Filol. linguíst. port., n. 12(2), p. 271-294, 2010.

Aspectos morfológicos da terminologia da nanociência e nanotecnologia

Joel Sossai Coleti[*]
Gladis Maria de Barcellos Almeida[**]

RESUMO: Neste artigo, descrevem-se e analisam-se os aspectos morfológicos do conjunto de termos mais frequentes da terminologia da Nanociência e Nanotecnologia, de forma a cooperar com a fixação e maior compreensão da terminologia da Nanociência e Nanotecnologia, áreas centrais das atividades de pesquisa, desenvolvimento e inovação no mundo todo, como também enriquecer a descrição morfológica do português a partir de um léxico especializado, de maneira a contribuir com o refinamento de conhecimento linguístico útil para aplicação em sistemas híbridos de extração automática de terminologias (aqueles que combinam métodos estatísticos e conhecimento linguístico) considerados mais eficazes.

Palavras-chave: Morfologia; terminologia; nanociência; nanotecnologia.

ABSTRACT: In this paper we describe and analyze the morphological aspects of the set of most frequent terms of Nanoscience and Nanotechnology terminology in order to contribute with the establishment and better understanding of the terminology of Nanoscience and Nanotechnology, which are central areas of research, development and innovation activities worldwide, as well as enriching the morphological description of Brazilian Portuguese through the creation of a specialized lexicon, thus contributing with the development of linguistic knowledge that is useful for application in hybrid automatic terminology extraction systems (those that combine methods statistical and based on linguistic knowledge) considered more effective.

KEY-WORDS: Morphology; terminology; nanoscience; nanotechnology.

[*] Programa de Pós-Graduação em Linguística – Universidade Federal de São Carlos. E-mail: gladis.mba@gmail.com.

[**] Universidade Federal de São Carlos (IC). E-mail: joelscoleti@gmail.com

1. Considerações iniciais

O estudo das unidades terminológicas, que compõem a terminologia da Nanociência e Nanotecnologia (doravante N&N), insere-se no projeto *Terminologia em Língua Portuguesa da Nanociência e Nanotecnologia: Sistematização do Repertório Vocabular e Elaboração de Dicionário-Piloto – NanoTerm*[1].

Nano-, prefixo grego, remete a *nánnos* ("de excessiva pequenez") ou *nânos* ("anão"), adotado na 11ª Conferência Geral de Pesos e Medidas (resolução nº. 12 de 1960) e equivale a um multiplicador 10^{-9}, ou seja, um bilionésimo da unidade indicada, assim, um nanômetro corresponde a 10^{-9} metros (1 nm = 10^{-9} m)[2]. Apenas para que se tenha uma idéia dessa pequeneza, o diâmetro de um fio de cabelo humano mede cerca de 30.000 nanômetros, já "um minúsculo vírus, invisível a olho nu, apresenta-se como uma incrível entidade com cerca de 200 nm" (Toma & Araki, 2005). É desse mundo do *muito pequeno* que tratam a Nanociência e a Nanotecnologia.

A origem do estudo das nanotecnologias é atribuída ao físico Richard Feynman. Feynman, em discurso para a Sociedade de Física Americana em 1959[3], alertou para a possibilidade de fabricar, algum dia, materiais e dispositivos de acordo com as especificações de seus átomos. Entretanto, o termo ***Nanotecnologia*** só viria a ser cunhado em 1974 por Norio Taniguchi, da Universidade de Tóquio, para diferenciar os trabalhos de engenharia no domínio da microescala dos trabalhos em submicroescala, os quais ele denominou ***nanotecnologia*** (Senai, 2004).

As nanotecnologias resultam de trabalhos multidisciplinares, principalmente nas áreas de Química, Física, Biologia, Ciência de Materiais, Medicina, Engenharia e Computação. Fazem parte dos estudos em N&N as capacidades de medir, manusear e organizar a matéria em nanoescala, já que nessa escala a matéria apresenta propriedades, fenômenos e processos únicos (Senai, 2004).

---

[1] Projeto apoiado pelo CNPq (processo 400506/2006-8) e desenvolvido pelo Grupo de Estudos e Pesquisas em Terminologia (GETerm) da Universidade Federal de São Carlos (UFSCar), em parceria com o Núcleo Interinstitucional de Linguística Computacional1 (NILC), sediado na Universidade de São Paulo (USP), campus São Carlos.

[2] Segundo o Dicionário Houaiss da Língua Portuguesa. Disponível em: http://houaiss.uol.com.br/.

[3] O discurso de Feynman foi escrito em uma placa de ouro pela técnica da *nanolitografia*. O texto possui apenas 400 *nanômetros* de largura, assim, em um único fio de cabelo seria possível inserir entre 125 e 250 cópias do texto (Toma, 2004).

No início dos anos 80, com a invenção de novas técnicas de microscopia, desenvolvimentos em nanotecnologia se transformaram em realidade. Em 1985, a primeira descoberta significativa foi feita por pesquisadores da Universidade de Rice, EUA, quando observaram a formação do *Carbono 60,* base para o desenvolvimento dos atuais nanotubos de carbono (Senai, 2004).

Nanociência e Nanotecnologia (N&N) são atualmente áreas centrais das atividades de pesquisa, desenvolvimento e inovação (nos setores de Saúde, Meio Ambiente, Agropecuária, Transportes, Informática, Comunicações, Educação e outros) em crescente expansão no mundo todo. Investimentos aplicados nessa área de conhecimento aumentam a cada ano. Para acompanhar esse desenvolvimento científico e tecnológico, além de investimentos financeiros expressivos e formação de recursos humanos especializados, é preponderante a sistematização de repertórios vocabulares em língua portuguesa (LP). O que se observa é a presença maciça de produtos terminológicos em língua inglesa, mas, ainda assim, limitados em abrangência e profundidade. Daí a necessidade de se voltar os olhos para a terminologia em língua portuguesa de uma área tão nova e promissora.

Ao mesmo tempo espera-se contribuir com a construção de conhecimentos linguísticos úteis para aplicação em sistemas híbridos de extração automática de terminologias (aqueles que combinam métodos estatísticos e conhecimento linguístico) considerados mais eficazes.

A terminologia da N&N, obtida a partir do projeto NanoTerm, é composta por 3.069 termos. Em razão desse número, foi preciso estabelecer critérios para delimitação, de maneira que as análises pudessem ser realizadas num conjunto representativo, mas não exaustivo. Após a aplicação desses critérios (a serem explicitados na seção 3), chegou-se a 295 termos para descrição e 28 para análise.

Os resultados obtidos pela descrição, analisáveis individualmente ou por meio de padrões estatísticos, se implementados em sistemas híbridos de extração, certamente contribuirão com o melhoramento destes sistemas. As análises permitiram detectar padrões morfológicos recorrentes.

