Mapas Conceituais: geração e avaliação

Cláudia Camerini Corrêa Pérez, Renata Vieira

Programa Interdisciplinar de Pós-Graduação em Computação Aplicada (PIPCA) –
Universidade do Vale do Rio dos Sinos (UNISINOS)
Av. Unisinos, 950 – 93.022-000 – São Leopoldo, RS – Brasil
{claudiap, renata}@exatas.unisinos.br

Abstract. *This paper presents an Information Extraction prototype for the construction of conceptual maps from Brazilian Portuguese texts. An evaluation of the prototype is presented.*

Resumo. *Este artigo apresenta um protótipo de Extração de Informação de textos escritos na Língua Portuguesa do Brasil para a construção de mapas conceituais. Uma avaliação do protótipo é apresentada.*

1. Introdução

O presente artigo apresenta um protótipo de Extração de Informação para construir mapas conceituais a partir de texto da Língua Portuguesa do Brasil, que possam representar o conhecimento de um domínio.

A Extração de Informação (EI) é a tarefa de extrair informação relevante de um texto em linguagem natural, e apresentá-la em uma estrutura formal. A tarefa de EI a partir de texto pode envolver uma seqüência de etapas, como segue: tokenização[1], análise léxico-morfológica, análise sintática, análise semântica, e resolução de co-referência. Para efetuar essas etapas nosso protótipo utiliza um conjunto de ferramentas de Processamento de Linguagem Natural descritas no decorrer deste trabalho.

Em Pérez (2003) foi apresentado um primeiro experimento para extração semi-automática de conhecimento de textos jornalísticos, onde foram selecionadas estruturas sintáticas simples. Em Pérez (2004) uma avaliação preliminar do protótipo foi apresentada. Aqui avaliamos a geração automática de mapas conceituais, utilizando medidas de abrangência e precisão, o grau de conectividade dos grafos, e através da análise dos erros, apresentamos propostas para melhorar os resultados do protótipo.

Este artigo está estruturado conforme segue. Na seção 2, é apresentado o conceito de Mapa Conceitual, formalismo utilizado na representação do conhecimento. A seção 3 apresenta um conjunto de ferramentas que auxiliam no processamento de linguagem natural. A seção 4 apresenta o método para extrair automaticamente de textos as estruturas sintáticas que compõem os mapas, seguindo as etapas de EI. Os resultados e a avaliação do protótipo proposto são expostos nas seções 5 e 6.

[1] Do inglês *tokenization*.

2. Mapas Conceituais

A fundamentação teórica de mapas conceituais está baseada na teoria de Aprendizagem ou teoria de Assimilação, desenvolvida por David Ausubel (1980). Em sua teoria, Ausubel explica como o conhecimento é adquirido e em que forma este fica armazenado na estrutura cognitiva do indivíduo. A estrutura cognitiva pode ser descrita como um conjunto de conceitos, organizados de forma hierárquica, que representam o conhecimento e as experiências adquiridas por uma pessoa. Conceito é um termo que representa uma série de objetos, eventos ou situações que possuem atributos comuns. Conseguir definir conceitos e os relacionar é indício de aquisição de conhecimento que se obtém através de uma aprendizagem significativa [Moreira, 1987].

Baseado nessa teoria, Novak (2004) desenvolveu a metodologia de Mapa Conceitual (MC), procurando representar como o conhecimento é armazenado na estrutura cognitiva de uma pessoa. Nos MCs, conceitos são escritos em retângulos, e o relacionamento entre conceitos é indicado por uma linha entre dois retângulos. Para evidenciar o porquê de um relacionamento entre conceitos, palavras de ligação são colocadas nas linhas, formando proposições simples que mostram o significado do vínculo. Uma proposição é a união entre dois ou mais termos conceituais (conceitos) unidos por palavras de ligação, formando uma unidade semântica. É a criação desta unidade semântica que tem maior valor, ou seja, que dá a razão de ser de um MC, uma vez que a unidade semântica afirma ou nega algo de um conceito. A Figura 1 [Novak 2004] é um exemplo de um MC que descreve a estrutura de mapas.

