

# ConPor: um modelo de geração conceitual para o português

Lucia Specia, Lucia Helena Machado Rino

Departamento de Computação, Centro de Ciências Exatas e de Tecnologia – UFSCar  
Rod. Washington Luiz, km 235 – 13565-905 – São Carlos – SP – Brasil

{lspecia,lucia}@dc.ufscar.br

***Abstract.** This paper describes ConPor, a conceptual generator that maps Brazilian Portuguese sentences onto UNL, the semantic representation language of the UNL Project. A prototype system for this model is implemented and evaluated. The results show that the system has a good performance for the specific domain and the limited set of linguistic phenomena contemplated.*

***Resumo.** Este trabalho descreve o ConPor, um gerador conceitual que converte sentenças do Português do Brasil para a UNL, a linguagem de representação semântica do Projeto UNL. Um protótipo para esse modelo é implementado e avaliado. Os resultados mostram que o sistema apresenta um bom desempenho para o domínio específico e o conjunto limitado de fenômenos lingüísticos contemplados.*

## 1. Introdução

O ConPor (CONceitualização do PORTuguês) é um gerador conceitual que mapeia sentenças do PB (Português do Brasil) em sentenças ou códigos da linguagem de representação semântica UNL (*Universal Networking Language*) (UNL, 2001). Esse mapeamento é baseado na estrutura sintática da sentença e é realizado por meio da identificação das dependências intra-sentenciais correspondentes, em um processo denominado *mapeamento sintático-conceitual*. Apesar de se assemelhar a outras propostas de análise semântica (ou geração conceitual) (por exemplo, Mitamura and Nyberg, 1992; Dorr and Voss, 1996; Beale et al., 1996; Bean et al., 1998; Germann, 1998; Czuba et al., 1998; Romacker et al., 1999), principalmente na seqüencialização das análises sintática e semântica e no aspecto composicional da análise semântica, o ConPor é o único sistema para a interpretação do PB que utiliza a UNL como linguagem de representação semântica.

O ConPor pode ser utilizado em qualquer aplicação fundamental de PLN (Processamento de Línguas Naturais) que seja dependente de representações conceituais, por exemplo, na tradução e na sumarização automática. Sua proposta original está focalizada, entretanto, no contexto de tradução automática do Projeto UNL (Uchida et al., 1999), o qual utiliza a linguagem UNL como interlíngua. Para contemplar tal comunicação multilíngüe, esse projeto prevê o desenvolvimento ou a customização de dois processos envolvendo cada LN (Língua Natural): a codificação de LN em UNL e a decodificação da UNL em qualquer LN. O ConPor pretende, assim,

apresentar uma proposta similar à da codificação LN-UNL do Projeto UNL, produzindo códigos UNL que podem ser decodificados para outras LNs.

Diferentemente dos tradutores automáticos atualmente disponíveis para o PB, que produzem resultados muito aquém do esperado por não processarem significativamente os aspectos lingüísticos, o ConPor realiza uma análise profunda das sentenças em português, baseada em um léxico semântico, para produzir as sentenças na linguagem UNL. Essa linguagem é descrita na Seção 2. O modelo de geração conceitual do ConPor é apresentado na Seção 3. Uma avaliação preliminar do sistema é discutida na Seção 4. Considerações finais sobre este trabalho são apresentadas na Seção 5.

## 2. A linguagem UNL

A linguagem UNL expressa o significado de uma sentença por meio de conjuntos de relações binárias entre conceitos, cuja sintaxe é dada por  $RL(UW_1, UW_2)$ , sendo RL (*Relation Label*) o rótulo que indica a relação semântica e UW (*Universal Word*) um conceito universal. UWs podem ser genéricas, tais como *book*, *run* ou *John*, ou complexas, indicando variações no significado, por exemplo, *book(icl>room)*. Neste caso, há uma relação de hiperonímia entre *book* e *room*, indicada por *icl*. Essa UW representa, assim, o conceito *to book a room*, ou seja, “reservar um quarto”. Atributos adicionais de um conceito podem ser indicados por meio de ALs (*Attribute Labels*), os quais são geralmente relacionados às características morfossintáticas (tais como tempo e aspecto) dos componentes que remetem a esses conceitos. ALs complementam UWs por meio do símbolo “@”. RLs são indicados por mnemônicos, por exemplo, *agt* para “agente”, *mod* para “modificador”, *obj* para “objeto” e *ptn* para “parceiro”. As relações binárias são identificadas na superfície textual pelas relações sintáticas entre os componentes sentenciais (os quais, via de regra, serão associados a UWs). A Figura 1 ilustra um código UNL para a sentença (1) “A lua crescente ocorre amanhã”.

```
obj(occur.@entry, moon.@def)
mod (moon.@def, crescent)
tim(occur.@entry, tomorrow)
```