Para o desenvolvimento desta pesquisa, utilizou-se a tipologia dos processos de formação de palavras sugerida por Alves (2007), a qual será explicitada na seção 2. Na seção 3, descrevem-se os procedimentos metodológicos adotados para a execução do trabalho. Nas seções 4 e 5, apresentam-se, respectivamente, a "Descrição" e a "Análise", nas quais se apresentam os resultados obtidos. Por fim, são estabelecidas as considerações finais na seção 6.

2. Processos de formação de palavras

Alves (2007) divide em seis os processos de formação neológica, a saber: neologismos fonológicos; sintáticos; conversão ou derivação imprópria; semânticos; os formados por empréstimos; e, finalmente, agrupados no último item estão a truncação, a palavra-valise, a reduplicação e a derivação regressiva.

Segundo a autora, neologismo fonológico é a criação de um item léxico cujo significante seja totalmente inédito, ou seja, tenha sido criado com base em nenhuma palavra já existente. Neologismo sintático é a combinatória de elementos já existentes no sistema linguístico, neste processo, estariam os neologismos formados por derivação e composição. A conversão ou derivação imprópria é um tipo de formação lexical pelo qual uma unidade léxica sofre alterações em sua distribuição sem que haja manifestação de mudanças formais. Neologismo semântico é a criação de um novo elemento a partir de uma transformação semântica manifestada num item lexical. Neologismo formado por empréstimo é a utilização de bases de língua estrangeira.

E, no último item, incluem-se, segundo a autora, processos menos produtivos, mas que também contribuem para o enriquecimento lexical da língua, são eles: truncação, palavra-valise, reduplicação e derivação regressiva. A truncação é um tipo de abreviação em que uma parte da sequência lexical, geralmente a final, é eliminada. Palavra-valise é um tipo de redução em que duas bases são privadas de parte de seus elementos para constituírem um novo item léxico. Reduplicação é um recurso morfológico em que uma mesma base é repetida duas ou mais vezes a fim de constituir um novo item léxico. E, finalmente, derivação regressiva é um tipo de criação lexical em que ocorre a supressão de um elemento, considerado de caráter sufixal.

Nas subseções seguintes, apenas os neologismos sintáticos, a conversão e os formados por empréstimos serão apresentados detalhadamente, já que apenas esses processos foram produtivos no vocabulário da N&N.

2.1 Processos Sintáticos: Derivação e Composição

Processos sintáticos formam novas palavras a partir da combinatória de elementos já existentes na língua. "São denominados sintáticos porque a combinação de seus membros constituintes não está circunscrita exclusivamente ao âmbito lexical" (Alves, 2007: 14).

Os processos sintáticos são subclassificados em *derivação* (formados por derivação prefixal, derivação sufixal e derivação parassintética) e *composição* (podendo esta ser subordinativa, coordenativa, satírica[4], sintagmática, por siglas

[4] Não aplicável a um vocabulário técnico.

ou acronímica). Esta última consiste na "justaposição de bases autônomas ou não-autônomas. A unidade léxica composta, que funciona morfológica e semanticamente como um único elemento, não costuma manifestar formas recorrentes, o que a distingue da unidade constituída por derivação" (Alves, 2007: 41).

A seguir, serão apresentados os tipos de derivação (prefixal, sufixal e parassintética) e de composição (subordinativa, coordenativa, sintagmática, por siglas ou acronímica).

A derivação prefixal consiste na união de um prefixo a uma base, possibilitando o surgimento de variados significados. Prefixos são "partículas independentes ou não-independentes que, antepostas a uma palavra-base, atribuem-lhe uma idéia acessória e manifestam-se de maneira recorrente, em formações em série (ALVES, 2007: 15). Observem-se alguns exemplos retirados da terminologia da N&N: ab*sorção,* ad*sorção,* bio*tecnologia,* de*composição,* infra*vermelho,* semi*condutor,* nano*estrutura,* micro*esfera,* etc.

A derivação sufixal é a união de um sufixo a uma base. Sufixo constitui um "elemento de caráter não-autônomo e recorrente, [que] atribui à palavra-base a que se associa uma idéia acessória e, com frequência, altera-lhe a classe gramatical" (Alves, 2007: 29). Observem-se alguns exemplos retirados da terminologia da N&N: *acopla*mento*, anodiza*ção*, armazena*gem *de hidrogênio, catalisad*or*, condutiv*idade *iônica, cristal*ito*, retarda*nte *de chama,* etc.

A derivação parassintética é aquela "em que o prefixo e o sufixo juntam-se simultaneamente a uma base nominal (...). Nesse processo, é fundamental que os dois afixos incorporem-se ao mesmo tempo à palavra-base" (Alves, 2007: 40). Exemplo de parassíntese poderia ser a palavra a*chocolat*ado.

A composição subordinativa geralmente ocorre entre "dois substantivos, em que o primeiro exerce o papel de determinado e o segundo, de determinante" (Alves, 2007: 41). A autora destaca que, ao contrário do que postulam as gramáticas, o segundo elemento (determinante) pode variar, de acordo com o número. A autora reforça que a subordinação lexical também pode combinar outras classes de palavras, tais como:

- verbo + substantivo: *abaixa-voz, lava-pés, tira-gosto,* etc.
- adjetivo + substantivo: *alta-costura, baixa-falésia, longa-metragem,* etc.;
- substantivo + adjetivo: *abelha-africana, badejo-preto, caba-cega,* etc.;
- numeral + substantivo: *dois-amores, três-irmãos, três-marias,* etc.;
- substantivo + preposição + substantivo: *abóbora-do-mato, caldo-de-cana, cabeça-de-chave,* etc.

A composição coordenativa ocorre "sempre entre bases que possuem a mesma distribuição" (Alves, 2007: 44), de maneira que "a função sintática da coordenação é exercida pela justaposição de substantivos, adjetivos ou membros de outra classe gramatical" (Alves, 2007: 44), como nos exemplos: *badejo-ferro, inhambu-relógio, maníaco-depressivo,* etc.

Alves (2007) diferencia a composição subordinativa da coordenativa da seguinte forma: os componentes justapostos ou coordenados que integram uma composição coordenativa "não manifestam relação de subordinação do tipo determinado/determinante. As bases que compõem a nova unidade lexical desempenham a mesma função que a do elemento recém-formado e associam-se cumulativamente a fim de formarem esse neologismo" (Alves, 2007: 45).

A composição sintagmática é formada por membros que, originalmente, integravam um segmento frasal, os quais se encontram "numa íntima relação sintática, tanto morfológica quanto semanticamente, de forma a constituírem uma única unidade léxica" (Alves, 2007: 50). Ainda de acordo com a autora, "os membros integrantes do composto sintagmático conservam as peculiaridades flexionais de suas categorias de origem" (Alves, 2007: 51).

Ao diferenciar o neologismo criado por este processo dos itens léxicos compostos, Alves (2007) estabelece: "a unidade lexical sintagmática encontra-se ainda em vias de lexicalização. Por isso, não costuma ser unida por hífen. O item léxico composto, ao contrário, é geralmente transcrito com essa marca gráfica" (Alves, 2007: 51) e afirma também que "outro critério, que também revela a lexicalização de um sintagma, supõe o caráter fixo de seus membros integrantes".