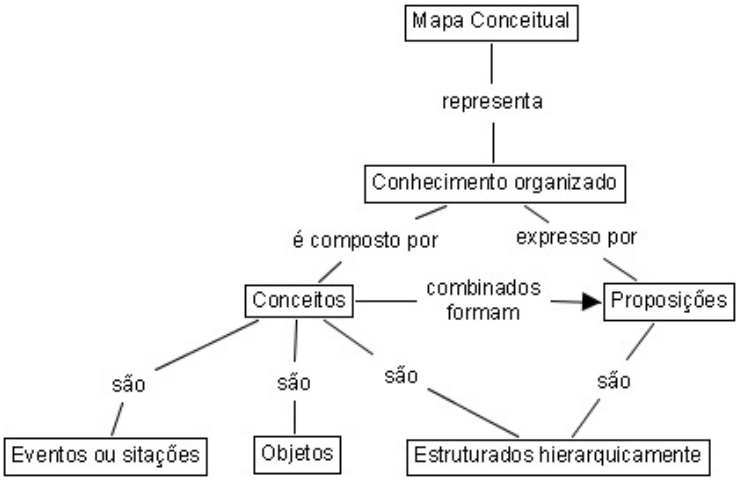

Figura 1. Exemplo de mapa conceitual adaptado de [Novak 2004]

Os MCs vêm sendo utilizados nas mais distintas áreas, tendo diferentes finalidades, como na aprendizagem, na avaliação, na organização e na representação de conhecimento. Giraffa (2002) apresenta uma proposta metodológica para auxiliar na avaliação da aprendizagem de alunos de cursos do tipo Web Based Training (WBT). O processo de avaliação de forma geral baseia-se na comparação do MC que foi utilizado pelo autor do WBT na estruturação e organização do curso, versus o MC gerado pelo aluno. Cañas (2002) apresenta um módulo para a ferramenta Cmap Tools que extrai informações automaticamente de um MC e pesquisa na Web por conceitos que podem ser relevantes para o contexto do mapa. Esses conceitos são sugeridos ao usuário da ferramenta como possíveis candidatos a serem incluídos no mapa. A ferramenta Cmap

Tools, desenvolvida pelo IHMC - University of West Florida [IHMConcept 2004], permite aos usuários construir, navegar, compartilhar mapas conceituais e inserir links para texto, figuras, vídeos e sons.

Assim como um MC pode ser usado com propósitos educacionais para acompanhar o processo de aprendizagem, representando a estrutura cognitiva, consideramos que um MC possa refletir o conhecimento essencial daquilo que está expresso em um texto. Além disso, procuramos verificar se a aplicação de algumas técnicas de Processamento de Linguagem Natural (PLN) podem auxiliar na construção automática ou semi-automática dos mapas conceituais.

A seção 3 apresenta as ferramentas de PLN que auxiliam na análise sintática dos textos e na seleção de estruturas sintáticas para a composição dos termos e relacionamentos dos MCs

3. Ferramentas utilizadas

Para a realização do experimento, foi utilizada uma parte do corpus NILC (Núcleo Interinstitucional de Lingüística Computacional), composto por 7 textos didáticos sobre Ciências da Editora Scipione do ano de 1983.

A EI dos textos foi realizada com o auxílio de ferramentas de PLN. O parser PALAVRAS, desenvolvido para o português por Eckhard Bick [Bick, 2000], realiza as etapas de tokenização, processamento léxico-morfológico, e a análise sintática. PALAVRAS é parte de um grupo de parsers do projeto VISL (Visual Interactive Syntax Learning), do Institute of Language and Communication da University of Southern Denmark. O parser recebe como entrada o conjunto de sentenças de um corpus, e gera a análise sintática das sentenças.