Figura 1. Código UNL da sentença (1)

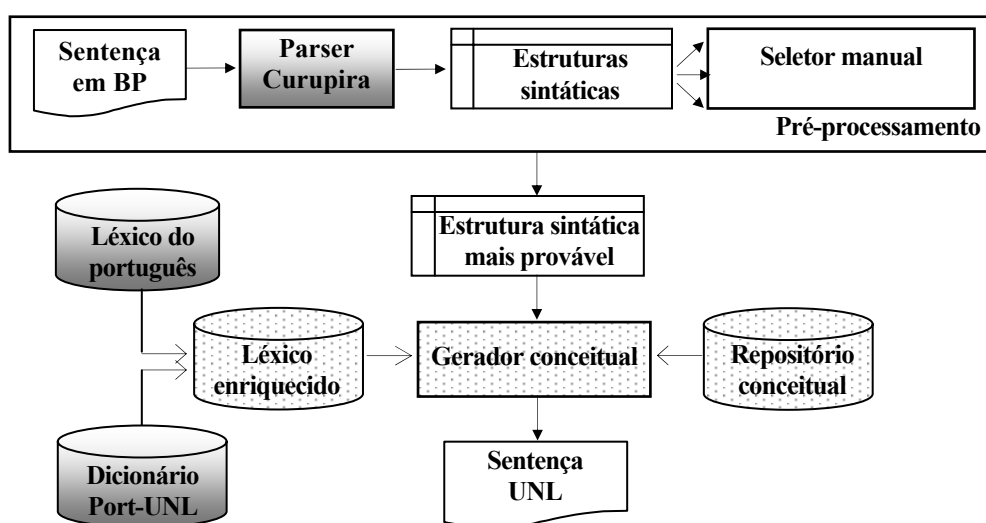
## 3. O sistema ConPor

A arquitetura do ConPor é ilustrada na Figura 2. Os módulos em cinza indicam recursos e processos já existentes no NILC<sup>1</sup> e reutilizados nesse sistema. Os módulos pontilhados indicam os componentes específicos do ConPor. O *parser* Curupira (Martins et al., 2002) ainda está em desenvolvimento e, apesar da sua abrangência, não utiliza informações semânticas, podendo gerar estruturas sintáticas incorretas ou múltiplas para uma única sentença. Considerando somente estruturas sintáticas adequadas para manter o foco no mapeamento sintático-conceitual, o ConPor admite como entrada somente a estrutura sintática mais provável, que é selecionada manualmente dentre as estruturas fornecidas pelo Curupira. O *léxico do português* (Nunes et al., 1996) e o *dicionário português-UNL* (Dias-da-Silva et al., 1998) provêem informações para o *léxico*

<sup>1</sup> Núcleo Interinstitucional de Lingüística Computacional de São Carlos ([www.nilc.icmc.sc.usp.br](http://www.nilc.icmc.sc.usp.br)).

*enriquecido*, um léxico semântico com informações determinantes do mapeamento sintático-conceitual, incluindo os traços semânticos dos substantivos e a estrutura argumental dos verbos. Atualmente, ele contém 514 itens lexicais, definidos com base em um *corpus* de exemplos de 122 sentenças de horóscopos diários, extraídas de jornais brasileiros (público-alvo genérico, portanto).

O *repositório conceitual* contém os modelos de mapeamento sintático-conceitual, os quais alimentam o gerador conceitual. Eles incluem dois componentes: (a) uma gramática de regras de projeção, que mapeiam os constituintes sintáticos em UWs, anotando-as com seus papéis semânticos e (b) um conjunto de *templates* de relacionamento, que determinam os RLs apropriados entre cada par de UWs. A identificação dos papéis semânticos é a tarefa mais importante do mapeamento, em oposição à escolha lexical e à identificação das relações UNL. A escolha lexical é simplificada, uma vez que o léxico enriquecido armazena somente entradas não ambíguas, seguindo uma abordagem similar à das linguagens controladas (Arnold, 1993). A identificação dos RLs corresponde, basicamente, a uma paráfrase dos papéis semânticos usando a terminologia da UNL. Em função da natureza dos dados, a aplicação desses *templates* de relacionamento é também não ambígua: cada par de papéis semânticos indica uma única relação UNL.



