

**CONHECENDO UM *SOFTWARE* TIPO *PARSER*: INSTRUMENTO
AUXILIAR PARA A PESQUISA TERMINOLÓGICA E SUA UTILIZAÇÃO
PELO ESTUDANTE DE LETRAS TENDO EM VISTA A ELABORAÇÃO DE
GLOSSÁRIOS ESPECIALIZADOS.**

Maria José Bocorny Finatto
Universidade Federal do Rio Grande, Instituto de Letras, Projeto Termisul,
Brasil.
mfinatto@terra.com.br

Resumo: Este trabalho, após uma breve apresentação da trajetória dos estudos terminológicos desenvolvidos no Brasil e do papel que desempenham na formação atual do estudante de Letras, apresenta um *software* tipo *parser*, relatando um ensaio de sua aplicação sobre uma amostra de *corpus* lingüístico determinado. Produzido originalmente para fins de análise lingüística e ensino-aprendizagem de idiomas, tal instrumento ainda é pouco familiar à maioria dos alunos de graduação de Letras, embora seja bastante utilizado em diferentes pesquisas lingüísticas, também naquelas que pretendem a obtenção de ferramentas que agilizem o reconhecimento semi-automatizado de terminologias técnico-científicas em diferentes tipos de textos escritos. Tais pesquisas têm sido aplicadas principalmente à obtenção de glossários especializados, o que pode ser uma alternativa de trabalho para o graduado em Letras. O ensaio apresenta uma avaliação inicial do programa quanto à classificação morfológica de uma amostra de textos acadêmicos de Química, tendo sido procedida uma primeira abordagem do reconhecimento de aspectos morfossintáticos de unidades lexicais nominais que integram o *corpus*. Tendo em vista a utilização de outros *softwares* de marcação lingüística, os resultados indicam alguns princípios e critérios a considerar em uma análise mais abrangente sobre comportamento e configuração das terminologias químicas no âmbito da pesquisa terminológica de perspectiva lingüística.

1- INTRODUÇÃO – O CENÁRIO DOS ESTUDOS TERMINOLÓGICOS

O estudo das linguagens da técnica e da ciência é, entre nós brasileiros, relativamente novo. Entretanto, de certo modo, podemos dizer que já experimentou desenvolvimento na década de 80, quando tinham muito destaque, no panorama dos estudos lingüísticos, as pesquisas de línguas estrangeiras instrumentais, que investigavam, com fins didáticos e pedagógicos, as peculiaridades do inglês usado na ciência e na tecnologia.

Há mais ou menos de dez anos, a Terminologia, uma nova disciplina e área de estudos, entendida aqui simplesmente como estudo das linguagens e termos técnico-

científicos, foi introduzida no panorama da pesquisa acadêmica da área de Letras. A introdução deveu-se a professores brasileiros, de formação lingüística, que haviam estado em contato com os estudos terminológicos do Canadá, principalmente na província de Quebec, selecionados pelo programa de bolsas da Associação Brasileira de Estudos Canadenses¹.

A envergadura dos estudos terminológicos canadenses impressionava a todos, pois a pesquisa, a construção de bancos de dados e a elaboração de produtos terminográficos em suporte eletrônico eram velozmente movidas pelas necessidades prementes de tradução de uma comunidade em situação de bilingüismo. Naquele país, nossos professores observaram a implementação e o funcionamento de uma infraestrutura modelar apoiando a tradução e a constituição de uma verdadeira “indústria da língua”, orientada pelo direito do cidadão ao acesso a qualquer informação lingüística fornecida por instituições governamentais, destacando-se, no cenário, o desenvolvimento de *softwares* auxiliares à tradução e reconhecimento de terminologias técnico-científicas.

Além disso, perceberam um trabalho organizado, situado no ambiente de uma política lingüística para a implantação e administração de usos terminológicos em empresas, comércio e serviços em geral, sem que fossem desrespeitadas diferenças de usos e costumes lingüísticos de uma sociedade que emprega o francês, inglês e, ainda, línguas autóctones.

Atualmente, no Brasil, contamos já com o reconhecimento, inclusive institucional, da importância dessa nova matéria de estudos, principalmente nos currículos de nossos cursos de formação de tradutores, visto que cerca de 70% da demanda tradutória nacional localiza-se na chamada *tradução técnica*. Não obstante, mesmo para o estudante de Letras que não se dedique à tradução, uma formação terminológica possibilita outras perspectivas de trabalho, tais como o reconhecimento e adequação de processos de terminologização em materiais didáticos, redação e revisão de textos técnicos e científicos, participação em equipes de elaboração de material instrucional e de elaboração de glossários e dicionários especializados.

