Redes Neurais e Correção Gramatical do Português: um estudo de caso

Andréia Gentil Bonfante
Maria das Graças Volpe Nunes
{andreia, mdgvnune@icmsc.sc.usp.br}
Departamento de Ciências da Computação e Estatística
Instituto de Ciências Matemáticas de São Carlos
Universidade de São Paulo
Caixa Postal 668
13560-970 São Carlos -SP

Abstract

This work aims to investigate the performance of neural networks when used for the task of detecting grammatical errors in texts written in brazilian portuguese. The absence or the incorrect use of “crase” was used as a case study in this work. The results were compared to the ones obtained from the application of ReGra, a brazilian grammar checker symbolist tool. Two models of neural networks were built (Backpropagation and Elman) through the use of the Sttutgart Simulator. The goal of the work is not to determine the better paradigm for this task, but to analyse the performance of the involved models in order to integrate them in an architecture which can capture their best characteristics.

Resumo

Neste trabalho, estudamos a viabilidade da implementação de técnicas conexionistas quando aplicadas na correção gramatical do Português do Brasil, através de um tipo de erro bastante freqüente, que é o relativo ao uso de crase, tanto em casos de ausência, quanto de uso incorreto. Para isso, foram construídos dois modelos de redes neurais (Backpropagation e Elman) com o uso do Simulador SNNS. Os resultados são comparados com os obtidos de testes com ReGra, uma ferramenta de correção gramatical de textos em português, que é um modelo típico de implementação simbolista tradicional. A meta deste trabalho não é determinar o melhor paradigma para esta tarefa, mas analisar o desempenho de modelos envolvidos a fim de integrá-los em uma arquitetura que aproveite suas melhores características.

1. Introdução

Juntamente com a evolução dos processadores de textos, as ferramentas que a eles se incorporam também têm que se aperfeiçoar. As ferramentas que auxiliam o processo de escrita estão cada vez mais poderosas e eficientes, principalmente as de pós-processamento de texto. Com o propósito de melhorar o trabalho do usuário, elas contribuem para a melhoria estrutural do texto, fazendo revisões finais, como o controle da pontuação, verificação ortográfica, apontam erros mecânicos de digitação (controle de abertura e fechamento de parágrafos, etc.), erros gramaticais (uso de crase, concordâncias verbal e nominal, etc.) e ainda geram dados estatísticos do texto (número e tamanho de sentenças, parágrafos, etc.) [Aluísio-95].
No Brasil, ainda são poucas as ferramentas que tratam erros gramaticais, destacando-se o revisor DTS [Reference Software-94] e o ReGra [Nunes-96]. Pela própria complexidade em se modelar a língua portuguesa, a revisão gramatical automática se torna um desafio ao lingüistas computacionais. Um processo até então tradicional de construção de tais ferramentas dá-se através da filosofia simbolista, que lança mão da implementação de regras de produção a serem seguidas e validadas durante o processo de verificação. No ReGra, a implementação conta com o auxilio das ATNs, redes de transição aumentadas [Woods-70], e de uma DCG, gramática de cláusulas definidas [Pereira & Warren-80] para o Português do Brasil. Algumas alternativas surgem ao processo descrito acima, notadamente através do uso de sistemas conexionistas, representados pelas Redes Neurais, na resolução de problemas de Lingüística, em casos como reconhecimento de sentenças [Servan-Schreiber et al.-89], problemas de ambigüidade ([Mayberry & Miikkulainen-94], [Mendes & Eizirik-89]), classificação de sentenças [Lawrence et al.-96] e etiquetagem sintática (atribuição de classe sintática às palavras individuais) [Marques & Lopes-96].
Descrevemos aqui um trabalho que teve como objetivo a observação do comportamento das redes neurais no processo de revisão gramatical do Português. E, como estudo de caso, projetamos dois modelos de redes neurais, um parcialmente recorrente, representado por uma rede Elman [Elman-90], e um modelo backpropagation, que é um dos modelos mais tradicionais usados em aplicações envolvendo redes neurais, tendo ambos como função a detecção de erros com relação ao uso da crase na língua portuguesa do Brasil, tanto em casos de presença incorreta, quanto de ausência.
Nas seções seguintes detalhamos o problema a ser tratado, comentamos a implementação simbolista do ReGra, que serviu como base para algumas comparações e cujo *tagger* (classificador sintático) foi utilizado na classificação das palavras para a implementação das redes, em seguida, a implementação conexionista realizada e, por fim, os resultados dos testes realizados seguidos de sua análise e conclusões.