**Figura 2. Arquitetura do ConPor**

As regras de projeção são composicionais e são definidas de acordo com as gramáticas livres de contexto (Chomsky, 1965), com conotações gerativas e transformacionais. Além da estrutura sintática, elas são governadas principalmente pela estrutura argumental dos verbos, de acordo com o modelo de Borba (1990)<sup>2</sup>. A granularidade da decomposição das estruturas sintáticas de entrada e, conseqüentemente, das próprias regras de projeção, é baseada nas variações gramaticais encontradas no *corpus* de exemplos utilizado para delinear a construção dos módulos específicos do ConPor. Sua aplicação é realizada de maneira hierárquica, i.e., das regras de projeção mais genéricas para as mais específicas. O processo termina quando as regras não podem mais ser

<sup>2</sup> Derivado do modelo de Chafe (1970) para a língua inglesa.

decompostas e o único passo restante é a escolha das UWs e a atribuição dos papéis semânticos correspondentes para os itens lexicais (consistindo, portanto, do acesso ao léxico enriquecido). Um importante aspecto dessas regras é seu não-determinismo, i.e., a possibilidade de aplicação de várias delas para uma mesma entrada. Isso permite diferentes combinações de regras parciais e, com isso, mesmo estruturas sintáticas não previstas na modelagem do ConPor (contudo, licenciadas) podem ser contempladas por combinações de regras parciais. A Figura 4 ilustra simplificadaamente um conjunto das regras de projeção que se aplicam à geração conceitual para a estrutura sintática da sentença (1), reproduzida na Figura 3<sup>3</sup>.

```
sent([suj(sn([aadn(det(a)),s(lua),aadn(adj(crescente)))])),pred(sv(vi(ocorre))),sadv(adv(amanhã))])
```

**Figura 3. Estrutura sintática da sentença (1)**

```
t_processo → sv_processo, sn_paciente, adv_tempo
sv_processo → verifica_traços_verbo(classe = processo, predicacao = intransitivo),
              recupera_traços_verbo(Tempo, RestSuj),
              recupera_conceito_verbo(UnIV),
              retorna(evento:UnIV.@entry.@Tempo)
```

**Figura 4. Algumas regras de projeção para a sentença (1)**

Essas regras representam apenas parte do modelo de mapeamento para a sentença (1). A regra mais geral, *t\_processo*, mapeia os constituintes dessa sentença (sujeito, predicado e sintagma adverbial) em seus conceitos UNL, juntamente com seus papéis semânticos. Para tanto, ela é decomposta em outras três regras de projeção: *sv\_processo*, *sn\_paciente* e *adv\_tempo*. A regra de projeção para o sintagma verbal (*sv\_processo*), por exemplo, acessa o léxico enriquecido para verificar as restrições do verbo (classe e predicação), recuperar seus traços relevantes (tempo e restrições de seleção argumental) e seu conceito UNL, associando a ele os ALs pertinentes (*@entry* e *@present*) e atribuindo o seu papel semântico (evento). A estrutura intermediária resultante da aplicação dessas regras é ilustrada na Figura 5, mantendo o aninhamento sintático, de modo a preservar as dependências contextuais da sentença original.

```
[[paciente:moon.@def., modificador:crescent], [[evento:occur.@entry.@present],
                                                [tempo:tomorrow]]]
```

**Figura 5. Estrutura intermediária da sentença (1)**

Para identificar as relações UNL entre os componentes dessa estrutura intermediária e, finalmente, produzir a sentença UNL, são aplicadas as regras de relacionamento, que buscam os pares de UWs em uma mesma subestrutura, para combiná-los em uma relação UNL. Por exemplo, para a estrutura intermediária da Figura 5, a regra identificada pelos papéis semânticos “paciente” e “modificador” designa o RL *mod* (e, portanto, uma relação de modo entre os conceitos), resultando no código UNL

<sup>3</sup> sent = sentença, suj = sujeito, sn = sintagma nominal, aadn = adjunto adnominal, det = determinante, s = substantivo, adj = adjetivo, pred = predicado, sv = sintagma verbal, vi = verbo intransitivo, sadv = sintagma adverbial, adv = advérbio.