Em terminologias, Alves (2007) aponta para a alta frequência de itens léxicos sintagmáticos, resultantes "de uma indecisão em relação à designação de uma nova noção. A denominação em forma de sintagma pode vir a ser substituída por uma única base ou o sintagma pode chegar a cristalizar-se e inserir-se no léxico da língua" (Alves, 2007: 54). Com relação às terminologias, a autora postula que "o vocabulário de uma tecnologia ou de uma ciência em formação condiciona o surgimento de unidades lexicais sintagmáticas em que se observa o empréstimo de termos de disciplinas conexas" (Alves, 2007: 55). Observem-se alguns exemplos retirados da terminologia da N&N: *amostra de espinélio dopada, banda de condução, cadeia polimérica, diâmetro da nanoesfera, eletrólito polimérico gelificado,* etc.

As composições por siglas ou acronímia são, segundo Alves (2007), um tipo especial de composição sintagmática, resultantes do processo de economia discursiva. "O sintagma é reduzido de modo a tornar-se mais

simples e mais eficaz no processo da comunicação" (Alves, 2007: 56). Apesar de apresentarem características variadas, "de maneira mais frequente, o neologismo é constituído pelas iniciais dos elementos componentes do sintagma", podendo também "decorrer da união de algumas sílabas do conjunto sintagmático" (Alves, 2007: 56).

Alves (2007) ressalta que "o neologismo formado por sigla, ao ser empregado pela primeira vez, apresenta-se frequentemente explicado por meio de todo o sintagma ou de sua definição", sendo que, "em seguida, passa a ser usado independentemente do sintagma que o originou" (Alves, 2007: 57).

2.2 Conversão

Conversão é processo que "designa um tipo de formação lexical pelo qual uma unidade léxica sofre alterações em sua distribuição sem que haja manifestação de mudanças formais" (Alves, 2007: 60). Em outras palavras, é quando há alteração da classe morfológica. Exemplo: *o consorciado pode participar de sorteio todos os meses*, em que o item *consorciado* é empregado como substantivo.

2.3 Empréstimo

Alves (2007) aponta diferentes modos de integração dos empréstimos à língua receptora. Um deles constituiria o estrangeirismo, que corresponde ao item léxico estrangeiro empregado em uma língua, mas sentido como não integrante ao acervo lexical do idioma. É um tipo de item lexical facilmente encontrado em vocabulários técnicos, isto porque "o termo estrangeiro é empregado para cobrir uma lacuna de denominação na língua vernácula, o que vem a refletir a origem do desenvolvimento tecnológico e científico de um domínio de especialidade" (Oliveira, 2007: 52). Alguns exemplos poderiam ser: *software, hardware, shopping center,* etc.

Outra maneira de integração do estrangeirismo é a tradução dos itens, de forma a facilitar a recepção dos textos, como nos exemplos: *arquivo* < *file, rascunho* < *draft, imprimir* < *print,* etc.

A adaptação gráfica, morfológica ou semântica do item léxico estrangeiro é considerada a fase neológica, já que é nesse processo que o item léxico estrangeiro se integra à língua receptora (Alves, 2007). No entanto, "a incorporação ortográfica de uma unidade lexical estrangeira ao sistema português não constitui uma regra" (Alves, 2007: 77). Um exemplo seria *scanner* > *escâner*.

Um terceiro "modo de integração de uma formação estrangeira a outro sistema linguístico é representado pelo decalque (...) [que] consiste na versão literal do item léxico estrangeiro para a língua receptora" (Alves, 2007: 79). Exemplos[5] poderiam ser: *ton sur ton* > *tom sobre tom, peau d'ange* > *pele de anjo, shirt-dress* > *vestido-camiseta,* etc.

3. Procedimentos metodológicos

Nesta seção, são apresentadas e discutidas as etapas da pesquisa. Inicialmente, serão relatadas a extração, seleção, lematização e validação dos termos. Na sequência, serão abordados os procedimentos de identificação dos processos de formação de palavras e a sistematização dos dados. Será, ainda, apresentado o recorte para análise. E por fim, a metodologia de análise.

3.1 Extração, Seleção, Lematização e Validação dos Termos

A partir do corpus[6] em língua portuguesa de N&N, elaborado entre 2006 e 2007, foram extraídos de forma semiautomática os termos que compuseram a terminologia, objeto de análise desta pesquisa.

O corpus de N&N está distribuído pelos seguintes gêneros textuais: científico, científico de divulgação, informativo, técnico-administrativo e outros. Compõem o corpus 1.057 textos de 57 fontes diferentes, totalizando 2.739.621 palavras. O corpus foi finalizado em 12 de julho de 2007.

A extração de candidatos a termos é um procedimento semiautomático que é revisto pelo linguista e avaliado pelo especialista de domínio e, por isso, é denominado **semi**automático, justamente por necessitar da interferência humana. São extraídos candidatos a termos simples (*unigramas*) e complexos (formados de duas, três ou quatro palavras, chamados respectivamente de *bigramas, trigramas* e *tetragramas*).

O trabalho do linguista se faz presente desde a elaboração da *stoplist*[7], já que mesmo que se possa aproveitar *stoplists* disponíveis na web, recomen-

---

[5] Exemplos retirados do artigo "O neologismo por empréstimo no vocabulário da moda", de Emilia Maria Peixoto Farias (Universidade Federal do Ceará), disponível em http://www.filologia.org.br/vcnlf/anais%20v/civ2_12.htm (acesso em 17/01/2010).

[6] O corpus do projeto NanoTerm foi elaborado por Joel Sossai Coleti e Daniela Ferreira de Mattos, ambos com bolsa PIBIC/CNPq.

[7] A *stoplist* é formada por uma lista de palavras que devem ser evitadas pelo programa na geração dos candidatos a termos, pois não são relevantes para o intuito da pesquisa.

da-se a adaptação dessa lista para o corpus com o qual se trabalha, para que se possam alcançar melhores resultados. Outra tarefa do linguista é a revisão da lista de candidatos a termos gerada no método estatístico.

Na revisão da lista, feita antes da validação pelo especialista de domínio, procede-se a uma limpeza dos falsos candidatos a termos. Os *candidatos a termos* constituem itens léxicos que se comportam nos seus respectivos contextos como termos, mas cuja autenticidade será validada posteriormente. A validação dos candidatos a termos pode ser feita das seguintes formas: 1) pela comparação da lista de candidatos com uma lista de itens léxicos provenientes de um corpus de referência (corpus da língua geral); 2) pela submissão da lista de candidatos à análise de especialista do domínio; 3) pela utilização dos dois procedimentos sequencialmente, ou seja, comparam-se as listas e, após a comparação, submete-se o resultado à apreciação do especialista (Almeida & Vale, 2008: 484).