2. Um tratamento conexionista da crase

Para a tarefa de detecção de erros de crase, foram estudados dois modelos de redes neurais: o modelo Multilayer Perceptron e o modelo Elman de rede parcialmente recorrente [Elman-90]. As redes neurais foram construídas e treinadas com a ajuda do simulador SNNS (Stuttgart Neural Network Simulator) [Reference

Software-95]. Para a obtenção dos padrões de treinamento, ou seja, dos exemplos de ocorrência, contamos com a ajuda de um programa que rastreia parte do banco de textos (*corpus*), construído pelos projetistas do ReGra[1], de onde foram retirados os exemplos aleatoriamente. Assim, as redes foram treinadas com casos em que se usa a crase e também com contra-exemplos, que são os casos onde não se deve usar, sendo assim abordados casos de ausência de crase.

A partir da localização da ocorrência, é retirada da posição uma janela mínima de sua vizinhança, com somente uma palavra antes e uma depois. Com a janela selecionada, esta é submetida ao classificador sintático (*tagger*) do ReGra, tendo como resultado as classes sintáticas e demais traços dessas palavras. A classificação ocorre entre treze classes principais: substantivo, artigo, preposição, numeral, verbo, pronome, adjetivo, nome próprio, abreviatura, conjunção, advérbio, prefixo e sigla. Estas classes possuem subdivisões tais como gênero, número, regência (no caso verbal e nominal), transitividade (no caso de verbos), etc.

Para a codificação, utiliza-se somente as informações previamente consideradas como relevantes: as classes substantivo e adjetivo necessitam das informações quanto ao gênero, número e se regem ou não a preposição ‘a’; as classes numeral, nome próprio e abreviatura necessitam de informações quanto ao gênero e número; a classe pronome possui, além das informações de gênero e número, indicação de ser ou não pronome possessivo; a classe verbo contém informações sobre a transitividade nos casos de transitividade direta e/ou indireta, se está em sua forma infinitiva ou não, e sua regência quanto à preposição ‘a’; a classe advérbio possui informações de modo ou lugar; e por fim, classes que só necessitam de indicação de aparecimento, que são artigo, preposição, conjunção, prefixo e sigla.

O resultado, então, é a obtenção de um vetor de cinqüenta e três posições representando nas vinte e cinco primeiras posições (com cada uma delas indicando presença (1) ou ausência (0) das características mencionadas acima) a palavra anterior à ocorrência do padrão ‘a(s)’, ‘à(s)’, ‘aquele(s)’, ‘àquele(s)’, ‘aquela(s)’, ‘àquela(s)’, ‘aquilo’, ‘àquilo’; nas três posições seguintes, as características da ocorrência, e nas vinte e cinco posições restantes, a codificação da palavra que aparece após a ocorrência.

A título de ilustração, tomemos como exemplo o segmento de frase:

“... *reagiu à onda de boatos* ....”.

O *tagger* do ReGra obtém de seu léxico as seguintes informações:

reagiu: *verbo (intransitivo e transitivo indireto); rege (a, contra, sobre); não infinitivo.*
à: *crase; singular.*
onda: *substantivo; feminino; singular; não rege ‘a’.*

Ambas as redes neurais foram treinadas e validadas com os mesmos conjuntos, sendo que o conjunto de treinamento consistiu de 1576 exemplos e o conjunto de validação 500 exemplos. As estruturas das redes também são coincidentes. Os neurônios estão distribuídos em duas camadas, sendo a entrada constituída de

---

[1] Este corpus é composto por centenas de textos extraídos de jornais, revistas, livros, dissertações, teses e redações de vestibular, constituindo um conjunto de aproximadamente 30 milhões de palavras [Nunes et al.-96b]

cinqüenta e três neurônios, a camada oculta vinte e seis e a camada de saída um neurônio.
A seguir, mostramos os resultados com alguns testes feitos com as redes neurais propostas, implementadas como descrito acima.