A partir daí, revela-se também a importância do trabalho e das relações interdisciplinares: como estudar a linguagem de uma determinada área de

¹ Para maiores detalhes, veja MACIEL (1999).

conhecimento, ciência ou tecnologia sem a colaboração de seus especialistas ou sem algum tipo de contato com seus princípios básicos? Como proceder análises e descrições linguisticamente adequadas, em prazos viáveis, das linguagens especializadas sem o auxílio das ferramentas de processamento de textos?

Na direção das investigações inter e multidisciplinares dinamizadas pelos estudos terminológicos de cunho acadêmico, podemos dizer que já contamos, no Brasil, com muitos trabalhos importantes de reconhecimento terminológico na área da economia², da pesquisa agrônômica, do direito ambiental³, entre outras. Neste sentido, acreditamos que, para estudar, como no nosso caso, as terminologias da Química em português, a partir de uma perspectiva lingüística, será preciso considerar, além, é claro, da sua matéria lingüística propriamente dita, a sua história, visão de mundo enquanto ciência empírica, usos, sistema de conceitos e modos de expressão. Sendo fundamental, enfim, contar com o diálogo e com a cooperação dos químicos para que possamos, como lingüistas, movimentarmo-nos razoavelmente por seus textos e por sua linguagem, quer tenhamos ou não como interesse mais imediato a tradução.

Nesse âmbito, como bem sugere Alain Rey (REY, 1979, p.23) empreender apenas uma análise morfossintática das terminologias científicas seria algo paralisante. Essa imobilidade, conforme entendemos, seria o resultado indesejável de desvincular o “formato” lingüístico das terminologias do seu ambiente maior de significação, constituído no escopo de uma linguagem, que no caso, convencionamos chamar especializada. Por isso, julgamos importante frisar também aqui que as terminologias em geral não podem mais ser confundidas com meros conjuntos de nomenclaturas ou listas de etiquetas denominativas, pois constituem, antes de tudo, linguagens e também são usos diferenciados de uma língua, estando essas duas dimensões sujeitas à ação de diferentes forças.

Sem perder de vista tais especificidades, acentua-se cada vez mais a demanda do processamento de linguagens técnico-científicas, nas quais estão em relevo as terminologias, com vistas ao estabelecimento não mais de bases de dados, mas de verdadeiras bases de conhecimento. Nesse cenário, muitos dos instrumentos já disponíveis para o *Processamento da Linguagem Natural* mostram-se úteis para quem

² Veja Alves (1998), obra produzida no âmbito de pesquisa sobre neologismos terminológicos.

³ Para maiores detalhes, veja a página do Projeto TERMISUL: www.ufrgs.br/termisul

pesquisa na área da Lingüística de Corpus e também para quem repertoria e organiza terminologias técnico-científicas.

Entretanto, conforme vemos, ainda são tímidos o acesso e a familiaridade dos estudantes de Letras com diferentes recursos informatizados que se ocupem de algum tipo de identificação ou classificação da linguagem natural. Entre diferentes recursos, trazemos aqui uma aplicação de um *software*, produzido originalmente para fins de ensino-aprendizagem de idiomas, ao reconhecimento preliminar da terminologia de textos especializados de Química, tratando de caracterizar, em primeiro lugar o tipo de texto de ensaio, depois, uma situação de uso do software com alunos de graduação e, por último, o funcionamento básico do instrumento e suas perspectivas de utilização pelo estudante que se interesse pela pesquisa em Terminologia.

2- O TEXTO ESPECIALIZADO DE QUÍMICA

Embora não sejam abundantes os estudos específicos sobre as terminologias, linguagens ou textos especializados de Química em suas diferentes instâncias, contamos com o valioso trabalho de Godly (1993), cujo título, em português, seria “*Métodos e princípios terminológicos na área de conhecimento da Química*”. Esse autor nos explica algumas de suas peculiaridades, mostrando que, se os químicos inclinam-se a utilizar fórmulas e símbolos entre si durante a experimentação de laboratório, por outro lado, nas suas publicações, objetivando o entendimento de pessoas distanciadas de determinado experimento e com graus de conhecimento variados da matéria, tendem a empregar uma linguagem e uma terminologia moldadas por princípios mais ou menos estabelecidos por sua comunidade. Isso, segundo o autor, em função de vários fatores, gera um tipo de texto de acesso bastante problemático.

Tentando esclarecer-nos sobre causas de uma “obscuridade” do texto especializado de Química, Godly (*op. cit.* p.142) salienta a sua impenetrabilidade crescente, mesmo ao cientista interessado, se esse não possuir determinada especialização ou conhecimento. Além disso, aponta que a construção e o estabelecimento dos termos químicos seriam os fatores principais dessa problemática, visto que possuem diferentes perfis e pertencem a diferentes tipos: nomeiam substâncias, reações, aparelhos, compostos, fenômenos e inter-relações, sem contar

aqueles emprestados da Física, Matemática e Biologia, com significados iguais ou distintos de suas área de origem, e, ainda, aqueles termos oriundos da linguagem do dia a dia.