Para a realização do procedimento de extração semiautomática, foram avaliadas três ferramentas computacionais, com o apoio do NILC, são elas: o programa Unitex, a ferramenta de extração presente no Corpógrafo e o Pacote NSP.

O Unitex[8], desenvolvido na Universidade Marne-La-Vallée (França) por Sébastien Paumier, consiste num conjunto de programas que permite o processamento de grandes quantidades de textos, em diversas línguas. Uma característica que o diferencia de outros programas que trabalham com corpus é o fato de funcionar com base em dicionários eletrônicos de cada uma das línguas que o integram. Esses dicionários possibilitam ao usuário do programa a realização de buscas pela forma exata, pela forma canônica e também por categorias gramaticais. Além disso, o programa permite a combinação desse tipo de busca com a busca por formantes. Essas características fazem com que ele possa ser particularmente útil em buscas de construções complexas (Almeida & Vale, 2008). Entretanto, é capaz de gerar apenas listas de unigramas, sendo que para a obtenção de termos complexos é preciso que cada item léxico de alta frequência seja testado no concordanciador,[9] observando as demais palavras que estão em seu entorno, o que torna a tarefa

---

Integram esta lista preposições, conjunções, pronomes, artigos, verbos modais, determinados nomes próprios, etc.

8 Mais informações podem ser obtidas em: http://www-igm.univ-mlv.fr/~unitex/.

9 O concordanciador é uma ferramenta que permite listar as ocorrências no texto de uma determinada palavra com seu contexto imediato (TERMISUL, http://www6.ufrgs.br/termisul/ferramentas.php).

altamente suscetível a erros e extremamente morosa, já que se manipula um corpus com 2.739.621 palavras.

Analisou-se em seguida a ferramenta de extração presente no ambiente web Corpógrafo[10], desenvolvido pelo Polo da Linguateca sediado na Faculdade de Letras da Universidade do Porto (FLUP), Portugal. O Corpógrafo é um gestor de corpus planejado para o desenvolvimento de pesquisas terminológicas, isto é, para a extração de termos e sua organização em bases de dados. Fornece um ambiente web integrado para o manejo de corpus, disponibilizando ferramentas para o seu processamento. Dentre as ferramentas que possui, estão: extratores, concordanciadores, contadores de frequência e ferramentas de pré-processamento de corpus, como as de limpeza e sentenciadores. Toda a funcionalidade do Corpógrafo está associada a um dos quatro ambientes de trabalho ou módulos: gestor de arquivos e de corpus, pesquisa em corpus, centro de conhecimento, centro de documentação. Em sua versão atual, o Corpógrafo proporciona maior suporte para a difusão dos produtos terminológicos, permitindo que se exportem bases terminológicas no formato XML para serem usadas em outras aplicações; que se gerem glossários em HTML usando uma base terminológica, além de se exportar o próprio corpus em XML para ser usado em outras aplicações ou por outros usuários (Almeida *et al.* 2006). No que se refere à ferramenta de extração, uma grande vantagem do Corpógrafo é o fato de possibilitar a geração de listas, desde termos simples até termos complexos longos, formados por várias palavras. Além disso, o ambiente mostrou-se extremamente amigável e de fácil manipulação. Entretanto, o grande sucesso alcançado pelo Corpógrafo é, ao mesmo tempo, um de seus maiores inconvenientes, já que a grande demanda de usuários atrelada à limitação do servidor web faz com que o ambiente torne-se instável, muitas vezes inacessível e sempre lento no que se refere ao carregamento dos recursos e o processamento dos resultados[11].

Avaliou-se então o Pacote NSP[12] (*N-gram Statistics Package*), implementado por Ted Pedersen, Satanjeev Banerjee e Amruta Purandare, da Universidade de Minnesota em Duluth, posteriormente eleito para a realização da extração dos candidatos nesta pesquisa.

O pacote NSP é capaz de gerar listas de candidatos a termos (de unigramas a pentagramas) por meio de medidas estatísticas simples (frequência)

---

[10] Disponível em: http://www.linguateca.pt/Corpografo/.

[11] Na época da avaliação, o Corpógrafo era disponibilizado pela Linguateca exclusivamente em sua versão *on-line*, no entanto, atualmente ele está disponível para *download*, podendo ser utilizado diretamente no computador do usuário.

[12] Maiores informações podem ser obtidas em: http://ngram.sourceforge.net/.

ou avançadas (combinatórias), além disso, o NSP pode ser utilizado localmente, ou seja, não necessita de servidores web ou de conexão. Entretanto, o Pacote NSP não é amigável, como os demais programas. Sua operação (sem interface gráfica) por meio de comandos específicos de linguagens de programação demanda significativo empenho do linguista[13].

A partir do corpus de N&N, foram gerados pelo Pacote NSP 587.615 candidatos a termos que, após revisados pelo linguista, foram enviados ao especialista de domínio[14] para validação. Depois da análise do especialista é que o candidato a termo passa de fato a ser tomado como termo. Como resultado deste processo, foram obtidos 3.069 termos. Como apenas *0,52%* dos candidatos foram confirmados como termos, pode-se comprovar a necessidade de se avançar em conhecimento linguístico para sistemas de extração a fim de que se obtenham índices de acerto satisfatórios.

Dada a grande quantidade de termos obtidos, foi adotado um critério estatístico para limitar a quantidade de termos a serem descritos. Optou-se por selecionar apenas 10% dos termos mais frequentes de cada n-grama. Após esta seleção, observou-se ainda a presença de erros (candidatos a termos equivocadamente validados como termos) provenientes do processo de extração, tais erros foram excluídos da lista, novos termos não foram incluídos (em substituição aos erros) para que se respeitasse o critério de frequência. Desta forma, foram submetidos à descrição 295 termos.

3.2 Identificação dos Processos de Formação Lexical e Sistematização dos Dados

Para a descrição dos processos de formação de palavras, após a extração dos termos, organizou-se uma grande tabela[15] no *Microsoft Excel* contendo 9 colunas com o seguinte conteúdo:

- coluna 1: todos os 295 termos;
- coluna 2: processo sintático de derivação prefixal;
- coluna 3: processo sintático de derivação sufixal;

---

[13] É preciso destacar que, para a correta utilização do Pacote NSP, foi fundamental a colaboração dos pesquisadores Profa. Dra. Sandra Maria Aluisio e Daniel Feitosa, ambos do NILC.

[14] O Prof. Dr. Osvaldo Novais de Oliveira Jr é o especialista de domínio integrante do projeto NanoTerm. O pesquisador é professor do Instituto de Física de São Carlos, Universidade de São Paulo e membro fundador do NILC.

[15] Devido às limitações de espaço, essa tabela não figura como apêndice.

- coluna 4: composição subordinativa;
- coluna 5: composição coordenativa;
- coluna 6: composição sintagmática;
- coluna 7: composição por siglas ou acronímica;
- coluna 8: conversão;
- coluna 9: estrangeirismo.