3. Resultados obtidos dos testes

As redes neurais apresentaram, no geral, resultados bastante satisfatórios, se levarmos em conta o fato de serem modelos matemáticos que trabalham com vetores representativos dos exemplos a serem treinados. Assim, somente extraindo regularidades estatísticas dos exemplos, elas não são capazes de tratar muitos casos de exceções, como por exemplo, no caso das preposições, que não aceitam crase, salvo no caso de 'até', que pode ser usado como 'até a'; ou então problemas de semântica, como no caso da palavra 'uma', que, quando tem papel de numeral, requer crase, e quando é artigo, não requer crase. No entanto, como o *tagger* classifica de acordo com a classe mais freqüente, e qualquer desambiguação é desconsiderada no modelo conexionista, toda ocorrência de numeral gera um falso erro, ou seja, aponta-se um erro que na verdade não ocorreu.
As tabelas a seguir refletem os resultados de testes com as duas redes implementadas: a rede backpropagation (BP) e a rede Elman (E). Na segunda coluna está o número total de segmentos de sentenças usados no treinamento. A terceira coluna apresenta a porcentagem de acerto da rede em relação ao conjunto total. Por acerto da rede entende-se resultado acima de 0.6 no caso de exemplo certo, e abaixo de 0.4, no caso de exemplo errado. O erro da rede está representado na quinta e na sexta colunas, sendo elas a porcentagem de classificações erradas e a porcentagem de exemplos em que houve dúvidas quanto à classificação, respectivamente. Classificação errada é quando a rede não consegue responder certo ao exemplo apresentado no teste ou então o exemplo não está no treinamento e a rede não conseguiu generalizar corretamente. Por 'dúvida' entende-se quando a resposta da rede estiver no intervalo [0.4..0.6].
A Tabela 1 apresenta resultados de testes feitos em um conjunto de segmentos contendo apenas casos de regência nominal, ou seja, substantivos e adjetivos que aparecem antes da ocorrência e que podem ser regidos ou não pela preposição 'a'. Dos 22 exemplos que a rede backpropagation classificou erroneamente, 10 exemplos (45.4%) ela respondeu conforme estava previsto em seu conjunto de treinamento (caso 1), ou seja, pode se tratar de exceções a esses exemplos[2], ou então problemas de semântica, que também não podemos resolver[3] ; 7 casos (31.9%) não estão no conjunto de treinamento e a rede não conseguiu generalizar corretamente (caso 2) e 5 exemplos (22.7%) que estavam no treinamento e a rede errou na classificação (caso 3). No caso da rede Elman, dos 44 exemplos que apresentam erro, 7 (15.9%) se inserem no caso 1 descrito acima; 10 (22.7%) estão no caso 2 e 27 (61.4%) estão no caso 3. O caso número 3 é talvez o mais

[2] Um exemplo de exceção que podemos citar é o caso da palavra 'até', que apesar de ser preposição e estas não aceitarem crase, ela aceita, quando significar 'até a'. Ex: "Leu o artigo até à última frase."

[3] Alguns exemplos de problemas de semântica: a palavra 'uma', quando tem papel de numeral, e não artigo ("Ela deve chegar à uma hora no lugar marcado."); verbos que podem ser tanto transitivos diretos quanto indiretos ("O profeta passou a mão pela barba.").

importante nesta avaliação, pois é ele quem aponta os erros mais graves das redes neurais, ou seja, os casos que estão previstos no treinamento mas não foram respondidos corretamente. Por isso, optamos por colocá-lo na tabela para que possa refletir mais os erros efetivos.

Tabela 1. Testes com exemplos de regência nominal.