A terminologia da Química, pelo que percebemos nesse estudo, em que pesem os esforços das denominações recomendadas por diversas instituições, é bastante heterogênea e, muitas vezes, parece comportar-se de modo assistemático. Não obstante, ainda que as nomenclaturas desempenhem um papel importante no reconhecimento da linguagem química, é importante percebermos que seus valores de significação, assim como sua conformação, comportam as variabilidades inerentes a qualquer sistema lingüístico. Assim, antes de examinar o material lingüístico, é importante levar em conta que a univocidade absoluta das terminologias, mesmo nas chamadas ciências exatas, é ainda e apenas um ideal.

3- UM EXEMPLO DE ATIVIDADE TERMINOLÓGICA PRÁTICA COM ALUNOS DE LETRAS: DO TEXTO ESPECIALIZADO À CONFECCÃO DE DICIONÁRIOS OU GLOSSÁRIOS

Feitas essas contextualizações sobre Terminologia e sobre o texto especializado de Química, imaginemos o seguinte exercício-desafio apresentado aos nossos estudantes, alunos da disciplina de Terminologia do curso de graduação em Letras, qual seja, o de propor um dicionário monolíngüe básico de termos de Química a partir de um *corpus* textual especializado, utilizando como primeira referência, não as suas percepções subjetivas sobre quais palavras seriam necessárias ou relevantes dicionarizar, mas seguindo apenas, como critério para a coleta de candidatos a entradas e verbetes, todos os substantivos e de todos os substantivos imediatamente acompanhados de um adjetivo que apareçam nos textos.

Nesse caso, se entregássemos aos alunos um texto em formato Word com cerca de 4000 palavras e solicitássemos, no ambiente de um laboratório de informática, uma identificação manual dessas unidades em um tempo bastante exíguo, certamente os estudantes logo perceberiam a necessidade de encontrar, através do computador, um modo mais ágil de classificação. Nesse ponto da atividade, apresentamos ao aluno um software tipo *parser* que, teoricamente, poderá listar, em poucos minutos e de maneira automática, todas as unidades que obedecem ao critério de coleta inicialmente proposto.

Daí à elaboração da lista de entradas do glossário seria apenas um passo, pensaria o estudante mais afoito, visto que pensará que a tarefa da listagem estará acabada e perfeita. Mas, cabe ao orientador da atividade mostrar que há muito a aprender com o funcionamento adequado e até inadequado desse instrumento e, ainda assim, também muito a lucrar com a agilização do trabalho, principalmente porque há outras e diferentes opções de instrumentos auxiliares, além do fato de que a análise estrutural de um *corpus* envolve diferentes estágios e refinamentos de processamento.

Nesse ponto do exercício, o orientador vincula a pesquisa terminológica, a teoria lingüística e a concepção desses instrumentos informatizados, que necessariamente envolveu conhecimentos gramaticais e lingüísticos. Esclarece ainda o professor aos estudantes que o *software* ao qual estão sendo apresentados foi produzido originalmente para fins de análise lingüística e ensino-aprendizagem de idiomas.

No cenário desta atividade aqui exemplificada, seria importante também relatar que tal instrumento ainda é pouco familiar à maioria dos alunos de graduação de Letras, embora seja bastante utilizado em diferentes pesquisas lingüísticas, também naquelas que pretendem a obtenção de ferramentas que agilizem o reconhecimento semi-automatizado de terminologias técnico-científicas em diferentes tipos de textos escritos. Tais pesquisas, como saberão os alunos desse exemplo fictício, têm sido aplicadas principalmente à obtenção de glossários especializados, mono ou multilíngües, o que pode ser uma alternativa de trabalho para o graduado em Letras, que é, enfim, a situação que o exercício-desafio pretende simular.

Procedida essa exemplificação de uma atividade didática com alunos de graduação, aos quais se apresenta um *parser* e algumas de suas possibilidades de utilização na pesquisa terminológica, voltamos ao relato de ensaio do instrumento, que deve ser compreendido como uma etapa possível de tal atividade.

4- OBSERVAÇÕES PRELIMINARES E METODOLOGIA DO ENSAIO

Nosso objetivo, num primeiro momento, para além de uma situação de aula, foi verificar se as características apontadas por Godly (op.cit.) se aplicariam ao texto especializado de Química em português. Para formar um *corpus* textual de observação, contatamos diversos professores universitários dessa área, no Brasil, e

solicitamos o seu auxílio numa pesquisa sobre o texto especializado de Química no sentido de nos fornecerem versões digitais de suas publicações em português, divulgadas em anais ou periódicos nacionais. Isso segundo, nosso planejamento, agilizaria sobremaneira o trabalho prévio de análise textual, visto que eliminaria uso de *scanners*, e nos daria alguma condição de armazenar um *corpus lingüístico* inicial para futuras investigações terminológicas.

