

# **Criação e Exploração de um Corpus Bilingue (IN-PT) sobre Células Estaminais**

**TERESA CACILDA GONÇALVES FERREIRA**

---

**Dissertação de Mestrado em Terminologia e Gestão da  
Informação de Especialidade**

MARÇO, 2010



Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Terminologia e Gestão da Informação de Especialidade, realizada sob a orientação científica da Professora Doutora Rute Costa.

## DECLARAÇÃO

Declaro que esta tese é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e na bibliografia.

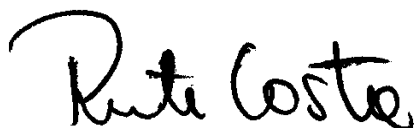
O candidato,

Teresa Ferreira

Lisboa, 31 de Março de 2010

Declaro que esta Tese se encontra em condições de ser apresentada a provas públicas.

A Orientadora,



Rute Costa

---

Lisboa, 29 de Março de 2010

**CRIAÇÃO E EXPLORAÇÃO DE UM CORPUS BILINGUE (IN-PT) SOBRE  
CÉLULAS ESTAMINAIS**

**CREATION AND EXPLOITATION OF A BILINGUAL (EN-PT) CORPUS ON  
STEM CELLS**

**TERESA GONÇALVES FERREIRA**

**PALAVRAS-CHAVE:** corpora electrónicos, estudos de tradução, tecnologias de informação e comunicação, células estaminais, corpus especializado bilingue, critérios de selecção de textos, terminologia, extracção terminológica, ferramentas de análise linguística.

**KEYWORDS:** electronic corpora, translation studies, information and communication technologies, stem cells, electronic bilingual corpus, text selection criteria, terminology, terminology extraction, tools for linguistic analysis.

**Resumo**

Apesar da compilação de corpus ser uma prática comum desde os séculos passados, na actualidade quando se fala em corpus fala-se necessariamente em corpus electrónicos, uma vez que é impensável o divórcio entre a criação e exploração de corpora e o avanço das tecnologias de informação e comunicação. O objectivo deste trabalho é compilar um corpus bilingue de língua de especialidade constituído por textos correspondentes a esta área de conhecimento específico: a terapia celular através de células estaminais, - pondo em prática algumas ferramentas informáticas para avaliar

o seu grau de eficácia no campo da tradução -, que por sua vez servirá para atingir um objectivo final: elaborar um glossário bilingue sobre células estaminais. Recorrer-se-á ao apoio das ferramentas electrónicas de análise linguística para exploração de corpora: O Corpógrafo e o Wordsmith Tools. Ilustrar-se-á em pormenor os passos que foram efectuados em cada uma das áreas do Corpógrafo e as tarefas complementares operadas no Wordsmith Tools. Analisar-se-á as vantagens das duas ferramentas e o interesse da sua complementaridade. Comentar-se-á o número de termos que se incluíram nas bases de dados e no glossário, as relações dos termos mais significativos e algumas ocorrências.

### **Abstract**

In spite of compiling corpora since the last two centuries, people cannot deny that nowadays the word corpora is connected to electronic corpora, as it is unthinkable to elaborate and exploit corpora without the advances of information and communication technologies. This work presents the criteria for the selection of the texts that will be used to elaborate an electronic bilingual corpus (English-Portuguese) related to a medical translation subfield. To elaborate a bilingual glossary on Stem Cell is our main purpose. The Corpógrafo and the Wordsmith Tools will be the electronic tools for linguistic analysis used to support our tasks of corpora exploitation. It will be explained in detail the steps that were taken in each one of the areas of the Corpógrafo and the complementary tasks operated in the Wordsmith Tools. The advantages of these two tools will be analysed as well as the interest of their complementarities. The number of terms included in the data bases and in the glossary will be commented, and the relations of the most significant terms and some of the occurrences will also be object of our attention.

## ÍNDICE

<b>INTRODUÇÃO.....</b>	<b>1</b>
<b>I. CORPUS VERSUS TRADUÇÃO.....</b>	<b>7</b>
I.1. CORPUS SINÓNIMO DE CORPUS ELECTRÓNICO .....	7
I.2. CRIAÇÃO E ANÁLISE DE CORPORA ELECTRÓNICOS .....	9
I.3. TIPOS DE CORPORA ELECTRÓNICOS.....	12
I.4. A UNIDADE MÍNIMA DE TRADUÇÃO.....	13
<b>II. A IDENTIDADE DA TERMINOLOGIA .....</b>	<b>18</b>
II.1. O CONCEITO E O TERMO QUE O DESIGNA.....	18
<b>III. METODOLOGIA .....</b>	<b>27</b>
III.1. CRIAÇÃO DO CORPUS ESPECIALIZADO BILINGUE.....	27
<b>IV. RESULTADOS.....</b>	<b>35</b>
IV.1. DESCRIÇÃO DOS RESULTADOS.....	35
<b>CONCLUSÃO.....</b>	<b>59</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>63</b>
I. REFERÊNCIAS CITADAS .....	63
II. FONTES TEXTUAIS DO CORPUS.....	66

### APÊNDICE : GLOSSÁRIO BILINGUE

**Caminhante, são teus rastos  
o caminho, e nada mais;  
caminhante, não há caminho,  
faz-se caminho ao andar.  
Ao andar se faz o caminho,  
e ao voltar para trás  
vê-se a senda que jamais  
se há-de voltar a pisar.  
Caminhante, não há caminho  
Somente sulcos no mar.**

**António Machado**

## Introdução

“An open mind is the best guide in linguistics, as in research in general and indeed in life itself.” (Stig Johansson, 1991: 6) Sirvo-me desta afirmação como ponto de partida da minha dissertação. Quando se inicia um projecto, este acaba por tomar conta de nós, ganha vida própria e guia-nos por rumos que pouco a pouco vamos desbravando. Levar dentro da *bagagem* uma mente aberta, uma forte determinação e um pouco de boa sorte à mistura é o suficiente para se penetrar em um “admirável mundo novo”.

Nos últimos anos, a leitura de inúmeros artigos em jornais e revistas, livros, panfletos publicitários e outros documentos sobre as células estaminais e suas aplicações terapêuticas tem sido uma constante. Nessa perspectiva, uma vez escolhido o domínio da tradução científica – principalmente a tradução médico-farmacêutica como aquela área que mais nos fascina e se quer explorar -, a temática específica do nosso projecto também não foi difícil. Pode dizer-se que as células estaminais são a nova fronteira da medicina e adivinha-se um futuro cheio de novas possibilidades devido às suas extraordinárias características. Estas células são especiais por duas razões fundamentais: a sua capacidade de auto-renovação e a sua indiferenciação, isto é, por serem células que não desempenham uma função específica no corpo e, assim, cujo destino ainda não está “definido”, podem programar-se em vários tipos de células existentes no organismo. Nesse sentido, os cientistas estão entusiasmados com a possibilidade destas poderem curar várias doenças crónicas e degenerativas, às quais não se consegue ainda dar resposta. E enquanto seres humanos que amamos a vida nós não queremos morrer, apesar de sabermos que é a única verdade irrefutável da nossa existência. Assim, se a investigação médica nos prolongar a estadia com qualidade tanto melhor.

Este trabalho específico pretende oferecer uma amostra representativa daquilo que representará a Dissertação de Doutoramento. O objectivo é compilar um corpus bilingue de língua de especialidade constituído por textos correspondentes a esta área de conhecimento específico: a terapia celular através de células estaminais, - pondo em prática algumas ferramentas informáticas para avaliar o seu grau de eficácia no campo

da tradução -, que por sua vez servirá para atingir um objectivo final: elaborar um glossário bilingue sobre células estaminais. Nesta fase, abordar-se-á as questões teóricas que se consideram prévias e necessárias para a elaboração e análise de um corpus.