Rede	total	acertos	erros	caso 3	'dúvidas'
BP	100	65%	22%	22.7%	13%
Elman	100	56%	44%	61.4%	0

Um dado que também devemos observar é a porcentagem de dúvidas ou omissões das redes neurais. Geralmente, não é interessante este comportamento da rede, mas para o caso específico de revisão gramatical, acreditamos que, enquanto a rede estiver com dúvidas em algum exemplo, pode ser um falso erro a menos que contamos em suas respostas erradas. Ou seja, é mais vantajoso omitir-se em uma determinada classificação do que indicar um erro ao usuário quando este não existe.
A Tabela 2 apresenta os resultados equilibrados dos testes feitos com um conjunto de segmentos contendo apenas casos de regência verbal, ou seja, somente verbos (transitivos diretos e indiretos e intransitivos), que pedem ou não o uso da preposição 'a'. Para este conjunto de testes a rede BP 'errou' em 66 exemplos dos quais: 40 (60.6%) se inserem no caso 1 descrito acima; 17 (25.75%) estão no caso 2 e 9 (13.65%) estão no caso 3. A rede Elman 'errou' 53 exemplos, onde: 20 (37.7%) se inserem no caso 1 descrito acima; 20 (37.7%) estão no caso 2 e 13 (24.6%) estão no caso 3.

Tabela 2. Testes com exemplos regência verbal.

Rede	total	acertos	erros	caso 3	'dúvidas'
BP	200	63%	33%	3.65%	4.0%
Elman	200	68%	26.5%	24.6%	5.5%

A Tabela 3 apresenta resultados de testes feitos em um conjunto de segmentos contendo apenas casos onde aparecem palavras não abrangidas nos casos anteriores[4], ou seja, palavras que não possuem regência (preposição, pronome, artigo, etc.).
Para este conjunto de testes a rede BP 'errou' em 31 exemplos dos quais: 13 (41.9%) se inserem no caso 1 descrito acima; 17 (54.9%) estão no caso 2 e 1 (3.2%) está no caso 3. A rede Elman 'errou' 44 exemplos, onde: 18 (40.9%) se inserem no caso 1 descrito acima; 20 (45.45%) estão no caso 2 e 6 (13.65%) estão no caso 3.

Tabela 3. Testes com exemplos de palavras sem regência.

Rede	total	acertos	erros	caso 3	'dúvidas'
BP	100	54%	31%	3.2%	15%
Elman	100	52%	44%	3.65%	4%

Os testes seguintes foram feitos com exemplos preparados a partir das regras gramaticais da língua portuguesa [Martins-96].

[4] Estamos falando aqui, como nos testes anteriores, somente das palavras que aparecem antes da ocorrência.

A Tabela 4 apresenta resultados obtidos de testes com um conjunto contendo apenas sentenças corretas. Dos 112 exemplos que a rede BP não classificou corretamente, 69 (61.6%) estão no caso 1; 38 (33.9%) estão no caso 2 e 5 (4.5%) estão no caso 3. A rede Elman não classificou corretamente 130 exemplos, dos quais 33 (25.4%) estão no caso 1; 46 (35.4%) estão no caso 2 e 51 (39.2%) estão no caso 3.

Tabela 4. Testes com exemplos corretos preparados das gramáticas.

Rede	total	acertos	erros	caso3	'dúvidas'
BP	311	56.3%	36%	4.5%	7.7%
Elman	311	55.3%	41.8%	39.2%	2.9%

A Tabela 5 apresenta resultados obtidos de testes com um conjunto contendo apenas sentenças incorretas, de onde a rede BP se destaca com melhor desempenho. Dos 94 exemplos que a rede BP não classificou corretamente: 51 (54.3%) estão no caso 1; 36 (38.3%) estão no caso 2 e 7 (7.4%) estão no caso 3. A rede Elman não classificou corretamente 155 exemplos dos quais: 60 (38.8%) estão no caso 1; 42 (27%) estão no caso 2 e 53 (34.2%) estão no caso 3.

Tabela 5. Testes com exemplos incorretos preparados das gramáticas.