Com alguma dificuldade, em função de diferentes problemas, entre os quais a maciça presença da língua inglesa como veículo da produção intelectual nacional nessa área de conhecimento, fixamos, preliminarmente, para o presente trabalho, uma amostra de 10 textos nas condições acima citadas, cada um com um número médio aproximado de 2.000 palavras. Foram utilizados apenas textos completos, constantes de anais de congressos realizados no Brasil e de revistas acadêmicas nacionais. As revistas fonte foram validadas por professores e bibliotecários do Instituto de Química da UFRGS como as mais relevantes no panorama da investigação acadêmica.

Provisoriamente, os textos publicados obtidos junto aos autores, que nos permitiram seu uso nesta pesquisa, foram separados em dois grandes grupos temáticos: Grupo 1: **ensaios experimentais de Química**; e Grupo 2: **discussão de problemas de educação química**. Esta amostra está constituída por três textos do segundo grupo e por sete do primeiro grupo. Quinze autores, entre eles doze professores ou pesquisadores do Instituto de Química da UFRGS⁴, e os demais professores da FURG, UFMG e UNICAMP, colaboraram com o envio de seus textos publicados ou aceitos para publicação e nos prestaram valiosos esclarecimentos quanto à natureza dos temas, formatos e tipologias usuais dos documentos.

Os textos dos dois grupos, pelo que verificamos, apresentam características macroestruturais genéricas comuns ao texto acadêmico do tipo artigo: resumo, palavras-chave, introdução, metodologia, análise ou desenvolvimento, resultados e discussão ou conclusão. Entretanto, o texto do primeiro grupo, que se caracteriza pela descrição de uma experiência de laboratório, traz geralmente uma parte denominada *experimental* antecedendo o tópico *resultados e discussões*. Além disso, os textos do primeiro grupo tendem a ser menores do que os do segundo. Figuras, tabelas,

⁴ Agradeço a cooperação dos pesquisadores da área de Educação Química da UFRGS (AEQ), coordenada pelo Prof. Dr. Cláudio del Pino e, especialmente, ao pesquisador Marcelo Eichler que, além da colaboração com seus textos, indicou-nos e até forneceu valioso material sobre o perfil da linguagem química.

ilustrações, diagramas e referências bibliográficas não foram apreciados nesta primeira análise dos textos; eventuais erros de emprego de norma culta, tais como concordância ou regência, também não foram objeto de qualquer consideração.

No que se refere a um primeiro reconhecimento da terminologia empregada, o texto de experiência, do grupo 1, tende a apresentar um elevado índice de opacidade ao observador lingüista e ao estudante de Letras, oferecendo a sua leitura um alto nível de dificuldade em função da especificidade de significação de seus termos, principalmente substantivos e adjetivos. De outro lado, observamos que o texto de educação química oferece um menor nível de dificuldade de compreensão de leitura. Isso é que se pode observar pelos dois exemplos abaixo indicados:

“O sistema catalítico bifásico metiltrioxorênio (MTO)-H₂O₂/CH₂Cl₂ foi utilizado na epoxidação do poli (butadieno). Os resultados mostraram alta seletividade, com o grau de epoxidação (10-50%) podendo ser modulado de acordo com a quantidade de oxidante adicionada. As análises por GPC e TGA mostraram, respectivamente, a conservação do peso molecular e a não incorporação do rênio no polímero epoxidado”. **Resumo de artigo, 2, grupo 1 (RA2/G1).**

“Nesta atividade, inicialmente, foram debatidos com os alunos, por exemplo, os aspectos organolépticos e propriedades do leite, suas diferenças de qualidade, seu potencial nutricional e econômico. Apresentou-se para os alunos, 6 (seis) amostras de leite desconhecidas, que deveriam por eles serem identificadas, quanto a sua integralidade. Nas atividades para a determinação da densidade e do percentual de gordura das diferentes amostras, identificou-se o leite como uma mistura de substâncias.” **trecho de artigo, 2, grupo 2 (A2/G2).**

É claro que cada um dos dois tipos de textos envolve terminologia própria, sendo, acertado considerar cada um tão especializado quanto o outro. Entretanto, podemos observar que o texto do grupo 1 utiliza, em geral, termos com um diferenciado grau de especificidade semântica, de circulação mais restrita, tal como se poderia perceber, para os trechos acima citados, *grosso modo*, pela simples oposição dos sintagmas nominais **percentual de gordura** e **polímero epoxidado**.