Com a ajuda de um outro *milagre*, a Internet e as ferramentas informáticas, poderemos obter resultados importantes. Não há dúvida que a integração da Internet em qualquer ambiente laboral facilitou em grande escala muitas das tarefas que é necessário levar a cabo no âmbito do nosso trabalho. O campo da tradução não é uma excepção a este avanço. Quem ambiciona ser um tradutor especializado não pode, de modo algum, descurar os avanços que as tecnologias da informação e da comunicação têm sofrido nos últimos anos. Se há alguém que beneficiou com esta nova panorâmica, esse alguém foram os tradutores. Se não todos, pelo menos aqueles que perceberam que é imperativo ser actual e absorver o que de melhor a tecnologia tem para oferecer. É, portanto, crucial saber dispor das múltiplas ferramentas, como processadores e editores de texto, correctores ortográficos, tradutores automáticos, corpus linguísticos, dicionários, enciclopédias, entre outros, e aceder a qualquer tipo de documento, independentemente do lugar onde nos encontremos, simplesmente dispondo da ligação à Internet. Hoje em dia, a grande novidade é a “Web 2.0”. Já é possível ter acesso a todas estas ferramentas, sem ser necessário instalar nenhuma destas aplicações no computador.

A nova escola, ou disciplina, denominada Estudos de Tradução surge em meados dos anos setenta com o objectivo de formular uma teoria da tradução e de se fixar no estudo quase exclusivo dos processos de tradução. A partir do nascimento dos Estudos de Tradução com os trabalhos de Lefevere (1998), James S. Holmes (1988 [1972]), Gideon Toury (1995), Itamar Even-Zohar (1978), Chesterman (1997) e Catford (1965) critica-se o estado de apatia dos estudos de tradução, o seu divórcio das outras disciplinas, a sua concentração exclusiva nos estudos literários e a sua falta de critérios definidos e de estratégias metodológicas. Como se evidenciará no **Capítulo I** deste trabalho, segundo os autores mencionados em cima, o papel da tradução é cumprir uma função sócio-cultural dentro da comunidade receptora, vincando-se a necessidade de haver uma perspectiva interdisciplinar, de se produzir hipóteses e de se trabalhar para comprovar a sua validade. Exclui-se os pressupostos teóricos não comprováveis na prática do processo tradutivo. O texto traduzido é encarado como uma nova criação e um acto de comunicação única e renovada.

Parte-se para o **Capítulo II** com a convicção que qualquer trabalho terminológico deve basear-se na análise e estruturação dos conceitos e das suas relações, assim como na análise e estruturação dos respectivos termos e relações que reflectem os primeiros, uma vez que os elementos base da Terminologia são o conceito e o termo que o designa e as relações que se estabelecem entre eles. É, portanto, inegável que a identidade da terminologia é a unidade lexical especializada, “El objeto de la terminología, como hemos dicho, es la unidad terminológica” (Cabré, 2000). Apesar do objecto primordial de estudo da Terminologia ser o termo que designa o conceito, sobretudo o conjunto de termos técnicos e ou científicos, o seu entendimento e tratamento não é consensual na comunidade científica. Leia-se comunidade científica como o conjunto de pessoas acreditadas para veicular opiniões e produzir investigações sólidas e credíveis, em uma área específica que é objecto da sua formação. Se questionarmos o senso comum sobre o que é um termo, a resposta é imediata, termo é o mesmo que palavra, vocábulo. Contudo, entre a comunidade de terminólogos, linguistas, filósofos a simplicidade desaparece e o significado de termo é tudo menos evidente. Para alguns, termo é a denominação de uma noção unívoca, inequívoca, admitindo uma única interpretação. Para outros, os termos são entidades variantes, porque funcionam na diversidade da língua.

A diversidade de abordagens teóricas e metodológicas que rodeiam a Terminologia tem por base várias perspectivas distintas, todavia concentrar-nos-emos em duas: a perspectiva cognitiva e a perspectiva linguística. Nesta perspectiva, convém apresentar as particularidades inerentes às teorias da terminologia. Este trabalho não tem a pretensão de valorizar nenhuma das abordagens apresentadas, ou destacar este ou aquele enfoque como o mais apropriado de entre as diferentes correntes terminológicas. Poderá, no entanto, dar indícios e apresentar argumentos de que o interesse poderá assentar em abordagens diferentes. Não há necessariamente duas perspectivas distintas, uma conceptual (que tem por finalidade representar o conhecimento de um domínio através de conceitos estruturados e as relações que os unem) e outra linguística, podendo haver complementaridade de metodologias. Como afirma Costa - “the methodological coming and going present at different stages of the terminology workflow is what finally unchains the acknowledge complementarity of the theoretical approaches” (2005:116). O importante é definir que critérios foram estabelecidos e se

estes foram os mais adequados aos objectivos e necessidades daquele projecto único e específico.

O alargamento de horizontes a que se tem assistido ultimamente no âmbito da terminologia não pode ser descurado. A rápida evolução e transformação da ciência terminológica é um facto observável nos últimos anos como consequência do rápido e revolucionário avanço das tecnologias de informação e comunicação e da sua abertura a novos modelos: a pragmática, a perspectiva sócioterminológica, as metodologias em terminologia textual, o tratamento automático de corpora de especialidade. “La linguistique de corpus a imposé une reforme des méthodes et des conceptions en raison du développement de la gestion informatisée des écrits et de l’apparition de nouveaux outils langagiers” (Gaudin, 2005:3). O termo, o conceito, a fraseologia especializada, a definição terminológica e o texto especializado integraram o seu universo de interesses. A Terminologia assumiu-se, assim, como uma ciência interdisciplinar, uma vez que vários campos do saber como a Linguística, a Lexicografia, as Ciências Sociais são chamados a dialogar. A estreita ligação entre terminologia e cultura é, também, um facto crucial. “A pertença social a uma língua e às suas variações é parte inseparável da identidade do falante. Língua e fala não se opõem, sociedade e indivíduo também não” (Belo, 2001:195).

O estabelecimento de uma metodologia prática para criar e analisar o corpus será a preocupação do **Capítulo III**. Apresentar-se-á os critérios de selecção dos textos que farão parte do corpus bilingue e tecer-se-á comentários acerca das suas especificidades. Neste trabalho, concentrar-nos-emos na elaboração de um corpus bilingue virtual e, nessa perspectiva, os textos que o compõem estão originalmente em formato electrónico. A selecção textual será feita atendendo às nossas necessidades e interesses particulares que se prendem com a extracção terminológica no âmbito do subdomínio biomédico já mencionado. Qualquer domínio de especialidade tem necessariamente novos termos a defini-lo. Os artigos científicos estão recheados de termos específicos. A selecção de textos elaborados por uma comunidade científica acreditada é imperativa, se se quer obter os artigos mais ricos em termos actuais e adequados. Os corpora representarão um período cronológico relativamente curto para garantir a actualidade. Uma vez que serão recolhidos tanto textos completos como fragmentos, os corpora serão mistos. A estratégia será compilar esses artigos e “limpá-los”, retirando