Assim foi possível observar o(s) processo(s) atuando em cada termo como também gerar dados estatísticos e gráficos.

Estabelecido o recorte, passou-se à seleção de cada termo a ser analisado. Esta seleção foi feita manualmente, a fim de se identificar os processos mais característicos, típicos ou relevantes.

4. Descrição

Como apresentado na seção 2, os processos sintáticos são divididos por Alves (2007) em derivação e composição. Os três diferentes tipos de derivação são: prefixal, sufixal e parassintética, mas como esta última não teve qualquer ocorrência, não será apresentada aqui. Já os tipos de composição são: coordenativa, subordinativa, sintagmática e por siglas ou acronímica.

4.1. Derivação

O processo sintático de derivação prefixal ocorre 53 vezes (de um total de 295 itens, recorte selecionado para análise, como já mencionado), muitas vezes integrando composições sintagmáticas. Observem-se alguns exemplos: *absorção, adsorção, biotecnologia, decomposição, dessorção, dielétrico, infravermelho, laser semicondutor, material nanoestruturado, microesfera, moagem reativa, preforma porosa,* etc. Registra-se a presença dos prefixos: *ab-, ad-, bio-, de-, des-, di-, infra-, micro-, nano-, pre-, re-, semi-.*

O processo sintático de derivação sufixal ocorre 79 vezes, muitas vezes integrando composições sintagmáticas. Observem-se alguns exemplos: *acoplamento, amostra de espinélio dopada, anodização, armazenagem de hidrogênio, barreira de potencial, cadeia polimérica, catalisador, condutividade iônica, controle escalar em malha, cristalito, dispositivo, eletrólito, espessura do filme, extrusora de rosca dupla, fóton, retardante de chama, semicondutor, silício poroso,* etc. Registra-se a presença dos sufixos: *-ada, -ado, -agem, -al, -ar, -ção, -dor, -eira, -eto, -ico, -(i)dade, -io, -ito, -ivo, -lito, -mento, -nte, -on, -or, -oso, -ura.*

4.2. Composição

O processo sintático de composição subordinativa ocorre 9 vezes, nos termos: *eletroquímica, equiaxial, espectroscopia, espectroscopia Raman, fotodetector, impedância eletroquímica, litografia, morfologia, precipitador eletrostático.*

O processo sintático de composição coordenativa ocorre 2 vezes, nos termos: *desvio padrão* e *sol-gel.*

O processo sintático de composição sintagmática ocorre 129 vezes. Observem-se alguns exemplos: *amostra de espinélio dopada, área superficial específica, armazenagem de hidrogênio, atividade catalítica, banda de condução, barreira de potencial, cadeia polimérica, campo de stokes, conformação por spray, controle escalar em malha, diâmetro da nanoesfera, difratograma de raios X, eletrólito polimérico gelificado, extrusora de rosca dupla, frequência do laser escravo, microscopia de força atômica, nanofita de SNO, retardante de chama,*etc.

O processo sintático de composição por sigla ou acronímica ocorre 27 vezes. Observem-se alguns exemplos: *AFM, DRX, DSC, filme de PET, imagem de MET, pastilha de SNO, PST, SPIN, TG,* etc.

4.3 Conversão

O processo de conversão ocorre 2 vezes, nos termos: *dielétrico* e *precipitado*, originalmente formas adjetivas que passaram a ser empregadas também como substantivos. Observem-se os exemplos extraídos do corpus:

> *Os micro-defeitos tais como as rugosidades nas interfaces do* ***dielétrico****, micro precipitados e cargas fixas localizadas, sempre estiveram presentes nos dispositivos microeletrônicos...* (Manoel Cesar Valente Lopes, 2000 [gênero Científico]);
>
> *A precipitação de um sistema multicomponente origina os óxidos mistos. O* ***precipitado*** *gerado deve ser filtrado, lavado e calcinado.* (Kírian Pimenta Lopes, 2000 [gênero Científico]).

4.4 Estrangeirismos

Os estrangeirismos ocorrem 18 vezes, na maioria das vezes em siglas: *AFM, conformação por spray, DNA, DSC, INGAP, LASER escravo, LASER mestre, MEMS, MOL, PST, SPIN,* etc.

4.5 Total de Ocorrências dos Processos de Formação de Palavras

Considerando o total de 295 itens léxicos analisados, obtêm-se os seguintes números apresentados na tabela 1.

Processo de formação	**Número de ocorrências**
composição sintagmática	129
derivação sufixal	79
derivação prefixal	53
composição acronímica	26
estrangeirismo	18
composição subordinativa	9
composição coordenativa	2
conversão	2
Total	**318**

Tabela 1: Total de ocorrências dos processos de formação de palavras

Há que se chamar atenção para o fato de que foram analisados 295 itens da terminologia da N&N, mas o total de processos observados é 318. Isso ocorre porque muitas vezes em um único item léxico observam-se mais de um processo, exemplos:

- *amostra de espinélio dopada*: incidem aqui os processos de derivação sufixal (no item *dopada*) e de composição sintagmática.
- *LASER escravo*: neste item, ocorrem os processos de sigla (*LASER*), estrangeirismo (*Light Amplification by Stimulated Emission of Radiation*) e composição sintagmática.

Os números acima podem ser ilustrados no gráfico 1, a seguir.

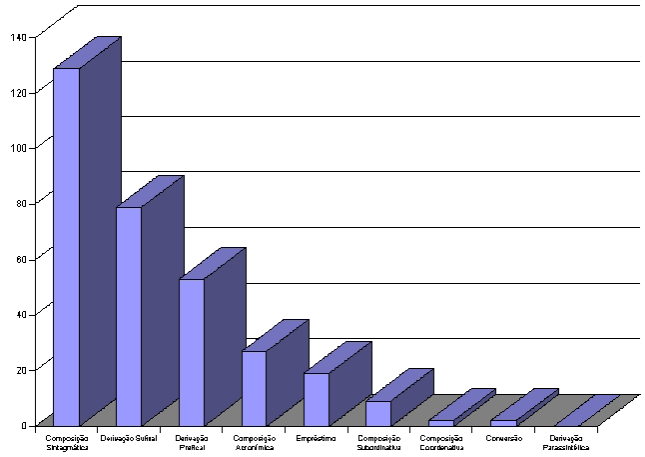

Gráfico 1: Quantidade de ocorrência dos processos de formação de palavras

Agrupando-se os dados pelos tipos de processo, obtêm-se os seguintes totais, apresentados na tabela 2.