A esse respeito, nos parece fundamental ressaltar que, apesar dessa diferenciação metodológica em dois grandes grupos temáticos, o texto de educação química, que integra tal *corpus*, é um texto que discute a formação, pertinência e a adequação de alguns conceitos químicos, sua abordagem e reprodução no ambiente escolar. Mas, muitas vezes, pelo que observamos, numa revista específica de educação química, publicada pela Sociedade Brasileira de Química, esse tipo de texto, à semelhança do que ocorre no grupo 1, também descreve e reproduz experiências de laboratório que

serão a base do trabalho com os alunos para a construção de determinados conceitos e princípios químicos.

Assim, vale ressaltar que qualquer tipologia textual aqui identificada tem valor provisório e não se pretende absoluta.

4.1 Análise automática dos textos da amostra

Todos os textos desta amostra foram, num primeiro momento, submetidos à análise do *parser* VISL, um analisador de textos em português e em outros idiomas produzido pelo grupo de pesquisa e desenvolvimento do Instituto de Linguagem e Comunicação da *Odense University*, disponibilizado gratuitamente para uso pelo endereço eletrônico <http://visl.hum.ou.dk>. O objetivo desse grupo de pesquisa é o desenvolvimento, na Internet, de ferramentas para o ensino de idiomas e para o procedimento de análises lingüísticas.

Ao nos aproximarmos, pela primeira vez, do programa, objetivamos verificar as possíveis condições de acesso do nosso estudante e também se a apontada complexidade do texto de Química seria traduzida, num primeiro momento, no âmbito de um classificador automático de palavras, por um alto escore de erros de identificação.

Um *parser* pode ser definido, de acordo com Hartmann (1998, p.106) como um programa de computador de procede automaticamente análises estruturais de frases em termos de suas categorias gramaticais e funções. Conforme Karlsson (1995, p.1), um *parser* realiza, assim, a atribuição automática de estrutura sintática e morfológica, mas não semântica, à produção textual escrita, constituindo um dos estágios de processamento de um corpus lingüístico.

A sistemática de análise de textos oferecida pelo programa em questão é baseada no formalismo de uma *Constraint Grammar* ou gramática de restrições transposta para um *software* que, para o português, possui um dicionário-base formado de artigos de jornais e de textos literários. O instrumento mostrou-se capaz de proceder análises bastante precisas de textos dos mais diferentes tipos, inclusive de alguns textos especializados da área jurídica.

Cada texto de nossa amostra de ensaio foi submetido a uma análise simplificada de classe gramatical de palavras (Flatmorf) e a uma análise de possibilidades morfológicas (Portmorf). A modalidade simplificada de análise procede apenas a

indicação da classe gramatical de cada palavra do texto por meio de um código de cores. A segunda análise justifica morfossintaticamente a indicação de cada classe, aponta radicais, desinências e afixos e ainda apresenta outras opções ou possibilidades de classificação morfológica. Observamos, a seguir, o feitiço da análise de possibilidades morfológicas para a seqüência “**O sistema catalítico bifásico metiltrioxorênio**”:

***o**

[o] DET M S <*> <art> <dem>

[o] PERS M 3S ACC <*>

[o] N M S <*> <let>

[O] N M S <*> <mass> <ABBR>

sistema

[sistema] N M S <ac> <am>

catalítico

[cata] ADJ M S <DERS -al [LOC]> <DERS -ita [HV]> <DERS -ico [ATTR]>

[cata] ADJ M S <DERS -al [LOC]> <DERS -ito [HV]> <DERS -ico [ATTR]>

[cata] ADJ M S <DERS -al [LOC]> <DERS -ite [STATE]> <DERS -ico [ATTR]>

[cata] ADJ M S <DERS -al [LOC]> <DERS -í [DIM]> <DERS -ico [ATTR]>

bifásico

[bifásico] ADJ M S <n>

"<metiltrioxorênio ALT xxxio>"

[metiltrioxorên] N M S <DERS -io [COLL]>

[metiltrioxorên] N M S <DERS -io [COLL]>

[metiltrioxorêna] N M S <DERS -io [COLL]>

[metiltrioxorêno] N M S <DERS -io [COLL]>

Legendas, em inglês, da ferramenta VISL⁵, para os tags de análise morfológica:

- <let> means 'light root' (used, for instance for letter names, which do not derive)
- <ac> is a semantic marker and means 'abstract, countable'
- <am> is a semantic marker and means 'abstract, mass'
- **DERS** means 'derivation via suffix'
- [HV] is a language code marker, meaning 'high level, scientific'

⁵ Agradecemos aqui ao Sr. Eckhard Bick, responsável pela arquitetura do *parser* VISL para o português, pelas informações e esclarecimentos prestados.

language'

- **DIM** means 'diminutivo'
- **ALT** means that the root in question has been altered by the parser in order to encounter a lexicon entry. In this particular case, only a dummy entry, xxx, has been found.
- **COLL** refers to the type of semantic change a suffix induces in its word, here 'collective' (for instance, 'mulheria')
- **ATTR** means 'attributive' and is also used as a suffix marker
- **PATR**, a suffix marker, too, is used for words designating people coming from a certain origin, place or background (like "patronyms"). Normally, 'eno' attaches to place names designating inhabitants, but here, is is used for a chemical root.