cabeçalhos, tabelas, imagens, figuras, agradecimentos, referências bibliográficas, informação sobre os autores e o que quer que seja necessário para isolar os termos relevantes no âmbito do objecto de estudo. A temática do nosso trabalho exige-nos rigor científico no tratamento de dados. Nessa ordem de pensamento, o *Corpógrafo* e o *Wordsmith Tools* foram as ferramentas informáticas seleccionadas para apoiar as tarefas que se descreveram em cima. O *Corpógrafo* é uma plataforma destinada à análise de corpora com o objectivo de produzir diversos recursos linguísticos em suporte informático. Através dela será possível compilar e organizar o nosso corpus pessoal, bem como classificar e categorizar os textos seleccionados e estudar, pesquisar e extrair os conhecimentos obtidos. O *Corpógrafo* será muito útil, já que integra num único ambiente de trabalho várias ferramentas (edição e pré-processamento de ficheiros, pesquisa, gestão e organização do conhecimento) que habitualmente se encontram dispersas. O *Wordsmith Tools*, escrito por Mike Scott, publicado pela Oxford University Press, é um programa com dez anos de existência (desde o seu lançamento em pequena escala) e que já vai na sua quarta versão. Para além de flexível e fácil de usar, o programa coloca à disposição do investigador uma série de recursos muito úteis na análise de vários aspectos da linguagem. O dito programa proporciona uma forma metódica de utilizar determinadas ferramentas informáticas para identificar ou detectar os candidatos a termo, proporcionando mais dados sobre os ditos termos, ajudando a compilar e a elaborar as suas definições e oferecendo exemplos contextuais. Uma vez que em muitos aspectos são idênticos, optar-se-á por se usar principalmente o *Corpógrafo*, apelando-se ao *Wordsmith Tools* sempre que se considere necessário complementar tarefas realizadas. No **Capítulo IV**, analisam-se os resultados com os quais se pretende elaborar um glossário bilingue sobre células estaminais que responda positivamente na procura de vocabulário específico e informação fiável e organizada por parte de tradutores, intérpretes, redactores científicos ou outro tipo de pessoas interessadas em células estaminais.

Produzir-se-á, assim, um projecto em construção, o que significa que não é um trabalho acabado, fechado, mas sim um que se pretende expandir, partindo do pressuposto objectivo que é necessário tratar um projecto em geral e uma língua em particular como uma entidade dinâmica. Neste sentido, esta abordagem implica reformulações constantes e projectos em constante expansão, acompanhando todas as

actualizações linguísticas e terminológicas que decorrerão dos avanços científicos efectuados nesta área.

## **I. Corpus versus Tradução**

Apesar da compilação de corpus ser uma prática comum desde os séculos passados, na actualidade quando se fala em corpus fala-se necessariamente em corpus electrónicos, uma vez que é impensável o divórcio entre a criação e exploração de corpora e o avanço das tecnologias de informação e comunicação. Os computadores cada vez mais económicos, acessíveis, práticos e intuitivos favorecem essa aproximação e, se a isso juntarmos a ligação à Internet, é possível percorrer o mundo sem deixar a cadeira onde se está sentado. Nas palavras de Frank Austermuhl (2001:1), “perhaps more than other professionals, translators are feeling long-term changes brought about by the information age.” Assim, a pergunta pode um tradutor ser bem sucedido sem usar ferramentas electrónicas, só pode dar origem a uma resposta: “For translators there is no longer any question of whether or not to use computers and networks. The use of information and communication technology (ICT) is *a fait accompli* in the lives of today’s language professionals” (2001:7)

No presente capítulo, abordar-se-á todas estas questões que se consideram prévias e necessárias para a elaboração de um corpus.

A investigação que aqui se apresenta tem como objecto de estudo a terminologia das células estaminais pertencente ao domínio médico.

### **I.1. Corpus Sinónimo de Corpus Electrónico**

Para quem não está familiarizado com a palavra *corpus/ corpora*, ela vem do latim e significa compilação de documentos ou informações relativos a uma disciplina ou tema. A leitura atenta dos artigos “Localización de Recursos y Compilación de Corpus via Internet: Aplicaciones para la Didáctica de la Traducción Médica Especializada” (2004) e “Traducir com Corpus: de la teoría a la práctica” (2002), de Gloria Corpas Pastor, revelou-se fundamental, uma vez que os artigos se ocupam das possibilidades que a Internet oferece para a localização de recursos e a compilação de corpus com vista à tradução de textos médicos. A autora descreve passo a passo os

procedimentos e as estratégias que se podem seguir para criar e utilizar de forma efectiva um corpus *ad hoc*. No primeiro momento de abordagem ao estudo dos corpora, esta orientação foi preciosa. Muitas outras obras foram lidas desde essa altura, todas elas apontando para a criação e exploração dos corpora como uma linha de trabalho que contém resultados de especial interesse para investigadores das áreas da linguística e da tradução.

Muitas são as pessoas que estão convictas que a compilação de corpus é um fenómeno novo, popularizado pela massificação do acesso aos computadores pessoais a partir da década de oitenta. De facto, na actualidade quando se fala de corpus fala-se necessariamente de corpora electrónicos ou informatizados, uma vez que se compõem de textos electrónicos – uma sucessão de caracteres ASCII, bytes e impulsos electrónicos, manipuláveis através de programas informáticos. Estes textos de carácter electrónico mudaram, completamente, a ideia tradicional de texto: entidade física tocável, com páginas que se sucedem num contínuo informativo. Assim, é necessário compreender que a informação está cada vez menos ligada ao objecto físico que a contém (seja um livro, um jornal, um documento, um ecrã de computador). O avanço tecnológico cria novas necessidades e altera paradigmas estabelecidos ao longo do tempo, acelerando e auxiliando esta transição. Ao mesmo tempo, o desenvolvimento no campo da informática e o aparecimento do mercado de computadores cada vez mais práticos e económicos, possibilitou a automatização das ciências da documentação e tornou o acesso à informação mais democrático. Contudo, deve acrescentar-se que considerando que a acessibilidade à informação pressupõe o uso de estratégias de tratamento da informação, no intuito de minimizar as dificuldades com que o usuário se confronta, para encontrar e entender a informação disponibilizada, pode dizer-se que todos os indivíduos têm acesso à informação, mas nem todos têm acessibilidade à mesma. Nesta perspectiva, os rígidos bancos de dados hierárquicos convencionais têm vindo a ser substituídos por um mundo de informação dinâmica, em permanente mudança e expansão, tornando os operadores informáticos cada vez mais próximos do vocabulário e das necessidades dos usuários.

Ao contrário do que muitos possam pensar, existiam corpora antes da era dos computadores, mas estes eram compilados por investigadores individuais para serem usados em projectos pessoais. A compilação de corpus constituiu-se uma prática

comum desde os séculos passados para as disciplinas como a lexicografia e o ensino das línguas. A título de exemplo trata-se de colecções de textos de línguas mortas (latim, grego, sânscrito) recompilados com o objectivo de estudar as línguas. O projecto inicial do *Oxford Dictionary* de Murray (1880) consistiu no estudo de um corpus formado por milhões de fichas de papel onde estão patentes exemplos reais, recolhidos à mão, dos diferentes usos e contextos de todas as palavras da língua inglesa. Na Alemanha, Kading analisou manualmente no ano de 1897 um corpus de onze milhões de palavras para estudar combinações de letras e sequências de letras em Alemão. É, todavia, inegável que o novo contexto tecnológico e social contribuiu de maneira decisiva para o ressurgimento e fortalecimento da investigação linguística baseada no uso de corpora. Assim, com o aparecimento das novas tecnologias de informação e comunicação (TIC), a sua utilização tem vindo a generalizar-se nas áreas de Linguística, do Processamento de Linguagem Natural e da Tradução. A propósito desta associação do uso de corpora com o aparecimento dos computadores e das ferramentas informáticas, Chelo Vargas Sierra, Professor do Departamento de Filologia Inglesa da Universidade de Alicante, declara no artigo “El Proceso Terminográfico Multilingüe Con *Wordsmith Tools*” (2006: 85): “La vinculación del ordenador a la compilación y creación de grandes corpóra ha hecho que hoy en día el término corpus contenga en sí la característica de electrónico y, por ende, analizable computacionalmente.”