Tipo de processo	Número de ocorrências
Processos Sintáticos – Composição	166 ocorrências
Processos Sintáticos – Derivação	132 ocorrências
Empréstimos	18 ocorrências
Conversão	2 ocorrências

Tabela 2: Produtividade dos processos de formação de palavras

Os dados podem ser ilustrados no gráfico 2:

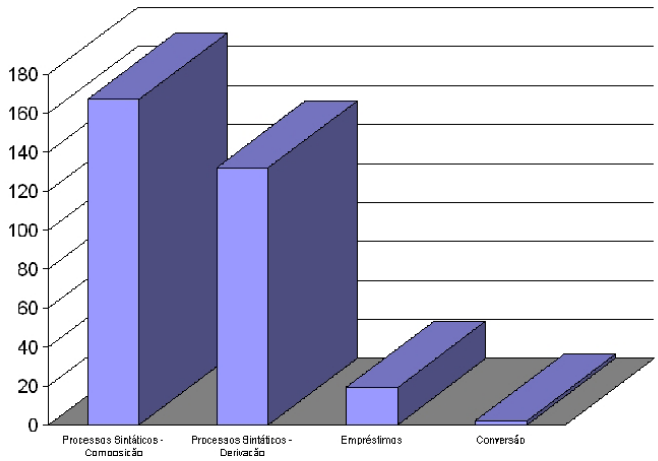

Gráfico 2: Produtividade dos tipos de processos de formação de palavras

5.0 Análise

A partir da descrição morfológica elaborada de acordo com a tipologia de processos de formação de Alves (2007), procedeu-se à análise dos termos selecionados.

5.1. Processos Sintáticos: Derivação

Conforme já apresentado na seção 4, dos 132 casos de derivação, 79 correspondem à sufixação e 53 à prefixação.

Com relação aos casos de derivação sufixal, destacam-se os sufixos: *-ado,; -agem, -ção e -mento; -ico; -or; -dade; -al.*

A sufixação com a partícula *-ado* indica formas verbais nominalizadas, que ocorrem com função adjetiva em n-gramas de duas ou mais partes ou com função substantiva (Almeida & Vale, 2008). Alguns exemplos: *amostra de espinélio dopada, cátodo preparado* e *precipitado.*

Os sufixos *-agem, -ção* e *-mento* indicam que o termo é um processo (Almeida & Vale, 2008). No grupo analisado, correspondem a 30 sufixos de um total de 79. Exemplos podem ser: *armazenagem, dopagem e moagem; anodização, calcinação, caracterização; acoplamento, espalhamento, tunelamento.*

A sufixação com a partícula *-ico* indica um processo de adjetivação, segundo Correia & Lemos (2005). Na terminologia analisada (considerando a variação *-ica*) há 8 ocorrências (5 bigramas, 2 trigramas e 1 unigrama) que

comprovam a afirmação de Correia & Lemos (2005). Exemplos: *cadeia polimérica, campo elétrico* e *condutividade iônica*.

O sufixo *-or* comumente indica um agente, personalizável ou não personalizável, com base em um radical verbal ou nominal, segundo o Dicionário Houaiss da Língua Portuguesa. *Catalisador, fotodetector* e *precipitador* são exemplos deste caso.

O sufixo *-dade* transforma uma base pertencente à classe dos adjetivos em um substantivo, segundo Correia & Lemos (2005). Na terminologia analisada, ocorre 1 caso (2 repetições) de sufixação com esta partícula (*condutividade* e *condutividade iônica*), sendo o substantivo *condutividade* formado a partir do adjetivo "condutivo" acrescido do sufixo *-dade*.

Já sufixo *-al* é, segundo Correia & Lemos (2005), responsável pelo processo de transformação de um substantivo em adjetivo. As ocorrências *barreira de potencial* e *transversal* validam esta afirmação.

No grupo dos casos de derivação prefixal, destacam-se os prefixos: *nano-, micro-, semi-, des-, infra-*. A prefixação com a partícula *nano-* é um dos mais produtivos processos sintáticos de derivação. Como já mencionado anteriormente, esse prefixo significa um bilionésimo da unidade indicada, assim, um nanômetro corresponde a 10^{-9} metros. Em todas as ocorrências analisadas, a presença do prefixo indica que pelo menos uma das dimensões do material se encontra em escala nanométrica. No entanto, a análise de toda a terminologia da N&N com prefixo *nano-* indica que este morfema também é adotado em sentido figurado (em contextos menos formais, como por exemplo, em textos do gênero informativo), referindo-se a dimensões reduzidas, mas não necessariamente em escala nanométrica, como indicaram Almeida & Vale (2008). São exemplos, entre outros, desse fenômeno os itens *nanoartesanato, nanopadronização* e *nanoperiodicidade*.

Almeida & Vale (2008) também apontam para o uso do prefixo ***nano-*** como forma presa (unida ou não por hífen) e como forma livre, exercendo em ambos os casos as funções de substantivo ou adjetivo. No recorte analisado, há uma ocorrência como forma livre e 17 ocorrências como forma presa. A forma livre manifesta-se no corpus tanto na forma de substantivo quanto na forma de adjetivo. As formas presas podem ser divididas entre as que ocorrem com função de substantivo e as que ocorrem com função de adjetivo, como se pode observar a seguir:

- função de substantivo: *nanociência, nanocompósito, nanoescala, nanoesfera, nanoestrutura, nanofio, nanofita, nanofita de sno, nanomaterial, nanômetro, nanopartícula, nanopartícula de sno, nanotecnologia, nanotubo, nanotubo de carbono;*

- função de adjetivo: *nanocristalino, nanométrico, escala nanométrica, filme nanoestruturado, material nanoestruturado.*

Assim como *nano-*, o prefixo *micro-* é utilizado para indicar a dimensão reduzida dos objetos, técnicas, equipamentos e outras bases que prefixa. O recorte terminológico em análise registra o prefixo atuando como forma livre (1 ocorrência) e como forma presa (16 ocorrências). Exemplos: *microeletrônica, microesfera* e *micrografia*.

Diferentemente das partículas *nano-* e *micro-*, o prefixo *semi-* não altera a dimensão do item da base, mas sim sua intensidade, ou seja, nas ocorrências analisadas (*laser semicondutor, material semicondutor, semicondutor*), a capacidade de condução da base é reduzida com o acréscimo do prefixo.

O prefixo *des-* adiciona à base o significado semântico de oposição. Assim, a *dessorção* é o processo oposto de *adsorção* ou *absorção* (todos termos presentes na terminologia).

O prefixo latino *infra-* (*infravermelho*) adiciona à base o significado semântico de localização. *Infravermelho* é, segundo o Novo Dicionário Eletrônico Aurélio versão 5.0, "relativo à parte do espectro situada antes do vermelho", sendo que a posição anterior ao vermelho é dada pelo prefixo que tem o significado de "abaixo". O termo formado pode ocorrer como substantivo ou como adjetivo. Outro exemplo é infraestrutura[16].