Na seqüência exemplificada, **O** foi reconhecido, na análise simplificada, como determinante (DET), **sistema** como substantivo masculino singular, **catalítico bifásico** como adjetivos masculinos singulares e **metiltrioxorênio** como substantivo masculino singular. Pelo que observamos, a análise morfológica simplificada também indica atribuição de caso e regência a cada unidade.

Sem entrarmos na apresentação mais detalhada do sistema de análise das palavras, que não é foco específico deste trabalho, podemos dizer que o índice observado de adequação das classificações, a despeito dos eventuais problemas apresentados, tal como se vê no exemplo da seqüência antes citada, confirmou-se como muito elevado. Salientamos, entretanto, que os textos de cada grupo foram submetidos apenas a essas duas possibilidades do *software*, que inclui outras modalidades seqüenciais de análise e de desambigüação, inclusive algumas puramente sintáticas.

Comentamos, a seguir, alguns dos resultados específicos da marcação de classe gramatical para adjetivos nos dois grupos de textos da nossa amostra preliminar.

5- COMPORTAMENTO DA CLASSIFICAÇÃO DOS ADJETIVOS PELO PARSER VISL

Conforme dissemos antes, o grau de adequação da classificação gramatical foi, para textos em geral, isto é, não terminológicos, bastante elevado. Inclusive, cabe

registrar que seu desempenho mostrou-se bom quando alimentado com alguns textos legais e jurídicos.

Entretanto, observamos que, nos textos do primeiro grupo de nossa amostra, ocorreram alguns problemas dignos de nota quanto à identificação de adjetivos. Além disso, no segundo grupo de textos, relativos à educação química, os problemas também se localizaram nesta classe de palavra, embora tenham ocorrido em número significativamente reduzido. É o que ilustramos abaixo com um trecho do artigo **A2/G1** do *corpus* inicial, no qual as unidades indicadas em vermelho foram classificadas como adjetivos:

“Essa seletividade pode ser aumentada pela diminuição da temperatura e/ou pela adição de uma base ao meio reacional^{3,4}. Além disso, o uso de H₂O₂ como oxidante é uma maneira limpa de se proceder a reação de epoxidação, uma vez que o **único** subproduto **resultante** é a água. No entanto, esse sistema nunca foi estudado na epoxidação de **polidienos**, como o poli(**butadieno**), poli(**isopreno**), etc. **Polidienos epoxidados** têm **grande** aplicabilidade **industrial**, devido às propriedades **adquiridas** após a modificação. Os polímeros em geral são moléculas **apolares**, relativamente **solúveis** em óleos e **solventes apolares**. Grupos **polares** (como **epóxidos**) na estrutura **polimérica** diminuem essa solubilidade. Entretanto, é **necessário** que o **polidieno** não seja totalmente **epoxidado**, pois quando o grau de epoxidação é muito **alto** há dificuldade em se tratar o material **obtido**, principalmente devido à sua **baixa** solubilidade nos **solventes usuais**. Em **baixos** graus de epoxidação (< 50%), os **polidienos** podem passar a ser utilizados como, por exemplo, em mangueiras para óleos ou juntas de motores para automóveis.” **trecho** do texto **A2/G1**

O artigo⁶ **A2/G1** da nossa amostra, intitulado “**Epoxidação de Polidienos Utilizando o Sistema Bifásico Metiltrioxorênio-H₂O₂/CH₂Cl₂**”, apresentou um dos maiores índices de classificação problemática de adjetivos. Num total de 1307 palavras analisadas, 172 foram indicadas como adjetivos. Desse conjunto de 172 unidades apontadas, houve erro em 28 unidades, o que representou, neste documento, uma média de margem de erro de 15% nos adjetivos reconhecidos. De outro lado, a margem de erro representou uma média aproximada de apenas 4% nos textos do grupo 2, sendo que os equívocos de enquadramento em classe gramatical, via de regra, incidiram em termos de estrutura morfológica semelhantes aos termos identificadas em **A2/G1**. Além disso, nos parece oportuno destacar que, em outro

⁶ Agradecemos aqui a gentil colaboração dos colegas professores do Instituto de Química da UFRGS, Departamento de Química Inorgânica: Paulo D. Marcico, Liane L. L. Freitas, Marly A. M. Jacobi, Márcia Martinelli, Annelise E. Gerbase e, em especial, o auxílio e esclarecimentos do colega José R. Gregório.

artigo⁷ do grupo 1, intitulado “**Estudo por Espectroscopia no Infravermelho da Interação Metal-Suporte em Pt/TiO₂. A Influência da Adsorção de Hidrogênio**”, o índice de adequação de classificação de adjetivos foi de praticamente 95%.