O Palavras Xtractor [Gasperin et al., 2003] foi outra ferramenta utilizada para conversão da saída do parser em arquivos XML. A ferramenta gera a partir da saída do parser, três arquivos XML, o primeiro arquivo XML contém as palavras do corpus com elementos <word> e atributos “id” que representam um identificador único de uma palavra. O segundo arquivo contém as categorias morfo-sintáticas das palavras (P*art-Of-Speech* - POS) e sua forma básica (sem flexões de gênero, número ou grau), onde, por exemplo, o elemento <art> um artigo, o elemento <n> um substantivo e o elemento <v> um verbo.

```
...
<chunk id="chunk_6" ext="subj" form="np" span="word_5..word_6">
  <chunk id="chunk_7" ext="n" form="art" span="word_5"/chunk>
  </chunk>
  <chunk id="chunk_8" ext="h" form="n" span="word_6"/chunk>
  </chunk>
</chunk>
...
```

Figura 2. Arquivo XML com as estruturas sintáticas do corpus – Chunks

O terceiro arquivo contém as estruturas sintáticas das sentenças, representadas por chunks. Um chunk representa a estrutura interna da sentença e pode conter sub-chunks, como ilustrado na Figura 2, onde o <chunk> pai (atributo id=“chunk_6”) é o

sintagma nominal “O homem” (atributo form=“np”) e sujeito da oração (atributo ext=“subj”) os <chunk> filhos (atributo id=“chunk_7” e id=“chunk_8”) são respectivamente um artigo (atributo form=“art”) e um substantivo e também núcleo do sintagma nominal (atributos ext=“h” e form=“n” de *head* e *noun*).

4. Seleção das Estruturas Sintáticas

A identificação das estruturas de interesse para a EI a partir dos textos está baseada no formato de mapas conceituais. Os mapas são representados por triplas (conceito – relação – conceito), que em textos da língua natural tendem a aparecer como sujeito – verbo – objeto. Atribui-se ao verbo a função de estabelecer a relação entre dois conceitos já que é a presença de um verbo que determina a existência de uma oração (Cegalla, 2000; Luft, 2000; Lima, 2002).

A seguir, encontra-se uma apresentação detalhada das estruturas consideradas para a formação de triplas: Argumento1 - Relação – Argumento2.

1) Conceitos

a) Argumento 1 (Sujeito)

- S-(N): o sintagma nominal que exerça a função de sujeito: resgata-se o núcleo e os adjetivos, se existirem. Exemplo: “0 *hormônio antidiurético* atua nos rins.”, o resultado é *hormônio antidiurético*;

- S-(PrR): o pronome relativo “que” exercendo a função de sujeito: recupera-se o núcleo do sintagma nominal que o antecede e os adjetivos, se existirem. Exemplo:“Daí partem nervos que se ligam a gânglios ...”, o resultado é *nervo*;

- S-(VPartic): o verbo no particípio que exerça a função de sujeito: resgate da forma normal para não assumirem a forma infinitiva (seqüestrados – *seqüestrar*; citados – *citar*). Exemplo: “*Citados* negam as acusações”, o resultado é *Citados*;

b) Argumento 2 (Complemento)

- C-(Od): o núcleo do sintagma nominal que exerce a função de objeto direto e o adjetivo se existir, resgate da forma canônica do núcleo do sintagma. Exemplo: “Homozigoto é o indivíduo que apresenta *genes iguais*...”, o resultado é *gene igual*.;

- C-(Oi): o sintagma preposicional que exerce a função de objeto indireto: resgata-se a forma normal de todo o sintagma. Exemplo: “A tireóide localiza-se *no pescoço*.”, o resultado obtido é n*o pescoço*;

- C-(AdjAdn): o adjunto adnominal ou o complemento nominal também pode constituir um terceiro elemento da tripla. Exemplo: “É ao conjunto dessas transformações que chamamos de *crescimento* corporal.”, o *crescimento* é extraído;

- C-(PredSuj): o predicativo do sujeito, complemento do verbo de ligação: resgata-se a forma normal de todo o predicativo do sujeito. Exemplo: “Castor é *meu pai branco*.”, o resultado obtido é *meu pai branco*;

- C-(AdjAdv): recupera-se a forma normal do adjunto adverbial. Exemplo: “Os impulsos nervosos caminham pelos *neurônios*”, é extraído *neurônios*;

- C-(AgP): o agente da passiva: recupera-se a forma normal de todo o agente da passiva. Exemplo: "... um feixe de luz que é captado *por um sistema de lentes*...", onde é resgatado, *por um sistema de lentes*.