*mod(moon.@def, crescent)*. Aplicando todas as regras de relacionamento apropriadas, o ConPor produz a estrutura conceitual da Figura 1.

#### **4. A avaliação do ConPor**

Na área de análise semântica, não se tem conhecimento de muitos trabalhos que abordem a avaliação *intrínseca* do desempenho do sistema, i.e., a avaliação das estruturas conceituais produzidas, independentemente do seu uso em qualquer aplicação (Sparck-Jones and Galliers, 1996). Dentre os trabalhos correlatos ao ConPor, somente Bean et al. (1998) e Romacker and Hahn (2000) realizam avaliações dessa natureza. A avaliação de Rosé (2000) também é considerável, apesar de seu trabalho seguir uma metodologia diferente de análise semântica, intercalando-a com a análise sintática. Bean et al. e Romacker and Hahn focalizam a qualidade das estruturas conceituais em relação a representações de referência, medindo a sua cobertura e precisão. Rosé focaliza a qualidade das estruturas conceituais decodificadas para a LN, quando comparadas às sentenças originais, medindo a sua fluência e a preservação da idéia central. A avaliação do ConPor contempla ambos os critérios, por meio da realização de dois testes, descritos abaixo, ambos elaborados com base em um mesmo *corpus* de teste com 80 sentenças de horóscopos em PB, selecionadas de 10 revistas e jornais diários. Esse *corpus* é completamente independente do *corpus* de exemplos, apesar de limitado às construções gramaticais contempladas pela gramática de projeção do ConPor.

##### **4.1. A qualidade das regras de mapeamento sintático-conceitual**

Esse teste pretendeu avaliar a qualidade das regras de mapeamento sintático-conceitual, comparando sentenças UNL geradas pelo ConPor para o *corpus* de teste com sentenças UNL de um *corpus* de referência, construído manualmente a partir do mesmo *corpus* de teste por um especialista em UNL. Essa comparação foi realizada por um juiz humano, usando três escores simbólicos: *sucesso* (para relações coincidentes com as do *corpus* de referência), *falha* (para relações sobressalentes, inexistentes no *corpus* de referência), e *falta* (para relações do *corpus* de referência não geradas pelo ConPor). Com base nos escores obtidos, foram calculadas as medidas de cobertura, precisão e *f-measure* (considerando os mesmos pesos para cobertura e precisão) para cada sentença UNL gerada pelo ConPor. Os resultados foram: 77 sentenças com cobertura e precisão (e, conseqüentemente, *f-measure*) iguais a 1, uma sentença com cobertura e precisão iguais a 0,67, e duas sentenças (que não foram codificadas) com precisão e cobertura iguais a 0. As codificações parcialmente incorretas são devidas à ambigüidade na definição das regras de mapeamento. A falta de codificação, por sua vez, ocorre em função da falta de regras adequadas para determinadas características semânticas do *corpus* de teste.

Adicionalmente, foram calculadas as médias (Md) para precisão, cobertura e *f-measure*. Os resultados foram: MdCobertura = 0,97, MdPrecisão = 0,99 e MdF-measure = 0,97. Esses valores altos são justificados porque (a) as regras de mapeamento permitem que o ConPor tome decisões claramente (elas se referem a construções simples e pequenas do PB, em função da natureza do domínio e do público-alvo); e (b) o *corpus* de teste contempla um conjunto limitado de fenômenos lingüísticos, incluindo somente estruturas sintáticas elementares, já previstas pelo ConPor. Embora (a) e (b) possam sugerir um teste tendencioso, é importante notar que os escores foram obtidos

pela comparação com as sentenças do *corpus* de referência, criado independentemente das decisões do ConPor.