Conforme ilustramos a seguir na figura 1, o erro de reconhecimento do termo **polidieno**, no texto **A2/G1**, pelo *parser*, deu-se em função da inadequada vinculação morfológica à base **polido**, marcado como adjetivo no dicionário-base. De outro lado, salientamos a correta classificação, no mesmo documento, de termos como **polímero**, **peróxido de hidrogênio** ou mesmo **metiltrioxorênio**, classificado heurísticamente, isto é, sem que houvesse um registro específico no dicionário-base. Como exemplos de termos incorretamente reconhecidos como adjetivos no texto **A2/G1** destacamos também: **olefinas**, **polibutadieno**, **epóxidos**, **solvente**, **cicloocteno**, **limoneno** e **oxirano**.

<p>polidieno [polido] ADJ M S <DERS -ia [ABSTR]> <DERS -eno [PATR]></p>

Figura 1 – análise morfológica VISL para o termo **polidieno**

6-OUTROS ELEMENTOS A CONSIDERAR NA ANÁLISE AUTOMATIZADA DO TEXTO ESPECIALIZADO

Ao que parece, os substantivos mais inadequadamente reconhecidos pelo *parser*, no texto especializado de Química têm, em relação ao padrão da linguagem não especializada, uma certa “fisionomia” adjetival em função, principalmente, dos seus sufixos característicos, o que se pode apontar como uma das marcas mais salientes e até mais conhecidas da morfologia dos termos químicos. Por outro lado, registramos que esse tipo de texto, mesmo o que descreve um experimento, não é carregado de fórmulas químicas e expressões numéricas ou equações, tal como alguém poderia imaginar. De todo modo, considerando-se o bom desempenho do *software* em outros

⁷ Agradecemos também aos colegas: professor Edilson V. Benvenuti do Laboratório de Superfícies Sólidas, Instituto de Química, UFRGS, Celso U. Davanzo do Instituto de Química da UNICAMP, Marilene M. Zepka (Departamento de Química - Fundação Universidade de Rio Grande), Adriana I. Wassermann, Carlos J. P. Gorga e Ricardo Baumhardt-Neto (Departamento de Química Orgânica - Universidade Federal do Rio Grande do Sul).

tipos de texto, surpreendeu-nos que ele tenha classificado também o termo **dioxina** como um adjetivo, mesmo quando marcado como sujeito de uma frase.

Ao nosso ver, para que empreenda uma análise em melhores condições desse tipo de texto, principalmente daqueles de maior grau de opacidade para o estudante de Letras, poderá ser útil fazer um levantamento prévio mais abrangente das especificidades morfológicas dos nomes químicos, adjetivos ou substantivos, e verbos. Neste particular, acreditamos que vale uma caracterização ou um reconhecimento prévio das características morfossintáticas de substantivos e adjetivos de Química. Ademais, no que se refere ao dicionário-base do analisador, como muitos desses termos provavelmente não constam do *corpus* de textos de jornais e de novelas do *parser*, provavelmente seria uma boa idéia incorporar à sua base as informações de um dicionário comum de língua. Um bom candidato seria o nosso *Dicionário Aurélio* (Ferreira, 1986) que, em sua primeira versão eletrônica, já contava com 130 mil verbetes.

Além disso, no sentido de aperfeiçoar o analisador, para o caso de textos especializados de Química, nos parece que seria interessante: a) utilizar também informações de dicionários terminológicos específicos, capazes de fornecer não só informações conceituais, mas também lingüísticas, com segmentos dos textos de ocorrência dos termos⁸; b) acrescentar ao *corpus* de textos, que gera o dicionário-base, um conjunto de textos especializados tais como os aqui coletados, principalmente os do grupo 1, visto que tendem a representar maior índice de dificuldade para o reconhecimento de unidades pelo *parser*. Isso acabaria qualificando o instrumento que, certamente, atingiria mais algumas condições necessárias para a futura obtenção de verdadeiros bancos de conhecimento, como os que Otman (1996) já refere a partir de *corpora* textuais tecnológicos e científicos, sem esquecermos de mencionar a possibilidade da confecção de hiperdicionários.