2) Relação

- R-(V): o núcleo do sintagma verbal: extraído em sua forma canônica, porém quando se tem uma construção composta, apenas o segundo verbo é extraído e sua forma é mantida (particípio ou gerúndio). Quando temos uma construção composta extraímos o segundo verbo, pois este possui maior peso semântico. Exemplo: "O citoplasma é *formado* por organóides...", o verbo *formado* é resgatado. Em "O pâncreas *produz* a insulina ...", a forma canônica do verbo é regatada.

- Advérbio: é resgatado o advérbio que antecede um verbo, pois a retirada do mesmo pode produzir um sentido oposto à tripla, observe: "A deputada *não* foi encontrada".

O aposto, apesar de sua importância, não foi utilizado. Sempre ligado ao aposto há um elemento que este explica. Para esse trabalho optou-se pelo primeiro elemento, abrindo-se mão do aposto. No exemplo "O diretor do Departamento Geral da Polícia do Interior, delegado Mário Covas (aposto), afirmou...", o elemento extraído é "diretor". Para a representação do conhecimento seria interessante observar a relação entre o elemento e a sua a explicação (diretor = delegado), com o mesmo efeito da resolução de co-referência.

Ao obterem-se as triplas, muitas vezes o pronome relativo "que" é extraído como sujeito em uma determinada oração. Com o objetivo de resgatar o elemento ao qual este pronome se refere, recupera-se o núcleo do sintagma nominal anterior mais próximo. Este procedimento nem sempre resgatará o referente correto, mas no corpus analisado trouxe um resultado satisfatório.

Na Tabela 1 são apresentadas regras para a formação das triplas e a estrutura que compõem as respectivas regras.

Tabela 1. Regras

Regra	Estruturas	Regra	Estruturas
1	S-(N) R-(V) C-(Od)	11	S-(VPartic) R-(V) C-(AdjAdn)
2	S-(PrR) R-(V) C-(Od)	12	S-(N) R-(V) C-(PredSuj)
3	S-(VPartic) R-(V) C-(Od)	13	S-(PrR) R-(V) C-(PredSuj)
4	S-(N) R-(V) C-(Od)(Oi)	14	S-(VPartic) R-(V) C-(PredSuj)
5	S-(PrR) R-(V) C-(Od)(AdjAdv)	15	S-(N) R-(V) C-(AdjAdv)
6	S-(N) R-(V) C-(Oi)	16	S-(PrR) R-(V) C-(AdjAdv)
7	S-(PrR) R-(V) C-(Oi)	17	S-(VPartic) R-(V) C-(AdjAdv)
8	S-(VPartic) R-(V) C-(Oi)	18	S-(N) R-(V) C-(AgP)
9	S-(N) R-(V) C-(AdjAdn)	19	S-(PrR) R-(V) C-(AgP)
10	S-(PrR) R-(V) C-(AdjAdn)	20	S-(VPartic) R-(V) C-(AgP)

Após a geração dos arquivos XML, aplicam-se folhas de estilos XSL[2] (*eXtensible Stylesheet Language*) sobre os arquivos de chunks, para extrair automaticamente dos textos as triplas para formação de MCs. Um documento XSL consegue transformar um documento XML em diversos formatos. Desta forma, é possível criar múltiplas representações da mesma informação a partir de vários documentos XSL.