## 4.2. A proximidade semântica

Com esse teste, buscou-se medir a similaridade entre o significado das sentenças do *corpus* de teste e o significado das sentenças em PB, decodificadas das sentenças UNL produzidas pelo ConPor. Essa similaridade é denominada, aqui, de *proximidade semântica*. Foram geradas por um especialista em UNL (que não teve acesso ao *corpus* de teste) 78 sentenças em PB, provenientes das sentenças UNL produzidas pelo ConPor. 27 juízes humanos, falantes nativos do PB, apontaram, então, a proximidade semântica de cada par de sentenças, usando três escores: 1, para nenhuma proximidade; 2, para uma proximidade mediana; e 3, para uma proximidade completa. Calculou-se, assim, a média dos escores dos juízes para cada sentença e a média geral, para todas as sentenças. Esse teste é menos estrito que o primeiro, apesar de mais subjetivo, uma vez que a proximidade semântica é indicada pela intuição dos juízes. Por essa razão, mesmo sentenças UNL consideradas parcialmente inadequadas no primeiro teste podem ser efetivas, do ponto de vista de similaridade de significado das sentenças originais.

Das 80 sentenças do *corpus* de teste, 49 foram imediatamente identificadas como idênticas a seus pares (e, receberam escore 3); as 2 duas sentenças não codificadas pelo ConPor receberam escore 1. Somente 29 delas foram, portanto, realmente submetidas aos juízes, resultando em: 10,3% com escore igual a 3; 75,8% com escore entre 2 e 3; e 13,8% com escore menor que 2. A maioria dos escores baixos refere-se a problemas da representação UNL, em particular, a problemas com os verbos “ser” e “estar”: ambos são representados em UNL pelo RL *aoj* (atributo do objeto). Os melhores resultados envolvem casos de variação de ordem dos constituintes na sentença, principalmente em conjunções. Os resultados médios envolvem também alguns casos com os verbos “ser” e “estar”, casos de ordem de advérbios e casos de variações de gênero (masculino e feminino). O problema relativo à variação de gênero se deve ao fato da UNL não prever tal representação. A variação de ordem de advérbios, bem como de outros constituintes da sentença, é devida à independência entre as relações binárias na sentença UNL: neste caso, elas podem ser decodificadas em qualquer ordem, para o PB. A flutuação de um advérbio em uma sentença, em especial, pode mudar consideravelmente seu significado. Por exemplo, a proximidade semântica atribuída ao par de sentenças (2) “Muitas vezes você só mostrará alegrias” e (3) “Você só mostrará alegrias muitas vezes” é igual a 2,2, pois o advérbio, em (2), parece indicar “frequência” e em (3), “quantidade”.

A média geral obtida com esse teste foi de 2,77, ou seja, o desempenho do ConPor, para a proximidade semântica sobre o *corpus* de teste, foi considerado acima da média, sendo que somente 7% dos 80 pares de sentenças avaliados tiveram significados considerados completamente distantes.

## 5. Discussão final

O ConPor é um analisador semântico que mapeia estruturas sintáticas de sentenças em português em sentenças UNL. Originalmente sugerida para a comunicação multilingual, por exemplo, em um ambiente de tradução automática, a interlíngua UNL, base do ConPor, torna-o um sistema versátil, que pode ser acoplado a qualquer aplicação de

PLN que envolva a representação conceitual. Desse modo, o ConPor pode ser usado tanto para a conceitualização em UNL a partir de estruturas sintáticas, quanto para a decodificação de representações UNL em alguma LN, bastando que se considerem um *parser* e um decodificador UNL-LN. Podem ser explorados, portanto, uma variedade de *parsers* para a derivação de estruturas sintáticas de sentenças em LN, visando, por exemplo, aqueles mais robustos que o utilizado no momento. Entretanto, não há ainda nenhum *parser* para o português que gere estruturas sintáticas com o nível de detalhamento desejado e que apresente um desempenho melhor que o do Curupira. Pode-se, também, considerar outros decodificadores UNL-LN e, com isso, usar o ConPor para gerar estruturas conceituais para quaisquer sentenças de qualquer LN.

Em relação a outros sistemas de metodologia similar, i.e., àqueles que utilizam a análise semântica orientada por modelos de mapeamento (por exemplo, Wilensky and Arens, 1980), o ConPor disponibiliza um mapeamento sintático-conceitual mais flexível, devido à composicionalidade das suas regras de projeção e de relacionamento. Ou seja, seus modelos de mapeamento não contemplam apenas estruturas conceituais completamente pré-definidas, como ocorre normalmente nos demais sistemas. Isto permite que o ConPor seja explorado mesmo para o mapeamento de estruturas sintáticas não previstas inicialmente.