## **I.2. Criação e Análise de Corpora Electrónicos**

É compreensível que as pressões e as exigências do mundo de hoje obriguem os profissionais de qualquer área e, neste caso particular, os tradutores a serem versáteis e adaptáveis às novas solicitações do mercado de trabalho cada vez mais competitivo, isto é, é necessário aprender a documentar-se de forma exaustiva sobre uma grande variedade de temas específicos, ser capaz de especializar-se em temas muito concretos em um breve espaço de tempo e reunir toda a documentação disponível sobre um tema, o que só será possível com o uso proficiente dos recursos informáticos e telemáticos. O século XXI não se compadece com a escassez de tempo; tudo acontece a um ritmo e uma velocidade alucinantes. Assim, um dos requisitos básicos do tradutor profissional

figura a necessidade de possuir bons conhecimentos de informática a nível do usuário, conhecer as tecnologias de informação (TIC) e saber manejar os programas de gestão terminológica e os sistemas de tradução por computador. A Internet é, assim, o recurso documental por excelência do tradutor, uma vez que é o canal de comunicação mais rápido e mais eficaz através do qual se acede a grande parte da produção científica e divulgativa do planeta. A quantidade de informação é tanta e tão abrangente que a questão não é encontrar o que se procura, a questão deve colocar-se de uma outra forma: como encontrar o que procuramos de forma rápida e fiável?

A possibilidade de aceder a um sem número de textos através da Internet e o desenvolvimento crescente de ferramentas informáticas específicas facilitaram a criação e a análise de corpora electrónicos. Há sinais que estamos apenas no início da era do estudo de textos informáticos. A linguística do corpus está a converter-se numa metodologia de investigação indispensável. Com um custo de reduzido, o computador tornou possível os estudos quantitativos que eram morosos, difíceis, ou impossíveis, de levar a cabo no passado. Contudo, a análise quantitativa é apenas um dos aspectos da potencialidade do trabalho informático. Como aponta Graeme Kennedy (1998: 5), na obra *An Introduction to Corpus Linguistics*, “Corpus linguistics is thus now inextricably linked to the computer which has introduced incredible speed, total accountability, accurate replicability, statistical reliability and the ability to handle huge amounts of data.” Assim, os corpora electrónicos caracterizam-se pela grande capacidade de sistematização no tratamento de grandes volumes de textos, substituindo o entediante trabalho manual de antigamente. Se o corpus está bem elaborado, os dados que dele se podem extrair são representativos de um determinado período, tipo textual, autor, subdomínio médico, etc. “Corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language” (Eagles, 1996). É evidente que quanto mais representativo for o corpus mais fiáveis serão os dados dele retirados. Para se trabalhar adequadamente com os corpora informáticos tudo o que se necessita é de software adequado. Há quem aponte a necessidade de uma escala mais abrangente de recursos electrónicos para corpora, mas é inegável que a evolução das tecnologias de tratamento de informação originou um crescente número de trabalhos relacionados com corpora textuais. Os corpora informáticos podem ser facilmente partilhados e, assim, podem ser sujeitos a estudos

mais alargados. Todavia, é abundante a construção dos corpora mais específicos, para suporte da pesquisa terminológica temporária, de actividades de tradução e mesmo de actividades de avaliação. Estes corpora, utilizados em tarefas muito específicas, são utilizados por um curto período de tempo, servindo um propósito, por exemplo, apoio a uma tradução, de modo que o esforço dispendido não é utilizado por outros utilizadores, resultando numa lamentável perda de recursos. A este propósito, François Rastier (2002:2) aponta que o corpus que se pretende elaborar deve ser desejado, quase amado sob a pena de resultar inútil:

Tout corpus suppose en effect une préconception des applications, fussent-elles simplement documentaires, en vue desquelles il est rassemblé: elle détermine le choix des textes, mais aussi leur mode de «nettoyage», leur codage, leur étiquetage; enfin, la structuration même du corpus. Allons plus loin, un corpus doit «être aimé»: s'il ne correspond pas à un besoin voire un désir intellectuel ou scientifique, il se périmé et devient obsolète.

Entende-se por corpus uma compilação mais ou menos extensa de fragmentos de textos e/ou textos e artigos completos em suporte electrónico, cuja característica fundamental é a representatividade do tema seleccionado na língua meta. “Corpus is a body of texts put together in a principled way, often for the purposes of linguistic research” (Stig Johansson, 1991: 3). Como refere Gloria Corpas Pastor (2004), só há pouco tempo é que os tradutores se interessaram pelo corpus como recurso para a tradução especializada, quando se começou a explorar as possibilidades dos corpora virtuais, mais concretamente os corpora *ad hoc*. Um corpus *ad hoc* é um corpus virtual, denominado *precision corpora do-it-yourself corpora*, que se compila pontualmente para a realização de uma determinada tarefa de tradução. Este tipo de corpus não compila um grande número de textos mas textos muito adequados, equiparados ao texto original quanto à temática, ao género e à variedade textual.

El objetivo central es la elaboración de un recurso rápido, fiable y a muy bajo costo que permita al traductor satisfacer las necesidades documentales específicas derivadas del encargo (conceptuales, terminológicas, estilísticas, discursivas, etc.), le proporcione modelos en la fase de reexpresión, le ayude a resolver problemas, a tomar decisiones y a validar las soluciones adoptadas en esse vaivén interpretativo que le lleva constantemente del texto origen al texto meta y viceversa. (Corpas, 2004: 236)

### I.3. Tipos de Corpora Electrónicos

Em primeiro lugar, deve distinguir-se os corpora representativos em oposição à livre compilação de textos para servir um objectivo específico. Existem muitos tipos de corpora: o corpus oral em formato áudio; o corpus de língua falada, transcrições de textos orais armazenados em suporte digital, como por exemplo o *The Lancaster/ IBM SEC corpus* e *The Machine – Readable Corpus of Spoken English*; o corpus de língua escrita, extraído de textos escritos e o corpus misto. Sara Laviosa, na obra *Corpus-based Translation Studies. Theory, findings, applications* (2002), estabelece uma classificação hierárquica organizada dos corpora em quatro níveis. A primeira classificação diz respeito ao tipo de textos compilados, pode falar-se em corpus de textos completos, incluindo textos na sua totalidade, corpus de fragmentos, que recolhe fragmentos de textos de extensão determinada, e os corpus mistos, recolhendo tanto textos completos como fragmentos. Os corpora sincrónicos em oposição aos corpora diacrónicos compõem a segunda classificação. Seguidamente, refere-se aos corpora especializados, compostos por textos correspondentes a áreas de conhecimento específicos especializados ou terminológicos e os corpora de referência ou gerais que abrangem uma grande variedade de géneros textuais, estilos e registos. Quanto às línguas representadas podemos caracterizar os corpora como monolíngues, que podem ser simples – podendo os simples ser constituídos por textos traduzidos ou textos originais – ou comparáveis, bilíngues (paralelos ou comparáveis) e multilíngues (paralelos ou comparáveis).

Neste trabalho, desviaremos a nossa atenção de todos os corpora que não sejam virtuais. Os textos seleccionados para a elaboração do corpus virtual estão originalmente em formato electrónico e a selecção textual realiza-se tendo em conta as necessidades e interesses particulares dos seus criadores, dependendo do grau de acessibilidade dos seus documentos. Apesar de este tipo de corpus ser muito útil, apresenta algumas limitações práticas, nomeadamente, a presença desigual que têm algumas línguas em relação a outras (facto que registaremos posteriormente como uma verdade empírica), o que resulta em uma maior dificuldade de localizar recursos e construir um corpus em outras línguas que não o inglês (língua por excelência da comunicação científica).