A seguir, a Figuras 3 mostra as triplas no formato próprio para a ferramenta de representação gráfica dos resultados. Para a representação gráfica das triplas no formato de mapas conceituais (Figura 4), foi utilizada a ferramenta Cmap Tools.

```
homem      ser                  o_mamífero_mais_evoluído_da_escala_zoológica
célula     ser                  a_pequena_unidade_do_ser_vivo
célula     reunir               grupo
tecido     formar               órgão
órgão      em_conjunto_formar   sistema
...
```

Figura 3. Triplas <Argumento1 – Relação – Argumento2 >

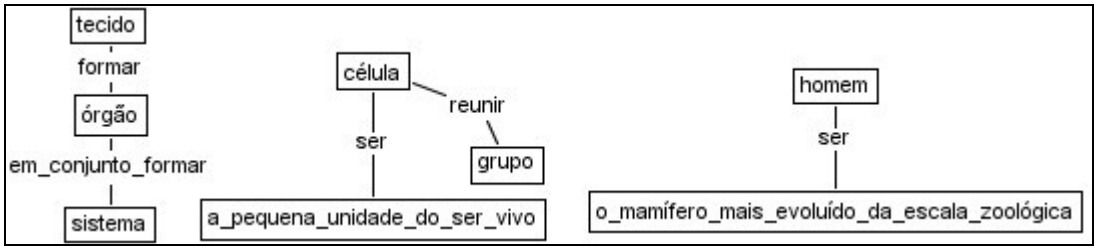

Figura 4. Parte do mapa conceitual extraído automaticamente

5. Resultados

Para a avaliação do protótipo os resumos dos textos didáticos foram apresentados a dois sujeitos para que a partir da leitura e interpretação dos textos mapas conceituais fossem gerados de forma manual. A leitura do resumo e não do texto completo foi adotada, uma vez que os resumos continham a informação principal a ser resgatada pela extração automática, os resumos foram fornecidos juntamente com os capítulos pela própria editora. As triplas geradas automaticamente, conforme as estruturas apresentadas na seção 4 são comparadas com as triplas identificadas pelos sujeitos. Na criação automática dos mapas foram utilizados os textos completos.

Na comparação entre os mapas manual e automático de um mesmo texto, existem muitos aspectos que podem ser observados, como por exemplo, dois mapas distintos podem retratar de forma correta o conhecimento sobre o mesmo domínio. A construção do mapa manual é fortemente associada a questões pessoais da visão do leitor sobre o conteúdo abordado. Isso dificulta a comparação entre os MCs construídos de forma manual e automática. Para facilitar a comparação dos resultados, foi sugerida, aos autores dos mapas manuais, a utilização das mesmas palavras que constam nos textos, por exemplo, procurando não trocar uma palavra que apareça no texto pelo seu

[2] Linguagem desenvolvida pelo W3C (*World Wide Web Consortium*) http://www.w3.org/Style/XSL/

sinônimo. Assim, conceitos e relações extraídos automaticamente podem ser comparados mais facilmente com os extraídos manualmente.

As triplas extraídas automaticamente de cada texto foram avaliadas utilizando as medidas de abrangência e precisão. Abrangência é a fração de itens recuperados relevantes, em relação aos relevantes na base de dados. Precisão é a fração de itens recuperados relevantes, em relação ao total de recuperados.

A Tabela 2 apresenta o número de sucessos, ou seja, triplas automáticas que coincidem com as triplas manuais, o número de triplas geradas automaticamente, mas que não constam nas triplas manuais (Outros), o número de triplas dos mapas construídos manualmente e os valores das medidas de abrangência e precisão das triplas automáticas em relação às triplas manuais (consideradas relevantes e obtidas pelos avaliadores humanos).