Rede	total	acertos	erros	caso3	'dúvidas'
BP	292	62.8%	32.1%	7.4%	5.1%
Elman	292	41.2%	53%	34.2%	5.8%

3.1. Estudo comparativo com o ReGra

Apresentamos aqui alguns testes feitos com conjuntos de testes que contêm apenas exemplos incorretos do uso ou não da crase. São três conjuntos, contendo respectivamente exemplos de regência nominal (Tabela 6), regência verbal (Tabela 7) e exemplos que não envolvem regência (Tabela 8). Os testes refletem os desempenhos da revisão do ReGra e das duas redes neurais propostas no caso específico do uso de crase.

Tabela 6. Testes com exemplos incorretos de regência nominal.

	exemplos	acertos	erros	'dúvidas'
ReGra	100	30%	51%	19%
BP	100	80%	14%	6%
Elman	100	45%	50%	5%

Tabela 7. Testes com exemplos incorretos de regência verbal.

	exemplos	acertos	erros	'dúvidas'
ReGra	200	29.5%	53.5%	17.0%
BP	200	67.0%	28.5%	4.5%
Elman	200	56.5%	36.5%	7.0%

Tabela 8. Testes com exemplos incorretos não contendo regência.

	exemplos	acertos	erros	'dúvidas'
ReGra	100	43%	56%	1%
BP	100	64%	28%	8%
Elman	100	37%	61%	2%

3.2. Limitações do experimento

As baixas porcentagens de acerto do ReGra são decorrentes do fato de o sistema optar por não indicar erro quando não tem certeza de uma situação. Assim, ele se omite em responder ao invés de indicar um falso erro. Por outro lado, isso não acontece nas redes neurais, que têm sempre que gerar valores de saída para os exemplos apresentados na entrada.

Se analisarmos os resultados apresentados pelas redes neurais isoladamente, acreditamos que ainda assim eles deixam muito a desejar. Talvez isto decorra do fato de termos optado por restringir a uma janela mínima a quantidade de palavras que usamos na codificação, ou seja, temos informações apenas de uma palavra antes e de uma palavra depois da ocorrência. Muitas vezes, devido a esse fato, deixamos de considerar porções relevantes do trecho que envolve a ocorrência. Considere, por exemplo, o caso em que o verbo regido pela preposição não esteja imediatamente à esquerda da mesma. Se houver um modificador (por exemplo, um advérbio) entre o verbo e a ocorrência, a codificação se altera completamente. Assim, se aumentássemos a quantidade de palavras a serem codificadas, talvez poderíamos obter resultados melhores.

Um outro problema que surge, e também gera impacto direto no resultado final, é quanto à classificação das palavras retiradas do léxico. Algumas classes de palavras que possuem várias classificações são hierarquizadas de acordo com a freqüência de uso das mesmas. Por exemplo, 'casa' pode ser substantivo ou verbo, sendo que o léxico fornece a primeira categoria como a mais freqüente. Quando temos casos em que categorias menos freqüentes ocorrem, a rede neural fica prejudicada, pois possui somente a classificação mais usada, e portanto, a errada para o caso. Vale lembrar que são os valores de atributos das palavras vizinhas à ocorrência que determinam os padrões aprendidos pelas redes neurais. ReGra, por outro lado, possui um módulo especialmente dedicado à desambiguação sintática, e que resolve muito satisfatoriamente este problema.

4. Análise dos resultados

Neste trabalho estudamos a viabilidade da implementação de técnicas conexionistas quando aplicadas na correção gramatical do Português do Brasil, através de um tipo de erro bastante freqüente, que é o relativo ao uso de crase. Tais erros são tratados pelo corretor gramatical ReGra, de modelo simbolista tradicional.

O ponto de partida foi um estudo das regras nas gramáticas, obtendo informações que julgamos relevantes para compor o vetor de codificação. A partir daí, tivemos noção de quais informações seriam necessárias quando da passagem dos segmentos pelo classificador sintático.

