela é usada.

2) A segunda regra consiste em permitir, no léxico, uma clara representação (e separação) daquilo que é possível mas só raramente utilizado, e do que é simultaneamente possível e realmente empregue.

## 7. Conclusão

Quais são os aspectos que devem ser captados, de modo a permitir-nos desenhar generalizações lexicográficas úteis? Concordou-se com uma lista de aspectos a diferentes níveis de descrição linguística

(morfológico, sintáctico, semântico). Na anotação dos corpora de acordo com estes aspectos emergiram interessantes correlações entre níveis diferentes, que devem ser codificados no léxico.

A análise de classes semânticas tornou possível identificar o núcleo comum de comportamento linguístico associado a toda a classe, bem como diferenciar desta última os aspectos específicos de cada verbo, ou de subclasses de verbos. Particular ênfase foi dada aos aspectos que se pensou terem um papel indispensável na tarefa de dsambiguação.

A última fase do projecto preconizava ir-se além do nível observacional, numa tentativa de generalização, começando pelas evidências coligidas e utilizando-as como base para uma modelação formal de entradas lexicais.

Só o uso de uma metodologia clara e bem definida, associado a uma abordagem classificatória que estabeleça classes de objectos com o mesmo comportamento, e que descreva classes individuais, permitiu a caracterização dos verbos de acordo com um esquema comum, que, por seu turno, pudesse ser traduzido no código formal requerido por um léxico de PLN, por exemplo, a Linguagem de Representação Typed Feature (TF) adoptada no projecto DELIS.

A representação no quadro do léxico TF levantou, no entanto, as questões e problemas principais surgidos quando se consideram os fenómenos mais relevantes resultantes da análise de evidências provenientes de corpora: i) como determinar o nível apropriado de abstracção dentro da hierarquia para cada tipo de informação, ii) como definir e representar todas as interacções possíveis entre diferentes tipos de informação, a diferentes níveis de descrição linguística, iii) como codificar tipos de informação que versões correntes do HPSG não manuseiam com facilidade, tais como informação semântica lexical complexa, combinatórias, preferências, prototipicidade, informação estatística, modificadores típicos que se combinam apenas com significados particulares de um verbo, etc. A maior parte destes aspectos e fenómenos – tal como se viu acima – não devem ser considerados numa base discreta, mas antes num contínuo. De forma a representar adequadamente informação lexical, qualquer formalismo lexical deveria, portanto, conseguir representar esta situação, de certo modo fluida.

## Notas

\* Tradução: Paula M. Neto.

† Tradução estritamente literal (lit.).

## Referências

- ALONGE A., CALZOLARI N., MONACHINI M. and ROVENTINI A., 1993, "WP2 – Italian Linguistic Annotation", DELIS Draft Report, Pisa.
- BINDI R., MONACHINI M. and ORSOLINI P., 1991, "Italian Reference Corpus – Key for Consultation", NERC Working Paper, Pisa.
- CORAZZARI O., MONACHINI M., ROVENTINI A. and CALZOLARI, N., 1996 "Speech Act and Perception Verbs: Generalizations and Contrastive Aspects", submitted to Euralex.
- FILLMORE C.J. and ATKINS B.T., 1992, "Towards a Frame-based Lexicon: the Semantics of RISK and its Neighbors", in A. Lehrer and E. Kittay (eds.), *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*, Hillsdale, N.J.: Erlbaum Associates, pp. 75-102.
- KRUEGER K. and HEID U., 1993, "On the DELIS Corpus Evidence Encoding Schema (CEES)", DELIS Internal Report, Stuttgart.
- HEID U., 1994, "Relating Lexicon and Corpus Computational Support for Corpus-Based Lexicon Building in DELIS", in *EURALEX94 Proceedings*, W. Martin et al. (eds.), Amsterdam, pp. 459-471.
- MONACHINI M., ROVENTINI A., ALONGE A., CALZOLARI N. and CORAZZARI O., 1994, "Linguistic Analysis of Italian Perception and Speech Act Verbs", DELIS Working Paper, Pisa.