Tabela 2. Análise de triplas - Automático e Manual (Sujeito 1 e Sujeito 2)

Corpus	Sujeito 1					Sujeito 2				
	Sucesso	**Outros**	**S1**	**Abr**	**Prec**	**Sucesso**	**Outros**	**S2**	**Abr**	**Prec**
Cie 1	12	31	25	48%	28%	9	34	23	39%	21%
Cie 2	26	185	42	62%	12%	10	201	23	43%	5%
Cie 3	5	47	20	25%	10%	9	43	27	33%	17%
Cie 4	6	11	30	20%	35%	3	14	13	23%	18%
Cie 5	14	91	32	44%	13%	12	93	17	71%	11%
Cie 6	22	109	37	59%	17%	14	117	30	47%	11%
Cie 7	13	76	39	33%	15%	11	78	19	58%	12%
Total	**98**	**550**	**225**	**44%**	**15%**	**70**	**578**	**152**	**46%**	**11%**

Observa-se nos primeiros resultados (sujeito 1) as medidas de abrangência do corpus Ciências que variam de 20% a 62%, apresentando média de 44%. A precisão foi baixa (15%), muitas triplas geradas (85%) não tinham correspondência no mapa manual. Uma explicação para isso é a utilização do resumo para a criação dos mapas manuais e do texto completo para os mapas automáticos. Por isso, aqui consideramos a medida de abrangência mais relevante para avaliar a metodologia proposta.

As observações apresentadas acima se confirmam para o segundo sujeito, como mostram os resultados da Tabela 2. Mesmo que haja uma diferença na interpretação dos textos para os dois sujeitos, o resultado se mantém uniforme, abrangência do corpus Ciências 44% e 46%, precisão 15% e 11%.

A extração utilizando somente a Regra 1 (Tabela 1), proposta inicialmente em Pérez (2003), originou 31 triplas. Com a inclusão das novas estruturas sintáticas, outras regras obtiveram resultados positivos, entre elas: Regra 12 com 25 triplas, Regra 18 com 14 triplas, Regra 6 e Regra 15 com 11 triplas.O experimento obteve um total de 109 triplas utilizadas com sucesso (isto é, casos em que houve correspondência com as triplas selecionadas por qualquer um dos sujeitos). Note que as 109 triplas correspondem a união das triplas com sucesso do sujeito 1 (98) e sujeito 2 (70) contendo algumas triplas em comum. Esse resultado mostra um ganho de 28% obtido pela inclusão das novas estruturas.

Analisamos também o grau de conectividade dos mapas, o número de grafos gerados para cada texto, a quantidade de conceitos interconectados. A Tabela 3 apresenta o número de grafos conforme a quantidade de conceitos do grupo e o total geral de grafos para cada texto. Para uma melhor compreensão das informações dessa tabela, tomamos como exemplo o texto Cie1. O texto possui um total de 18 grafos, onde 12 grafos são formados por 2 conceitos interconectados, 1 grafo formado por 3 conceitos interconectados, 1 grado formado por 4 conceitos, 2 grafos formado por 7 conceitos e o maior grafo interconecta 10 conceitos.

Tabela 3. Conectividade dos mapas

	Grupos de Conceitos													**Total de grafos**
Corpus	**2**	**3**	**4**	**5**	**6**	**7**	**9**	**10**	**11**	**14**	**18**	**62**	**97**	
Cie 1	12	1	1	1		2		1						18
Cie 2	36	4	2	1									1	44
Cie 3	16	3				1				1				21
Cie 4	7	2		1										10
Cie 5	34	5	2	2		2		1	1					47
Cie 6	15	5	2	1	2			1				1		26
Cie 7	24	9	2	1		1					1			38

Idealmente, os mapas construídos automaticamente deveriam corresponder a um grafo único com conceitos interconectados. Porém isso não acontece, há uma grande fragmentação como pode ser observado na Tabela 3. Uma interpretação para isso é o fato de que a geração automática baseia-se na forma das palavras e, em alguns casos, a conexão está no plano semântico (ver conceitos circulados da Figura 5). Nesse plano a conexão pode se dar diretamente como uma relação de sinônimo, por exemplo, *homem* e *indivíduo*, ou pode se dar indiretamente como uma relação entre as palavras que estão no mesmo campo semântico, por exemplo, *cérebro*, *pensamento_memória* e *a_decadência_mental*, que estão no campo semântico das funções cerebrais.