#### I.4. A Unidade Mínima de Tradução

O nascimento da Linguística de Corpus moderna pode fixar-se no início da década de 1960, quando se começaram a criar os primeiros corpora electrónicos de mais de um milhão de palavras como o *Brown Corpus*. W. Nelson Francis e Henry Kucera são os conhecidos compiladores do *Brown Corpus*. Estes investigadores aperceberam-se das possibilidades desta nova ferramenta de trabalho e tiveram a boa vontade de torná-la disponível a outros investigadores. Foi neste espírito que a *International Computer Archive of Modern English (ICAME)* se estabeleceu em 1977, com o objectivo de promover a cooperação entre os compiladores de corpus. Muita coisa mudou desde essa altura, mas este círculo de investigadores de corpus cresceu e esta partilha de informação e intercâmbio de ideias mantém-se e tem dado frutos. Na década de 70, o Generativismo chomskiano racionalizou a Linguística. Como consequência deste posicionamento teórico, rejeita-se o uso de corpora como fonte de dados, pois para a linguística chomskiana todo o corpus é parcial, uma vez que recolhe só um número limitado de exemplos de actuações e não recolhe a competência dos falantes. Isto é, o corpus oferece um número limitado de casos face ao conjunto infinito de possibilidades que a introspecção oferece à linguística.

Surge por essa altura uma nova escola, ou disciplina, denominada *Translation Studies*, com o objectivo de formular uma teoria da tradução que resolvesse problemas filosóficos e de se fixar no estudo quase exclusivo dos processos de tradução. Entretanto, Lefevere, teórico holandês da Tradução Literária, reage, no início dos anos setenta, ao estado de apatia dos estudos de tradução que só conhecem duas tendências: uma denominada abordagem científica preocupada apenas com questões linguísticas, rejeitando a especulação subjectiva e outra concentrada em questões literárias, rejeitando qualquer tipo de investigação linguística. Assim, Lefevere acusa os estudiosos desta área de serem escravos de ideologias, vivendo de costas voltadas para o progresso, concentrados nas suas teorias pessoais e desinteressados da destruição de obstáculos. Também James S. Holmes, no artigo “The Name and Nature of Translation Studies”, apresentado em 1972 em um congresso e passando despercebido até à data da

sua publicação em 1988, tece alguns comentários sobre o panorama das investigações eruditas sobre tradução, denunciando a falta de critérios definidos e de estratégias metodológicas (1988 [1972]: 68). Neste contexto, propõe uma disciplina empírica totalmente renovada e acrescenta que a função da tradução é cumprir um papel sócio-cultural dentro da comunidade receptora. O mais importante é “which texts were (and, often as important, were not) translated at a certain time in a certain place and what influences were exerted in consequence.” (Holmes, 1988 [1972]: 72) Na óptica de Holmes, os estudos de tradução não podem estar divorciados das outras áreas de conhecimento, vincando a necessidade de uma perspectiva interdisciplinar. A ideia subjacente das obras de Holmes é a necessidade de se produzir hipóteses e trabalhar no sentido de comprovar a sua validade. O Empirismo é o lema principal desta nova disciplina – *Translation Studies* – que visa a exclusão de pressupostos teóricos não comprováveis na prática do processo tradutivo. Ao contrário do que sucedia até então, o objectivo não era forçar teorias pré-existentes à tradução, mas estudar os aspectos específicos de cada tradução, encarar a tradução como dinâmica e histórica e aplicar os resultados à teoria linguística. A questão da tradução como texto dinâmico é para Baker (1993) a maior achega dos estudos de tradução, reafirmando a autora a necessidade do tradutor se descartar de abstracções e modelos teóricos.

As propostas de Holmes e dos autores da escola holandesa e belga foram positivas, mas não foi formalizada uma teoria com objectivos específicos. Toury soube integrar num modelo único todas estas propostas. Em 1978, o israelita Even-Zohar publica *Papers in Historical Poetics*, onde introduz o termo *polissistema* para se referir ao conjunto de todos os sistemas literários de uma cultura. O escritor declara que o texto traduzido não é simplesmente um texto assimilado, introduzido na cultura meta, mas um acto de comunicação única e renovada. É uma nova criação. Portanto, o texto traduzido assume normas específicas resultantes da sua relação com a língua e a cultura do texto meta. Partindo desta ideia Toury desenvolve em 1995 a sua proposta de *Descriptive Translation Studies*, qualificando os estudos descritivos como “target oriented” por ser precisamente no texto e na cultura meta onde a observação directa se direcciona. É na tradução que se encontram as decisões do autor e as normas que as condicionaram. Se se admitir que a tradução cumpre uma função social e que o tradutor desempenha um

papel preciso num dado momento concreto, teremos que aceitar que é importante procurar as circunstâncias que condicionaram as decisões do tradutor.

Começa a perceber-se que os Estudos de Tradução podem e devem ocupar-se de mais aspectos da tradução para além da sua recepção na cultura meta, tais como certos aspectos da tradução entre línguas concretas: a estruturação dos parágrafos, a organização interna das orações, a escolha e o uso das diferentes classes de palavras para verbalizar conceitos, entre outros. Ignorar estes aspectos pode contribuir para diminuir a qualidade da tradução.

A Estilística Comparada, que estuda as relações que se estabelecem entre duas línguas durante o processo de tradução e determinam as escolhas concretas do tradutor, passa a ser encarada como fundamental no processo tradutivo. Toda a investigação necessita de contraste e de comparação para identificar e caracterizar comportamentos específicos e estabelecer que comportamentos são universais. A Estilística Comparada nunca descarta o texto como unidade estrutural de tradução. A sua dedicação ao estudo das relações ou correspondências microestruturais que se estabelecem entre o texto traduzido e o texto original, tornam-na crucial. Nesta perspectiva, talvez se justifique a inclusão da estilística comparada nos estudos de tradução.

Segundo Tirkkonen-Condit, é recorrente utilizar uma metodologia desadequada aos objectivos perseguidos em muitos estudos empíricos que se realizam em tradução: “It is often methodology which turns out to be problematic.” (1989: 3-4) Wilss assinala a necessidade de determinar com clareza o objecto de estudo de qualquer investigação sobre tradução e optar por um ponto de vista selectivo para evitar a confrontação com um campo que é quase ilimitado (1988: 63). Toury (1995), por sua vez, insiste na contínua comprovação empírica das hipóteses de trabalho com o objectivo de reformular e eliminar erros. O modelo de Toury (1995) contempla a possibilidade de ampliar sucessivamente a amostra estudada, de modo a que as distintas hipóteses de trabalho que surjam da análise possam ser comprovadas e contrastadas num corpus de textos cada vez maior que estabeleça a validade relativa das conclusões parciais.