Com os segmentos de sentença devidamente classificados (uma palavra antes e uma depois da ocorrência), partimos para a composição de dois modelos de redes neurais - um modelo Backpropagation e um modelo Elman - que serviram de instrumento para o trabalho. Lançamos mão do simulador SNNS para a construção dos modelos, já que não pretendíamos nenhuma inovação e os modelos nele propostos satisfaziam a proposta inicial. Após vários testes, chegamos a parâmetros que julgamos ideais nas duas redes neurais, com o uso de apenas uma camada oculta com aproximadamente metade do número de neurônios da entrada e taxa de aprendizado com valor pequeno (0.01). Os ciclos utilizados no treinamento e o erro variaram de acordo com o modelo da rede.

Após as redes estarem treinadas, propusemos vários conjuntos de testes e analisamos resultados. Dois desses conjuntos foram usados como teste no ReGra para que pudéssemos também analisar seu comportamento.

Os testes mostraram que, de uma maneira geral as redes neurais conseguem desempenhar bem o papel de revisão gramatical, salvo problemas que não temos como resolver. Problemas como verbos que se classificam tanto como transitivos diretos ou indiretos, dependendo de onde estejam inseridos; palavras que podem ter vários significados ('casa', 'terra', 'assistir', 'aspirar', 'visar', 'caso', etc.) e outros problemas já relatados.

Cabe-nos analisar outros fatores que tradicionalmente têm influência direta no comportamento das redes neurais: o número de exemplos usados no treinamento; a representatividade desses exemplos, que por serem retirados de partes do *corpus*, podem não abranger todos os casos previstos nas gramáticas; a classificação da palavra que obtemos do léxico, que pode não estar de acordo com o papel que ela de fato exerce naquele contexto; a representação no vetor de codificação, que poderia conter um número maior de palavras que influenciam o fenômeno, etc.

Ainda que o número de exemplos usados (1576) nos pareça bastante razoável, tendo em vista outros experimentos nessa área, é possível, ainda que pouco provável, que um aumento nesse número cause um melhor comportamento das redes. Considerando que esses exemplos são tomados aleatoriamente no *corpus*, um acréscimo, em tese, aumenta a chance de selecionar um maior número de padrões.

O fato de o conjunto de exemplos de treinamento não ter sido formado segundo exemplos constantes das gramáticas não deve impactar tão fortemente o desempenho das redes, uma vez que o que nos interessa na revisão é a detecção de erros reais e mais freqüentes. Logo, a extração feita apenas do *corpus* nos parece acertada.

Quanto à classificação do léxico, os valores obtidos, ainda que algumas vezes problemáticos, representam o que de melhor poderíamos contar, uma vez que se trata do mais completo léxico do gênero para a língua portuguesa do Brasil.

Finalmente, o escopo de representação no vetor de codificação nos parece ser o fator que mais influenciou nos resultados obtidos. Acreditamos que se codificarmos uma região mais abrangente, próxima à ocorrência, estaremos fornecendo mais subsídios à rede neural para suas generalizações. A rigor, a representação de toda a oração envolvida nos parece a solução mais apropriada.

5. Conclusões

O que pudemos observar é que, relativamente sem trabalho algum de implementação, conseguimos chegar a resultados satisfatórios quando comparados com o trabalho de especificação das regras gramaticais usadas no ReGra. As redes neurais não poderiam ser auto-suficientes na revisão gramatical, já que esta precisa abranger muitos detalhes e, mais ainda, suprir o usuário com informações que o ajudem a não cometer mais aquele tipo de erro. Mas, acreditamos que o trabalho de verificação de corretores como o ReGra poderia ser minimizado com a ajuda de redes neurais.

Um cenário que nos parece possível poderia contar com duas fases na revisão gramatical: na primeira, redes neurais treinadas para diferentes classes de erros apontariam possíveis problemas; os casos mais "nebulosos" (valores de saída no intervalo $0.4 \leq x \leq 0.6$, por exemplo) passariam para a segunda fase, para uma verificação por métodos tradicionais. Trata-se de um experimento que envolveria sucessivas e cuidadosas etapas, e de resultados infelizmente ainda imprevisíveis. Isto facilitaria o trabalho do revisor simbolista, que trataria apenas casos mais problemáticos, e acreditamos também que haveria uma considerável diminuição no tempo de processamento, já que o tempo gasto na primeira fase de revisão com a rede neural seria dispendido apenas com a montagem do vetor de codificação, pois uma vez treinada a rede, o tempo gasto para codificar e analisar os exemplos é quase desprezível se comparado com o tempo necessário para as análises sintática e semântica usualmente necessárias nos modelos simbolistas.