Para uni-los em um único grafo, uma proposta é considerar as variações dos nomes núcleo, por exemplo, *transformação* e *transformações*, *menino* e *menina*. No plano semântico, considerar as relações semânticas diretas, por exemplo, *homem* e *indivíduo*, e indiretas, por exemplo, *puberdade*, *adolecência* e *período_mestrual*.

6. Conclusões

O artigo apresentou um protótipo de Extração de Informação para construir mapas conceituais a partir de texto e sua representação estruturada através de mapas conceituais.

O processo de EI seguiu as etapas de tokenização, processamento morfológico e léxico, e análise sintática; utilizou o *parser* PALAVRAS e a ferramenta Xtractor que gera elementos XML correspondentes à análise fornecida pelo *parser*.

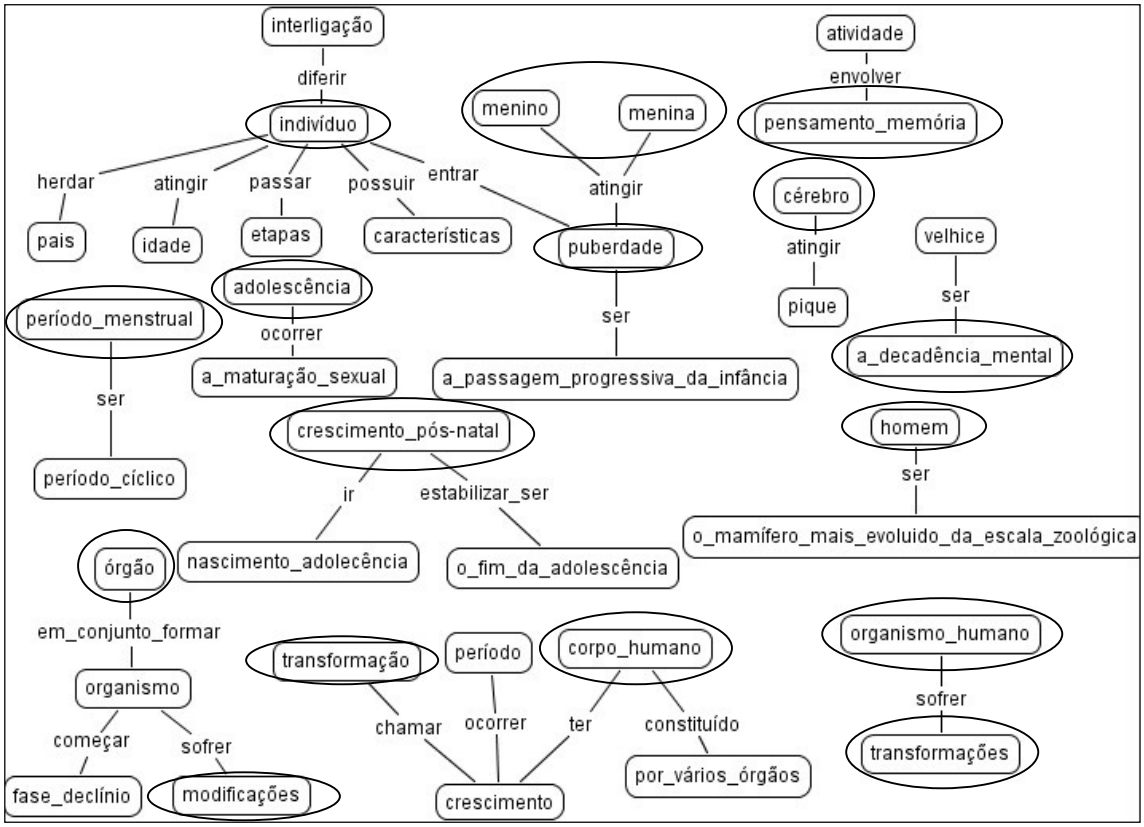

Figura 5. Grafos gerados a partir de Cie1

A seleção das estruturas sintáticas que podem corresponder aos nós e as arestas de um mapa conceitual foram propostas e avaliadas. Adotou-se para avaliação uma análise da similaridade entre as triplas dos mapas gerados automaticamente e os gerados manualmente, como base para o cálculo das medidas de abrangência e precisão das triplas. Observamos também o grau de conectividade dos grafos.