O processo que começa no TT (texto traduzido) tem como finalidade reconstruir o processo de tradução com o objectivo de identificar equivalências (microestruturas) entre as línguas implicadas, estabelecer o conceito de equivalência que determinou as

escolhas e decisões e descobrir o conceito de tradução que governou todo o processo tradutivo. Para Toury (1995), uma tradução é qualquer texto que seja aceite como tal na cultura receptora. Pode, assim, iniciar-se a análise do TT e reconstruir o processo de tradução. Surge, então, a necessidade de delimitar e definir uma unidade de tradução válida para todos os casos.<sup>1</sup> O tradutor não trabalha com textos inteiros de uma só vez, mas com segmentos (*manageable chunks*) que são unidades funcionais e operativas. Não se trata de unidades linguísticas (gramaticais ou sintácticas), mas de unidades funcionais (comunicativas e textuais) que se realizam através de uma ou mais unidades linguísticas. Para Chesterman, traduzem-se, antes de mais, as palavras que são portadoras de significado e as unidades básicas da comunicação (1997: 22). Num segundo momento, quando se começa a ver a tradução como uma interpretação ou adaptação do original ou mesmo como um acto de criação guiado pelo texto original, surge a ideia que a tradução não se deve fazer tanto palavra a palavra, mas sim sentido a sentido. Começam, assim, as discussões sobre que unidade mínima ou unidades suportam o significado da comunicação.

Catford (1965: 85) mantém que esta unidade é a oração já que é a unidade mais directamente relacionada com a função comunicativa. Outros autores propõem a ideia que a unidade de tradução é de natureza variável. O terceiro momento no desenvolvimento das teorias distintas sobre a tradução marca a mudança de perspectiva da tradução literária para a tradução de todos os tipos de texto elaborada não só por tradutores e críticos literários, mas também por filólogos e linguistas. A análise contrastável passa a ter uma grande relevância para o estudo da tradução já que esta se entende como uma forma de confronto entre dois sistemas linguísticos. Torna-se evidente que a existência de uma equivalência formal é uma quimera e abandona-se esta ideia pela ideia de uma relação que varia segundo a situação espaço-temporal, os objectivos da tradução. Em 1978, Holmes define o processo de tradução como um simples processo linear (*serial process*) e um processo estrutural. No plano estrutural, o tradutor extrai do TO um conceito mental ou abstracto, uma rede formada pelos conceitos e as relações que entre eles se estabelecem. Num quarto momento, a tradução é tida como um processo de comunicação e de reconstrução textual. A partir do

---

<sup>1</sup> Em Sánchez Trigo (2002), sumariam-se as principais propostas que se realizaram em relação com o conceito de unidades de tradução (capítulo 6).

nascimento dos Estudos de Tradução, com os trabalhos de Even-Zohar, Holmes, Toury e Lefevere, defende-se que a unidade fundamental de comunicação e, portanto, de tradução é o texto.

Na década de 80, começa-se a questionar o recurso à introspecção como única fonte de dados e a perceber que em muitas ocasiões a teoria chega mesmo a contradizer a actuação real dos falantes. Retoma-se o interesse pelos corpora textuais e são criadas associações e entidades e dedicadas à compilação de corpus e colecções de textos: *Association for Computational Linguistics (ACL)*, *Linguistics Data Consortium (LDC)*, *Consortium for Lexical Research (CLR)*, *European Language Resources Association (ELRA)*, entre outras. A Linguística de Corpus nas décadas de 80 e 90 pode definir-se como um ramo da linguística geral dedicada à análise de grandes corpora de textos automatizados mediante o emprego de ferramentas de análise textual especialmente criadas para cada projecto específico. Em 1999, Baker afirmava que ainda estávamos no início da investigação sobre corpus aplicada à tradução, a denominada *Corpus Translation Studies*. Pouco a pouco, o progresso dos estudos tradutológicos fundiu-se inevitavelmente aos avanços e potencialidades tecnológicas. Hoje em dia é inconcebível um corpus que não esteja em formato digital ou que não seja analisável por programas informáticos. Abordando as diferentes definições de corpus propostas nos últimos anos, nomeadamente o contributo de Leech (1992: 106), Hernández refere que corpus já é sinónimo de corpus electrónico.

(...) podemos considerar que un corpus no es más que una colección de texto en formato magnético, aunque Leech completa su definición recalando que la habilidad que poseen los ordenadores para buscar, recuperar, ordenar y hacer cálculos sobre cantidades masivas de texto nos ha brindado la oportunidad de comprender y de explicar el contenido de esos córpora de formas que no eran imaginables en la era que él denomina "pre-computacional". De hecho, dado que los avances tecnológicos van tan unidos al desarrollo de la lingüística de corpus tal y como hoy en día la conocemos, Leech argumenta que debe denominarse Computer Corpus Linguistics, ya que el término "lingüística de corpus" se usaba antes del advenimiento de los ordenadores digitales (Leech *ibid*). (Pérez Hernández, 2000)

## **II. A Identidade da Terminologia**

### **II.1. O Conceito e o Termo que o Designa**

Embora a Lexicografia e a Terminologia tenham muitas afinidades existem diferenças que merecem ser assinaladas. O léxico da língua geral é o conjunto de palavras conhecidas e dominadas por um indivíduo falante da respectiva língua. O léxico da língua de especialidade é um conjunto de termos conhecidos e dominados por um indivíduo especialista em determinado domínio de conhecimento. A especificidade das línguas de especialidade restringe o seu uso aos conhecedores de determinado conjunto de códigos, a denominada comunidade científica, muitas vezes não conhecidos por qualquer sujeito falante, “la terminologie proprement dite represente le bloc restreint de vocabulaire utilisé par des spécialistes communiquant entre eux” (Cabré, 1999:60). As línguas de especialidade possuem terminologias próprias, terminologias no sentido de conjunto de termos específicos de um determinado domínio, que não são do conhecimento geral. Embora a Terminologia faça uso de termos específicos em vez dos vocábulos de âmbito geral, ela também pertence à língua natural, entendida como linguagem como um todo, utilizada para a comunicação e expressão humanas, pois o indivíduo antes de ser um especialista é um falante natural da sua língua. A Terminologia pertence a uma língua de especialidade que por sua vez pertence à língua natural. Assim, a comunicação especializada é uma comunicação natural ocorrida dentro dos campos especializados. De entre os vários objectos de estudo da terminologia destacamos três: a fraseologia, a definição e o texto especializado.

O reportório lexical da Lexicografia não se esgota no universo referencial, ao passo que na Terminologia é centrado no universo referencial. Como abordaremos posteriormente com mais clareza, a metodologia básica da terminologia é, entre outras coisas, a elaboração de mapas conceptuais de cada área do conhecimento, podendo este caminho auxiliar a identificação dos termos e permitir a elaboração de definições desses mesmos termos. Um mapa conceptual estrutura-se pelo facto dos termos se inter-relacionarem lógica e ontologicamente.

De entre as teorias da terminologia que veiculam os estudos mais recentes de organização e representação do conhecimento, merecem destaque neste trabalho a Socioterminologia de Gaudin e a Teoria Comunicativa da Terminologia de Cabré.

A Teoria Geral da Terminologia foi a primeira a tratar sistematicamente a terminologia como uma disciplina, concebendo assim os primeiros alicerces da teoria moderna. Na década de 30, o engenheiro austríaco Eugen Wüster (1898-1977), preocupado em criar um instrumento de trabalho que desfizesse a ambiguidade da comunicação técnico-científica, publica em 1938 *The Machine Tool*, dicionário especializado na área de engenharia. Com uma preocupação puramente metodológica e normativa, Wüster traça as primeiras linhas da sua teoria terminológica. Deve-se, por isso, a ele a explicitação dos primeiros princípios de tratamento dos léxicos especializados, o que se convencionou chamar Teoria Geral de Terminologia (TGT). Duas décadas mais tarde, Wüster dedica-se a um estudo profundo da natureza dos termos, ambicionando a não ambiguidade das comunicações especializadas.