A avaliação intrínseca do ConPor indica que um bom desempenho das regras de mapeamento não garante um resultado (estruturas conceituais UNL) de significado próximo ao original: é possível que um mapeamento adequado conduza a uma proximidade semântica insatisfatória. Além disso, um mapeamento insatisfatório não necessariamente implica ausência de proximidade semântica. Dessa forma, a avaliação do ConPor demonstra que, para o mapeamento sintático-conceitual proposto, as medidas tradicionais de cobertura e precisão precisam ser complementadas com outras medidas, as quais devem contemplar diretamente o significado, como é o caso da medida de proximidade semântica definida. Diferentemente da proposta similar de Rosé, a avaliação realizada procura indicar que a fluência, para o caso do mapeamento em UNL, não é crucial na determinação da proximidade entre potenciais resultados automáticos e dados originais, pois nem sempre uma sentença não fluente implica a falta de proximidade semântica. Para demonstrar que esse aspecto é realmente secundário no ConPor, será necessário realizar testes mais aprofundados, considerando, inclusive, o acoplamento do ConPor a um sistema de decodificação automático. Entretanto, vale lembrar que a granularidade sentencial do ConPor pode ser um fator determinante da insignificância do fator fluência na correspondência entre os significados comparados, muito embora os únicos registros de avaliações de mesma natureza, na Linguística Computacional, também se refiram a unidades sentenciais.

Outra contribuição importante resultante da avaliação do ConPor diz respeito à identificação de problemas de especificação da própria Linguagem UNL: a UNL não fornece mecanismos para diferenciar conceitos que, na superfície textual, implicam distinções de gênero (masculino e feminino) ou de estado (caso dos verbos “ser” – estado permanente, e “estar” – estado temporário, ilustrados antes), muito embora eles sejam discriminantes das entidades referidas na cultura subjacente à LN em foco. A fragilidade na geração de estruturas UNL do ConPor, que permite a ambigüidade de interpretações de sentenças do português, evidencia, ainda, a subespecificação da UNL, corroborando estudos anteriores realizados no Projeto UNL-Brasil (Martins et al.,

2000): seria necessário revisar essa especificação, visando uma representação realmente interlingual, independente de quaisquer LNs. Entretanto, esta tarefa está fora do escopo de desenvolvimento do ConPor, no momento.

As principais limitações do ConPor residem em sua ausência de tratamento amplo dos fenômenos lingüísticos do português, sobretudo ao excluir o tratamento da ambigüidade lexical, resultando em códigos UNL relativamente discretos, os quais implicarão decodificações rígidas e, provavelmente, não condizentes com as variações naturalmente possíveis da LN. Embora essa limitação não seja refletida nos experimentos até agora realizados, extensões da versão atual do sistema devem ser consideradas, visando contemplar tal característica. A proposta imediata para isso é utilizar a base de conhecimento do Projeto UNL (Uchida, 2001), ainda em construção, a qual visa à representação de dois tipos de conhecimento extralingüístico: a probabilidade de dois conceitos co-ocorrerem, i.e., o grau de certeza de dois conceitos se relacionarem em um mesmo contexto, e a distribuição ontológica dos conceitos UNL, i.e., a dicionarização da hierarquia de conceitos universais. Informações de ambas as naturezas permitiriam refinar mais a correspondência de significado entre os itens lexicais dicionarizados e seu contexto de ocorrência e, portanto, aumentar a precisão da codificação UNL.

Outra opção, mais dependente de decisões subjacentes ao próprio ConPor, está em considerar conjuntos mais expressivos de traços semânticos, que possam subsidiar melhor as decisões de restrição de seleção nas regras de projeção e de relacionamento. Informações sobre a prioridade e a freqüência dos conceitos, já previstas no dicionário bilíngue UNL-Português original, também poderão ser utilizadas para buscar esse refinamento no mapeamento sintático-conceitual. Para isso, será necessário considerar novos *corpora* como base da modelagem, que incluam, por exemplo, sentenças maiores e mais complexas, contemplando um conjunto maior de fenômenos lingüísticos.