Observou-se que, mesmo que haja uma diferença na interpretação dos textos para os dois sujeitos, o resultado se mantém uniforme, abrangência do corpus Ciências 44% e 46%, precisão 15% e 11%. Em geral, a medida de precisão é baixa. Isso pode ser explicado pela utilização do resumo dos capítulos para criação dos mapas manuais. Pretende-se em um novo experimento utilizar os resumos do texto para geração automática dos mapas, e reavaliar quantitativamente os resultados.

Os resultados apresentados mostram que as regras propostas para EI do texto obtiveram um número considerável de triplas. Uma proposta de trabalho futuro é melhorar a analise das triplas que são geradas automaticamente, mas que não possuem similaridade entre as triplas geradas manualmente. Essas triplas podem conter informações relevantes do texto.

No experimento foram gerados automaticamente vários grafos para cada texto. Como trabalho futuro pretende-se realizar uma análise dos grafos para a construção de propostas para uni-los em um único grafo, semelhante ao mapa construído

manualmente. Uma observação já apresentada é considerar as variações dos nomes núcleo, e no plano semântico, considerar as relações semânticas diretas e indiretas.

Agradecimentos

Este artigo foi parcialmente realizado com apoio da Capes, CNPq, FAPERGS e UNISINOS.

Referências

Ausubel, D. (1980). Psicologia Educacional. - 2. Ed. – Rio de Janeiro: Interamericana.

Bick, E. (2000). "The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework", PH. D. thesis, Arhus University.

Cañas, A., Ford, K., Coffey, J., et al. (2002). "Herramientas para Construir y Compartir Modelos de Conocimiento Basados en Mapas Conceptuales". Revista: Informática Educativa, Vol. 13, No. 2, p. 145-158.

Cegalla, D. (2000). Novíssima gramática da língua portuguesa. São Paulo: Atual, 2000.

Gasperin, C., Vieira, R., Goulart, R. and Quaresma, P. (2003). "Extracting XML Syntactic Chunks from Portuguese Corpora". Proceedings of the Workshop TALN 2003 Natural Language Processing of Minority Languages and Small Languages.

Giraffa, L. Cabral, A. (2002). Avaliação de cursos WBT utilizando Mapas Conceituais. Anais do XII Simpósio Brasileiro de Informática na Educação (SBIE 2002), pp. 504-507.

IHMConcept Map Software a knowledge construction toolkit, Disponível em: http://cmap.coginst.uwf.edu/, acessado em novembro, 2004.

Lima, C. (2002). Gramática normativa da língua portuguesa. Rio de Janeiro: José Olympio.

Luft, C. (2000). Moderna gramática brasileira. Porto Alegre: Globo.

Moreira, M., Buchweitz, B. (1987) Mapas Conceituais – Instrumentos Didáticos e Análise de Currículo São Paulo: Moraes.

Novak, D. (2004). "The Theory Underling Concept Maps and How to Construct Them". Disponível em: http://cmap.coginst.uwf.edu/info, acessado em novembro, 2004.

Pérez, C., Gasperin, C., Vieira, R. (2003). Extração Semi-Automática de Conhecimento. Anais do Encontro Nacional de Inteligência Artificial (ENIA 2003), Campinas - São Paulo.

Pérez, Cláudia C., Vieira, Renata. (2004). Aquisição de Conhecimento a Partir de Textos para Construção de Mapas Conceituais. Workshop de Teses e Dissertações de Inteligência Artificial (WTDIA 2004). São Luís – Maranhão.

Potter, S. (2001). "A Survey of Knowledge Acquisition from Natural Language. Artificial Intelligence Applications Institute", internal project report of University of Edinburgh.