5.2 Processos Sintáticos: Composição

Alves (2007) propõe uma divisão que se mostra útil, qual seja, os itens que costumam ser grafados com hífen (mas não necessariamente) constituem os *compostos subordinativos ou coordenativos*, e aqueles que são grafados sem hífen, mas que já se encontram em vias de lexicalização, constituem os *compostos sintagmáticos.*

Como no grupo analisado não há nenhum caso de compostos justapostos com hífen, apenas os compostos aglutinados sem hífen, consideraram-se então os casos de *eletroquímica, equiaxial, espectroscopia, espectroscopia Raman, fotodetector, impedância eletroquímica, litografia, morfologia, precipitador eletrostático* (9 ocorrências) como sendo constituintes do grupo de compostos subordinativos, já que são termos formados por mais de um radical:

- eletr(i/o)- + química > *eletroquímica*

[16] A parte inferior de uma estrutura, segundo o Novo Dicionário Eletrônico Aurélio versão 5.0.

- equi- + ax(i)- + -al > *equiaxial*
- espectr(o)- + -scop- + -ia > *espectroscopia*
- fot(o)- + detector > *fotodetector*
- lit(o)- + -grafia > *litografia*
- morf(o)- + -logia > *morfologia*
- ele(c)tr(o)- + -stat(o)- + -ico > *eletrostático*

E os termos *desvio padrão* e *sol-gel* como os únicos casos de compostos coordenativos no grupo analisado.

Já os itens sintagmáticos, bastante produtivos em terminologias (129 ocorrências), correspondem a termos com a seguinte estrutura:

- N + A: *atividade catalítica, cadeia polimérica, campo claro, campo elétrico, campo escuro, campo magnético, campo óptico, capacidade específica, cátodo preparado, condutividade iônica, constante dielétrica, corrente contínua,* etc.
- N + N: *desvio padrão, efeito brillouin, espalhamento brillouin, espalhamento Raman, espectro Raman, espectroscopia Raman, luz síncrotron, sensor brillouin.*
- N + prep (+ det) + N: *armazenagem de hidrogênio, banda de condução, barreira de potencial, campo de bombeio, comprimento de onda, condição de anodização, conformação por spray, corpo de prova, diâmetro da nanoesfera, espessura do filme, resisténcia4 à corrosão,* etc.
- N + A + A: *área superficial específica, campo elétrico local, eletrólito polimérico gelificado.*
- N + A + prep + N: *controle escalar em malha, microscopia eletrônica de transmissão, microscopia eletrônica de varredura, microscópio eletrônico de transmissão, microscópio eletrônico de varredura.*
- N + prep (+ det) + N + A: *amostra de espinélio dopada, extrusora de rosca dupla, frequência do laser escravo, frequência do laser mestre, método do precursor polimérico, microscopia de força atômica, microscópio de força atômica.*
- N + prep + A + N: *moagem de alta energia.*
- N + prep + N + N: *difração de raios X, difratograma de raios X.*

Esses são os padrões morfossintáticos existentes no grupo de termos analisados. Note-se a grande produtividade das estruturas N + A e N + prep (+ det) + N, o que confirma os estudos terminológicos sobre a produtividade dessas formações em vocabulários especializados.

Outra grande incidência são as siglas ou composições acronímicas (26 ocorrências) que, no grupo analisado, constituem itens autônomos, como *DRX (= Difração de Raios X)* e *DSC (= Dye-sensitized Solar Cell)*, ou integram composições sintagmáticas, como *frequência do LASER (= Light Amplification by Stimulated Emission of Radiation)* e *imagem de MET (=Microscópio Eletrônico de Transmissão)*.

Comprova-se pelos dados obtidos a proposição de Correia (1998):

> Se quisermos comparar a produtividade dos mecanismos de formação de palavras nas terminologias com os da língua corrente, rapidamente verificamos que, no âmbito das linguagens científicas e técnicas, é muito frequente o recurso à composição, quer por temas, quer sintagmática (as também chamadas lexias complexas, na terminologia de Pottier), apresentando as unidades lexicalizadas, muitas vezes, uma dimensão bastante superior às da língua corrente (Correia, 1998: 70).

Em terminologias, Alves (2007) também aponta para a alta frequência de itens léxicos sintagmáticos, que "resultam, nesses casos, de uma indecisão em relação à designação de uma nova noção. A denominação em forma de sintagma pode vir a ser substituída por uma única base ou o sintagma pode chegar a cristalizar-se e inserir-se no léxico da língua" (Alves, 2007: 54).

5.3 Empréstimos

Embora seja esperado um número alto de empréstimos, sobretudo quando se trata de uma área que tem seu expoente de pesquisa e inovação em países de língua inglesa, como é o caso das áreas de N&N, não há uma ocorrência significativa, já que foram encontrados 18 casos de empréstimos no grupo analisado, dos quais 6 constituem repetições como *frequência do laser, frequência do laser escravo, frequência do laser mestre, laser, laser escravo, laser mestre, laser semicondutor*. De maneira que há, na realidade, 12 casos de empréstimos.

Via de regra, esses empréstimos constituem itens autônomos, como *chip* (= abrev. de microchip, do pref. ingl. micro- (v. micr(o)-) e ingl. chip, lit., 'lasca', 'fragmento'), ou integram composições sintagmáticas e/ou siglas, como campo de *stokes* e *AFM* (=Atomic Force Microscope).

Esses resultados confirmam a pesquisa[17] de Alves (disponível na Base de Neologismos do Português Brasileiro Contemporâneo), em que a autora

[17] "O projeto tem o objetivo geral de coletar e analisar a neologia do português contemporâneo do Brasil, observada em um corpus jornalístico, fornecendo subsídios para o estudo

demonstra que os estrangeirismos correspondem a 17% das unidades lexicais neológicas, contra 38% de casos de derivação e 37% de composição. O que comprova que, mesmo nessas terminologias, os processos mais produtivos de formação de palavras no português são os autóctones.

6. Considerações finais

Esta pesquisa teve como objetivo aprofundar, do ponto de vista morfolexical, o estudo de um grupo de termos que compõem a terminologia da N&N, obtida no Projeto NanoTerm, produto das etapas de extração, limpeza das listas de candidatos, lematização e validação. Em todas essas etapas iniciais, foram seguidos os procedimentos metodológicos sugeridos pela Linguística de Corpus e pela Terminologia.

Para esta pesquisa, fez-se necessário delimitar um conjunto terminológico para descrição e outro conjunto ainda mais limitado para análise, de maneira a pôr em evidência aspectos relevantes que seriam úteis para o aprimoramento de ferramentas de extração de termos baseadas em conhecimento linguístico.

Na sequência, realizaram-se os procedimentos de identificação dos processos de formação de palavras e de sistematização dos dados a partir da tipologia proposta por Alves (2007).