A chamada Terminologia Linguística está associada ao surgimento da Socioterminologia de Gaudin em 1993, atingindo a sua forma mais elaborada com a Teoria Comunicativa da Terminologia de Cabré. Nas palavras de Costa, “a Socioterminologia surge como reacção à Normalização e à necessidade de descrição dos fenómenos terminológicos reais, observáveis nos textos de especialidade e na comunidade científica de onde brotam.” (2001:199) De facto, a face linguística que a terminologia assumiu com Gaudin e depois com Cabré está em franca oposição com a Escola de Viena, representada pela Teoria Geral da Terminologia: “la terminologie n’est plus wüsterienne” (Gaudin, 2005:3). A TGT privilegiou a dimensão conceptual do termo, caracterizada como unidade cognitiva, oferecendo a possibilidade de dissociar conceito e forma, ou significado e significante. As teorias de fundamento linguístico rejeitam a dicotomia entre pensamento e linguagem. Citando Cabré “normaliser, dans cette conception de la terminologie, signifie standardiser por mieux comprendre; donc, appauvrir la langue en faveur de la communication” (1999:56). Os termos são elementos naturais das línguas naturais. Gaudin privilegia o uso social da língua e postula que a terminologia não pode ser estudada isoladamente, alheia ao contexto social, económico e comercial. “En tant que science sociale, elle étudie le comportement collectif des locuteurs notamment la circulation sociale des termes, de la

verbalisation des savoirs et des représentations et systèmes de valeur qui leur sont attachés” (2005:4). Na sua ordem de pensamento, a Socioterminologia possibilita, não descurando outras perspectivas que também o permitem, estudar o funcionamento dos léxicos de um domínio de conhecimento que são usados na comunicação entre os pares. Assim, a origem dos termos foi objecto da sua preocupação, propondo a eliminação do artificialismo e do idealismo da escola de Viena através do exame do contexto de léxicos especializados.”Il s’agit, de nos jours, d’élargir le champ de la terminologie, de dépasser son idéalisme et son logicisme, dans une perspective qui tienne compte des acquis divers, depuis le repérage automatique et ses implications au plan de la description linguistique, jusqu’ aux avancées de l’histoire des sciences, par exemple” (2005:8).

Veiculada no final da década de noventa, - momento em que a Teoria Geral da Terminologia ganha inúmeras críticas pelo seu carácter idealista -, a Teoria Comunicativa da Terminologia impõe-se como uma teoria descritiva de base linguística e perspectiva funcionalista focada no carácter comunicativo do termo. Cabré ressalta quais os aspectos que considera concorrentes à inadequação da TGT:

Las causas generales más importantes de esta aparente «crisis» pueden sintetizarse, en nuestra opinión, en dos: el *reduccionismo* com que se había concebido la materia y sus aplicaciones, y la excesiva *uniformización* a la se había querido someter la disciplina en aras de la universalidad de los principios que la sustentaban. (1999:69)

O esquecimento dos aspectos comunicativos e pragmáticos inerentes ao léxico das linguagens especializadas; a inflexibilidade no tratamento dos termos, conceitos e características; a crença idealista de que os conceitos preexistem às denominações dos mesmos; de que o conhecimento técnico-científico é universal; de que a estruturação de um campo de conhecimento não depende do seu contexto, são alguns dos aspectos da teoria wüsteriana que Cabré mais critica. “Wüster estableció para la terminologia un objecto de análisis y unas funciones de trabajo bastante restrictivos” (76). A TCT, ao contrário da TGT, enaltece as diferenças entre a Linguística e a Terminologia. A Teoria Comunicativa da Terminologia fundamenta os seus princípios no carácter comunicativo do discurso especializado, apresentando com isso uma reflexão a respeito da língua

efectivamente utilizada no âmbito especializado. Cabré (1999) definiu uma teoria generalizada levando em consideração que a terminologia não é uma disciplina autónoma mas sim interdisciplinar, integrando aspectos da Linguística, das Ciências Cognitivas e das Ciências Sociais, isto é, integrando uma teoria da língua (que considera as unidades terminológicas dentro da língua natural sem deixar de destacar o aspecto singular do seu carácter terminológico), uma teoria do conhecimento (que explica como se conceptualiza a realidade) e uma teoria da comunicação (que descreve criteriosamente os tipos de situações que podem dar origem ao processo de comunicação, contemplando a amplitude e a diversidade desse processo), e transdisciplinar, actuando em todas as disciplinas.

Cabré (1999) postula a necessidade de conceber uma teoria de carácter flexível do processo comunicativo dentro do universo especializado, apresentando três princípios e três condições que norteiam a sua reflexão teórica: Princípio da poliedricidade do termo – “las unidades terminológicas son inherentemente poliédricas, es decir, unidades que integran al mismo tiempo aspectos lingüísticos, cognitivos y sociales” (85). A autora esclarece que a Teoria Comunicativa da Terminologia não considera os termos como unidades isoladas que constituem o seu próprio sistema, mas sim unidades que se incorporam no léxico de um falante no momento em que se adquire competência comunicativa.

O princípio do carácter comunicativo da terminologia – “toda unidad terminológica persigue inmediata o remotamente una finalidad comunicativa” (85), seja de forma directa (entre especialistas) ou indirecta (através de traduções ou interpretações). Essa unidade pode exercer várias funções (referencial, expressiva, conotativa) quando integrada em um discurso. Os termos activam-se singularmente no seu papel pragmático (buscando sempre comunicar um conhecimento) e na adequação em um discurso.

O princípio da variação – “todo proceso de comunicación comporta inherentemente variación, explicitada en formas alternativas de denominación del mismo concepto (sinonímia) o en apertura significativa de una misma forma (polisemia)” (85). No processo comunicativo existem variações do tipo sinonímico, denominações distintas para o mesmo conceito, ou polissémicas, conceitos distintos

para uma mesma denominação. O termo é visto por Cabré como uma unidade de conhecimento, composta por uma forma e um conteúdo, sendo a forma a unidade lexical que denomina o conceito, o conteúdo. Para a terminologia um termo é uma unidade composta por um único conceito e uma única designação lexical, assim se uma designação lexical possuir dois ou mais conceitos distintos, trata-se de dois ou mais termos em relação de homonímia. O mesmo se aplica quando um conceito apresenta mais de uma designação lexical, trata-se de mais de um termo em relação de sinonímia. Isto é, os termos são unidades de forma e conteúdo e o conteúdo é sempre simultâneo à forma. Se houver alguma relação de sinonímia ou de homonímia, trata-se de termos distintos relacionados.

Condição de língua natural: a linguagem especializada é um subconjunto da língua natural, pois respeita o conjunto de regras gramaticais da língua natural. Embora a terminologia esteja sobre a égide de uma língua de especialidade, restrita a um público reduzido, o seu objecto é a unidade lexical, que não é palavra, pois só a situação comunicativa poderá decidir, “los aspectos lingüísticos, cognitivos y ontológicos de la terminologia quedarían integrados en la primera teoría del language natural como vertientes distintas de un mismo objecto de estudio” (83). O termo é uma unidade lexical que faz parte da língua geral e da gramática que descreve cada língua.

Condição de comunicação especializada: a terminologia é tematicamente produzida em situações profissionais e não adquire o seu significado do objecto da realidade, mas sim de estruturas consensuais e pré-estabelecidas.

Condição de especialização: o grau de especialização de um texto é pautado no modo como este veicula a sua temática. A forma é constante, mas o seu conteúdo depende do âmbito, do tema, do tipo de texto, dos interlocutores e da situação discursiva. O texto especializado passou a assumir um papel essencial para os estudos terminológicos, na medida em que em princípio termos e palavras não se distinguem. É nos textos e nas comunicações especializadas que as unidades lexicais e as estruturas fraseológicas adquirem o seu valor, porque é nele que estão contidos todos os elementos lingüísticos e extra-lingüísticos em uma interacção entre a língua e a vida social (Costa, 2001). Necessário será ir mais além dos conteúdos temáticos, sendo primordial reconhecer os mecanismos semióticos e pragmáticos que conferem propriedades

terminológicas a determinadas estruturas linguísticas. De qualquer modo, a abordagem terminológico-linguístico-textual não representa um fim em si mesma, mas tem valor funcional. É um requisito essencial para as pesquisas terminológicas.