Agradecimentos

As autoras agradecem a toda a equipe do NILC (Núcleo Interinstitucional de Lingüística Computacional) da USP/São Carlos pela valiosa colaboração que prestaram à realização deste trabalho. Este trabalho foi parcialmente financiado pela Itautec-Philco S.A., CNPq (#301365-91.1) e CAPES.

Referências

[Aluísio-95] Aluísio, S. M. Ferramentas de Auxílio à Escrita de Artigos Científicos em Inglês como Língua Estrangeira. Tese de Doutorado, IFSC - Departamento de Física e Informática, 1995.

[Bonfante-97] Bonfante, A.G. Uso de Redes Neurais para Correção Gramatical do Português: Um Estudo de Caso. Tese de Mestrado, ICMSC - Departamento de Computação, Março,1997.

[Elman-90] Elman, J.L. Finding structure in time. *Cognitive Science Society*, vol.14, pages 179-211, 1990.

[Lawrence et al.-96] Lawrence, S.; Fong, S.; Giles, L. Natural Language Grammatical Inference: A Comparison of Recurrent Neural Networks and Machine Learning Methods. In: *Symbolic, Connectionist, and Statistical Approaches to Learning for Natural Language Processing. Lecture Notes in AI series.* Springer Verlag, Berlin, 1996.

[Marques & Lopes-96] Marques, N.C.; Lopes, G.P. A Neural Network Approach to Part-of-Speech Tagging. In: *II Encontro para o Processamento Computacional do Português Escrito e Falado*, Curitiba, PR, 1996.
[Martins-96] Martins, R.T. O Tratamento Computacioal da Crase. *Relatório das estratégias de formalização de regras para o emprego da crase na norma-padrão do português do Brasil*, Nov.1996.
[Mayberry & Miikkulainen-94] Mayberry, M.R.; Miikkulainen, R. Lexical Disambiguation Based on Distribute Representations of Context Frequency. In*: Procedings of the 16th Annual Conference of the Cognitive Science Society*, 1994.
[Mendes & Eizirik-89] Mendes, S.B.T.; Eizirik, L.M.R. Um Modelo Neuronal Probabilístico para Solução da Ambigüidade do Português. *Revista Brasileira de Computação* Jul/Set vol.5 pp.45-51, COPPE, Rio de Janeiro, 1989.
[Nunes et al.-96] Nunes, M.G.V.;Ghiraldelo, C.M.; Montilha, G.; Turine, M.A.S.; Oliveira, M.C.F.; Hasegawa, R.; Martins, R.T.; Oliveira Jr.,O.N. Desenvolvimento de um Sistema de Revisão Gramatical Automática para o Português do Brasil. In: *II Encontro para o Processamento Computacional do Português Escrito e Falado*, Curitiba, PR, 1996.
[Pereira & Warren-80] Pereira, F.C.N.; Warren, D. Definite-clause grammar for Natural Languages Analysis. Artificial Intelligence, 13, 1980.
[Reference Software-94] Revisor DTS Software Ltda. User's Guide, 1994-1995.
[Reference Software-95] SNNS (Stuttgart Neural Network Software) V4.1. User's Guide, 1995.
[Servan-Schreiber et al.-89] Servan-Schreiber, D.; Cleeremans, A.; McClelland, J.L. Learning Sequential Structure in Simple Recurrent Networks. In: *Advances in Neural Information Processing Systems.* San Mateo, CA, 1989.
[Woods-70] Woods, W.A. Transition Network Grammars for Natural Language Analysis. *CACM, Special Issue on Computational Linguistics*, 13(2), October 1970, p.591-606;