Efetuou-se, então, a descrição morfológica de 295 termos. Os resultados obtidos confirmaram os pontos teóricos levantados na seção 2 como também aqueles que são consenso nos estudos dos processos de formação de palavras em vocabulários especializados. Nesse sentido, cumpre destacar os seguintes padrões morfolexicais observados:

---

da evolução do léxico português (variante brasileira) e para a elaboração de repertórios de unidades lexicais neológicas. Para a constituição da Base, foi utilizado um corpus constituído pelos jornais Folha de S. Paulo e O Globo e pelas revistas IstoÉ e Veja a partir de 01-93, observado segundo um sistema de amostragem (um veículo por semana). Nesses veículos, foram coletados neologismos de caráter vernáculo (derivação, composição, truncação, transferência semântica...) e de caráter estrangeiro. Foram consideradas como neológicas as unidades lexicais que não estão incluídas em um corpus de exclusão constituído por um conjunto de dicionários da língua geral. A Base conta, atualmente, com 13.568 unidades lexicais neológicas, que podem apresentar uma, duas ou mais ocorrências. O número de ocorrências da Base corresponde, atualmente, a 24.628." Para mais informações, consultar http://www.fflch.usp.br/dlcv/neo/.

Ø com relação à sufixação, destacam-se os seguintes morfemas: *-ado; -agem, -ção e -mento; -ico; -or; -dade; -al;* os quais poderiam gerar os seguintes padrões de busca, nos quais X representa a base léxica:

- X + *-ado*
- X + *-agem*
- X + *-ção*
- X + *-mento*
- X + *-ico*
- X + *-or*
- X + *-dade*
- X + *-al*

Ø com relação à prefixação, destacam-se os seguintes morfemas: *nano-, micro-, semi-, des-, infra-,* os quais poderiam gerar os seguintes padrões de busca, nos quais X representa a base léxica:

- *nano-* + X
- *micro-* + X
- *semi-* + X
- *des-* + X
- *infra-* + X

Ø com relação aos compostos subordinativos, embora sendo menos frequentes no grupo analisado, constatou-se grande incidência de morfemas greco-latinos, como se pôde observar nas formações:

- eletr(i/o)- + química > *eletroquímica*
- equi- + ax(i)- + -al > *equiaxial*
- espectr(o)- + -scop- + -ia > *espectroscopia*
- fot(o)- + detector > *fotodetector*
- lit(o)- + -grafia > *litografia*
- morf(o)- + -logia > *morfologia*
- ele(c)tr(o)- + -stat(o)- + -ico > *eletrostático*

Ø no que tange aos compostos coordenativos, observaram-se apenas as ocorrências *desvio padrão* e *sol-gel* no grupo selecionado para análise, demonstrando que não é um processo produtivo;

Ø no que se refere aos compostos sintagmáticos, constatou-se a grande produtividade das estruturas N + A e N + prep (+ det) + N. Em menor ocorrência, mas não menos importante, têm-se as estruturas: N + N, N + prep (+ det) + N + A, N + A + prep + N, N + A + A, N + prep + N + N e, finalmente, N + prep + A + N;

- Ø no que diz respeito às siglas ou formações acronímicas (26 ocorrências), observou-se que elas constituem itens autônomos ou integram composições sintagmáticas. Constatou-se, também, que correspondem a formas expandidas em português e em inglês, como em *DRX (= Difração de Raios X)* e *DSC (= Dye-sensitized Solar Cell)*;
- Ø com relação aos empréstimos, observou-se baixa frequência, entretanto, 100% dos casos (18) são provenientes da língua inglesa.

Integrando o item denominado *outros processos* por Alves (2007), está a conversão, que teve baixíssima ocorrência no grupo analisado, apenas 2 casos: *dielétrico* e *precipitado*, originalmente adjetivos que passaram a ser empregados como substantivos. Nota-se, pois, que esse processo, por não ser frequente, não parece ser útil para uma busca automatizada de termos num corpus.

Outro aspecto que foi observado diz respeito à densidade terminológica nos distintos gêneros textuais que compõem o corpus. O estudo revelou que a maior quantidade de palavras em um gênero não significa que neste mesmo gênero estará presente uma maior quantidade de termos. Os números demonstraram que o gênero *científico* apresenta a maior quantidade de palavras e também a maior quantidade de termos; no entanto, o gênero *informativo*, mesmo possuindo mais palavras do que o gênero *científico de divulgação*, apresenta menor quantidade de termos em comparação com o *científico de divulgação*; assim como o gênero *técnico-administrativo*, que tem mais palavras do que o gênero *outros*, apresenta menor quantidade de termos em comparação com *outros*.

Esses números confirmam a importância de se compilar um corpus que represente distintos gêneros textuais, mesmo em se tratando de uma pesquisa terminológica.

Concluído este trabalho, considera-se que os objetivos almejados tenham sido cumpridos. Entretanto, enumeram-se temas que poderão ser objeto de pesquisas futuras, tais como:

- ampliar a descrição deste vocabulário, inserindo mais termos de modo a atestar se os percentuais obtidos no que se refere aos processos de formação de palavras mais produtivos se mantêm;
- inserir os padrões morfolexicais observados no pacote NSP (programa utilizado para fazer a extração) e observar o quanto esse conhecimento linguístico melhora os resultados no processo de extração de candidatos a termos;

- replicar esse mesmo modelo de análise para outras terminologias, de maneira a observar se esses padrões se repetem ou se cada terminologia tem suas características próprias no que tange aos aspectos morfolexicais.

Referências bibliográficas

ALMEIDA, G. M. B.; OLIVEIRA, L. H. M.; ALUÍSIO, S. M. 2006. A Terminologia na era da Informática. Ciência e Cultura, Campinas (SP), v. 58, n. 2, p. 42-45. Disponível em: http://cienciaecultura.bvs.br/scielo.php?script=sci_arttext&pid=S0009-67252006000200016&lng= pt&nrm=iso. Acesso em 17/01/2010.

_____; VALE, O. A. 2008. Do texto ao termo: interação entre Terminologia, Morfologia e Linguística de Corpus na extração semi-automática de termos. In: ISQUERDO, A. N.; FINATTO, M. J. (orgs.). *As ciências do Léxico: Lexicologia, Lexicografia e Terminologia*. 1ª. ed. Campo Grande: Editora da UFMS, v. IV, p. 483-499.

ALVES, I. M. 2007. *Neologismo:* criação lexical. 3ª. ed. São Paulo: Ática.

CORREIA, M. 1998. Neologia e terminologia. In: MATEUS, M. H.; CORREIA, M. (coords.) *Terminologia: questões teóricas, métodos e projectos*. Cursos da Arrábida. Lisboa: Publicações Europa-América Ltda. p. 59-74.

_____; LEMOS, L. S. P. 2005. *Inovação lexical em português*. Lisboa: Edições Colibri.

OLIVEIRA, L. P. 2007. *A terminologia da Genética Molecular: constituição morfológica e estruturação semântica*. Dissertação (Mestrado). Universidade de São Paulo.

SENAI. 2004. *Nanotecnologias*. Brasília: SENAI.

TOMA, H. E.; ARAKI K. 2005. O *gigantesco e promissor mundo do muito pequeno*. Disponível em: http://cienciahoje.uol.com.br/materia/view/3440. Acesso em 18 de abril de 2007.