7-PERSPECTIVAS E CONSIDERAÇÕES FINAIS

Conforme vimos, ensaio do *parser* revela que há muito a investigar para relacionarmos de modo mais claro uma eventual complexidade do texto de Química com uma “fisionomia particular” de seus substantivos, cujos sufixos carregam valores

⁸ Um ótimo exemplo desse tipo de dicionário, na área da Economia, é o **Glossário de Termos Neológicos da Economia**, vide Alves (1998).

diferenciados, visto que foram preliminarmente classificados como adjetivos pelo programa. Os sufixos da terminologia química, como sabemos, servem para indicar, entre outras coisas, a valência química de substâncias e compostos. Além disso, prefixos e radicais de muitos termos desempenham um papel específico em função de conterem valores químicos sobrepondo-se aos valores lingüísticos usuais.

Isso mostra, tanto para o estudante que se aproxima dessa terminologia na situação de um exercício de aula, quanto ao pesquisador de experimenta *softwares* de marcação lingüística, a necessidade de contar com algum tipo de esclarecimento prévio, por exemplo, sobre funções e valores químicos e, fundamentalmente, sobre a “arquitetura” de sua nomenclatura, que pode ser, em muitos casos, “sinalizadora” desses valores e funções, constituindo-se em estruturas morfossemânticas determinadas.

De outro lado, uma análise abrangente do texto e da linguagem especializada de Química necessariamente terá que ir além da dimensão morfológica quando a feição de suas “palavras” e o valor constitutivo de suas subestruturas ou morfemas são moldados pelas especificidades de uma área de conhecimento ou ciência, mas ao mesmo tempo convivem, num mesmo texto, com outras unidades que não sofre uma mesma influência. Isso é o que também podemos verificar quando nos deparamos com alternância, por exemplo, entre ácido **muriático** e **ácido clorídrico**, pois a escolha de uma ou outra forma pode assinalar, entre outras coisas: determinada perspectiva ou filiação nos estudos de Química; envolver opção por denominações mais “científicas” ou “popularmente conhecidas”. Além disso, a escolha de determinada forma pode indicar a composição ou grau de pureza da substância em questão, somando-se a isso o fato de que, no primeiro caso, o valor semântico do sufixo é -ico é o mesmo da linguagem comum, enquanto que no segundo temos o acréscimo e sobreposição de um semantismo químico.

Finalmente, destacamos que os termos aqui apontados, inadequadamente classificados como adjetivos pelo *parser*, ilustram que é importante, para o caso do texto de Química, que o estudante de Letras, ao aproximar-se desse instrumento em funcionamento, tenha condições de entender não só como o programa é gerado, mas também como os termos químicos são cunhados, sendo capaz de resgatar valores semânticos simbolizados e carregados no *continuum* da linguagem de uma ciência.

Investigar a linguagem da Química é, para nós, lingüistas que se ocupam da comunicação técnico-científica, ainda um grande e importante desafio; este breve ensaio aponta apenas uma das muitas possibilidades de um longo percurso que já estamos trilhando rumo ao seu melhor entendimento. E, nesse percurso, o aperfeiçoamento da capacidade analítico-descritiva das ferramentas informatizadas é fundamental. Nossa intenção, para um futuro próximo, é explorar também com nossos estudantes as particularidades dessa linguagem e de sua terminologia na situação de comunicação estabelecida no âmbito dos livros didáticos de Química utilizados no Brasil, o que poderá interessar, de algum modo, as pesquisas sobre educação química.

8- BIBLIOGRAFIA CITADA

ALVES, Ieda Maria. (1998) (coord.) *Glossário de termos neológicos de Economia*. São Paulo: Humanitas, FFLCH/USP.270p (Cadernos de Terminologia).

HARTMANN, R.R.K.; James, GREGORY. *Dictionary of lexicography*. London:Routledge, 1998.

KARLSSON, F. *et al.* (eds.) *Constraint Grammar*. Berlin: deGruyter, 1995.

MACIEL, Anna Maria Becker.(1999) *A terminologia canadense vista por uma pesquisadora brasileira*. Conferência proferida no I Encontro de Integração de Estudos Canadenses e Produção Lingüística dos Alunos de Letras, promoção do Núcleo de Estudos Canadenses da UFRGS, Instituto de Letras, Porto Alegre, de 7 e 8 de julho de 1999.(inédito)

REY, Alain.(1992) *La terminologie. Noms et notions*. Paris: Presse Universitaires de France. 127p. (Que sais- je?).

GODLY, Ted. W. (1993). *Terminological principles and methods in the subject-field of chemistry*. In: SONNEVELD. Helmi B., LOENING, Kurt L. L. (ed.)**Terminology. Applications in interdisciplinary communication**. Amsterdam: Johns Benjamins, p.141-163.

FERREIRA, Aurélio Buarque de Holanda. (1986) *Novo dicionário de língua portuguesa*. 2.ed. ver. aum. Rio de Janeiro: Nova Fronteira, 1986, 1838p.

OTMAN, Gabriel. (1996) *Les représentations sémantiques en terminologie*. Paris: Masson, 216p.