## Referências

- Arnold, D.J., Balkan, L., Humphreys, R.L., Meijer, S. and Sadler, L. (1993) "Machine Translation: an Introductory Guide", Blackwells-NCC, London.
- Beale, S., Nirenburg, S. and Mahesh, K. (1996) "Semantic Analysis in the Mikrokosmos Machine Translation Project", In: Proceedings of the Symposium on Natural Language Processing, Kaset Sart University, Bangkok, Thailand, p. 294-307.
- Bean, C., Rindfleisch, T.C. and Sneiderman, C.A. (1998) "Automatic Semantic Interpretation of Anatomic Relationships in Clinical Text", In: Proceedings of the AMIA Annual Fall Symposium, Orlando, Florida, p. 897-901.
- Borba, F.S. (1990) "Dicionário gramatical de verbos do português contemporâneo do Brasil", Fundação Editora Unesp, São Paulo.
- Chafe, W. (1970) "Meaning and the Structure of the Language", University of Chicago Press, Chicago.
- Chomsky, N. (1965) "Aspects of the Theory of Syntax", The MIT Press, Cambridge.
- Czuba, K., Mitamura, T. and Nyberg, E. (1998) "Can Practical Interlinguas Be Used for Difficult Analysis Problems?" In: Proceedings of the 2nd Workshop on Interlinguas, Langhorne, Pennsylvania.



- Dias-da-Silva, B.C., Sossolote, C.R.C., Zavaglia, C., Montilha, G., Rino, L.H.M., Nunes, M.G.V., Oliveira Jr., O.N. and Alúcio, S.M. (1998) "The design of a Brazilian Portuguese machine tractable dictionary for an interlingua sentence generator", In: Proceedings of the III Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada, Edited by V.L.S. Lima, Porto Alegre, p. 71-78.
- Dorr, B.J. and Voss, C. (1996) "A Multi-Level Approach to Interlingual MT: Defining the Interface between Representational Languages", In: Natural Language Processing and Knowledge Representation, Edited by L. M. Iwanska and S. C. Shapiro, The MIT Press, Cambridge, 2000.
- Germann, U. (1998) "Making Semantic Interpretation Parser-Independent", In: Proceedings of the 2nd Workshop on Interlinguas, Langhorne, Pennsylvania.
- Martins, R.T., Hasegawa, R. and Nunes, M.G.V. (2002) "Curupira: um parser funcional para o português", Série de Relatórios Técnicos do NILC, NILC-TR-02-26, São Carlos, Dezembro, 43p.
- Martins, R.T., Rino, L.H.M., Nunes, M.G.V., Montilha, G. and Oliveira Jr., O.N. (2000) "An interlingua aiming at communication on the Web: How language-independent can it be?" In Workshop on Applied Interlinguas (Pre-Conference Workshop in conjunction with ANLP-NAACL2000), Seattle, Washington.
- Mitamura, T. and Nyberg, E.H. (1992) "Hierarchical Lexical Structure and Interpretive Mapping in Machine Translation", In: Proceedings of the COLING-92, Nantes.
- Nunes, M.G.V., Vieira, F.M.C., Zavaglia, C., Sossolote, C.R.C. and Hernandez, J. (1996) "A Construção de um Léxico da Língua Portuguesa do Brasil para suporte à Correção Automática de Textos", Tech.Rep. 42 (Relatórios Técnicos do ICMC), USP, São Carlos.
- Romacker, M., Markert, K. and Hahn, U. (1999) "Lean semantic interpretation", In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, Stockholm, p. 868-875.
- Romacker, M. and Hahn, U. (2000) "An Empirical Assessment of Semantic Interpretation", In: Proceedings of the 6th Applied Natural Language Processing Conference & the 1st Conference of the NAACL, Seattle, Washington, p. 327-334.
- Rosé, C.P. (2000) "A Framework for Robust Semantic Interpretation", In: Proceedings of the 1st Conference of the NAACL, Seattle, Washington, p. 311-318.
- Sparck-Jones, K. and Galliers, J.R. (1996) "Evaluating Natural Language Processing Systems: an Analysis and Review", Springer-Verlag.
- Uchida, H. (2001) "UNL Knowledge Base", UNU/IAS/UNL Center, Tokyo.
- Uchida, H., Zhu, M. and Senta, T.D. (1999) "The UNL, a Gift for a Millennium", UNU/IAS/UNL Center, Tokyo.
- UNL (2001) "The Universal Networking Language (UNL) Specification", UNU/IAS/UNL Center, Tokyo.
- Wilensky, R. and Arens, Y. (1980) "PHRAN - A Knowledge Based Natural Language Understander", In: Proceedings of the 18th ACL Meeting, Philadelphia, p. 117-121.