Apresentaram-se duas abordagens terminológicas: por um lado, segundo a Escola de Viena, os termos técnico-científicos não são aspectos naturais das línguas, caracterizando-se como unidades de conhecimento. Por outro, a corrente defendida por Gaudin e Cabré, reconhecendo que acima de tudo as terminologias são unidades lexicais e, como tal, partem dos sistemas linguísticos. A primeira abordagem postula a existência de uma fronteira entre o léxico de especialidade e o léxico comum. O termo técnico é uma unidade de conhecimento, antes de ser uma expressão linguística. A terminologia é assim uma língua artificial, idealista. As terminologias científicas e técnicas não pertencem à língua, estão fora da língua. Os termos técnico-científicos são meras etiquetas designativas ao serviço da ciência e da técnica. A terminologia expressa conceitos e não significados. A partir de Gaudin, a unidade lexical deixou de ser vista como um rótulo, passando a ser compreendida como um componente da língua. Contudo, não se pode negar a contribuição da teoria clássica da terminologia, uma vez que auxiliou a determinar o seu objecto específico – o léxico terminológico. Em virtude do destacamento específico do léxico geral a Terminologia adquiriu uma identidade própria em relação à Lexicografia.

Termo é linguisticamente falando um signo verbal que representa, denomina, designa, denota, sintetiza, expressa, especifica um conceito. Um termo é constituído por forma e conteúdo. O signo verbal mostra, no sentido de fazer ver, um conceito. Quando se consulta um texto, o investigador compreende os conceitos especializados por meio de termos especializados. O conceito manifesta-se por meio do termo, estando o primeiro contido no segundo. O conceito será a representação mental de algo real que é interpretado e mostrado pelo termo. A Teoria Comunicativa da Terminologia persegue uma terminologia baseada na linguagem natural e real – efectivamente usada nos ambientes especializados. A ideia de que o termo mostra o conceito parece ser análoga ao signo linguístico de Saussure (1992). Contudo, impõe-se aqui uma distinção fundamental. O conceito não é o significado. Saussure ilustra o signo como uma moeda

de duas faces, significado e significante: como uma unidade de conhecimento composta por um conceito (significado) e por um elemento lexical (significante) que o nomeia. Por outras palavras, significado e significante não se dissociam e, em Terminologia, o conceito é encarado de forma isolada, independente da sua denominação e em estreita relação com todos os outros conceitos que constituem um domínio de especialidade específico.

É evidente que a denominação do conceito é imperativa para que a comunicação seja possível. O homem racional não consegue viver num mundo caótico e desordenado, isto é, em um mundo onde as coisas que o rodeiam possuam formas sem “nomes” a elas associadas. Justamente devido à consciência que possui, o homem não conseguiria sobreviver em um espaço onde é difícil compreender o que o cerca. A tarefa de pensar o espaço exige que tudo seja nomeado e classificado. Os termos surgem da necessidade do homem dar inteligibilidade aos seres, animados e inanimados, do seu espaço e a sua formação implica as regras de formação de palavras de acordo com estruturas lexicais, morfossintáticas e fonéticas da língua. Quando se pensa os objectos que existem no mundo real, está a identificar-se as propriedades que os caracterizam. Conceptualizar os objectos implica a união mental, ou abstracção, das características desses objectos formando uma unidade de pensamento. Contudo, deve acrescentar-se que considerando que a acessibilidade a toda a área de saber específico em questão, no intuito de inferir informação para além da respectiva denominação, pode dizer-se que a comunidade de falantes de uma língua tem acesso à designação de um conceito, mas nem todos têm acessibilidade à mesma, porque o saber enciclopédico, o conhecimento específico e preciso é uma parcela importante nesta equação.

Considerando o objecto do nosso interesse actual: as células estaminais. O conceito de célula estaminal é constituído por traços conceptuais distintos. O público-alvo activa traços e ignora outros de acordo com o objectivo e interesse por este assunto. A preocupação do cientista prende-se com a aplicação terapêutica destas células e com que doenças se podem curar, usando terapias com transplante de células estaminais. O linguista pretende conhecer e estudar a língua de especialidade deste domínio com o intuito de extrair os termos específicos. Os pais preocupam-se em preservar as células estaminais do cordão umbilical do seu filho recém-nascido, porque acreditam que estas podem ser importantes para assegurar a sua saúde. A Igreja Católica

considera que a fertilização in-vitro é moralmente inaceitável e rejeita a investigação com células estaminais embrionárias. Cada uma destas pessoas atribui pesos diferentes a este domínio. Todos eles têm conceitos diferentes da mesma realidade.

É importante referir que a análise dos aspectos extra-linguísticos do termo não se prende apenas com o nível da conceitualização mas também com o nível semântico. Segundo Costa “os traços semânticos constituem o lugar das percepções” (2001:202). Esta análise é interessante sobretudo em sistemas linguísticos diferentes. O conceito é o mesmo, porém em virtude das diferentes percepções de uma mesma realidade, cada uma das comunidades linguísticas escolhe traços semânticos distintos do mesmo conceito. Temos assim diferentes significados contidos no mesmo significante. Pontos de vista distintos do mesmo objecto.

Em uma fase posterior, após a extracção terminológica da língua de especialidade do corpus sobre Células Estaminais, aplicar-se-á a metodologia seguida para realizar uma investigação aplicada em terminologia que deve basear-se nas Normas ISO do CT 37, fundamentalmente, nas normas 704, 1087 e 860. Assim, a análise dos conceitos e das suas relações e a análise dos respectivos termos e relações a eles inerentes devem ser a base do trabalho terminológico. Os padrões ISO, 704 (2000) e 1087-1 (2000), afirmam que o termo é a designação verbal de um conceito em um domínio específico. A Norma ISO 1087-1 (2000) define conceito da seguinte forma: “unit of knowledge created by a unique combination of characteristics.” Assim, os termos são a única via de comunicar as abstrações mentais. O termo está para língua, por ser a designação mental, assim como o conceito está para o pensamento, por ser a abstracção mental.

O sistema conceptual, que representa o conhecimento de um domínio de saber específico, e o respectivo sistema de termos é fundamental e inerente ao entendimento de todos os interessados na área seleccionada. Um sistema conceptual é organizado e classificado de acordo com os tipos de relações que se estabelecem entre os conceitos, que, de acordo com a Norma ISO 704 (2000), podem ser genéricas, partitivas e associativas. Assim como os termos podem ser simples, constituídos por uma só palavra gráfica, ou complexos, constituídos por mais de duas palavras gráficas. O termo é considerado monorreferencial quando designa um só conceito. É considerado

polissêmico quando designa dois ou mais conceitos. Dois termos são considerados sinónimos quando representam um mesmo conceito. São equivalentes em sistemas linguísticos distintos quando representam um mesmo conceito. A relação entre dois termos é hiperonímica quando a relação hierárquica que se estabelece parte do genérico para o específico e assim o primeiro impõe as suas propriedades ao segundo, criando uma dependência semântica. A relação entre eles é considerada hiponímica quando a relação hierárquica que se estabelece parte do específico para o genérico, permitindo que o primeiro possua as suas próprias características diferenciadoras, para além das propriedades semânticas impostas pelo segundo. A definição diz respeito a um texto cujo objectivo é representar e diferenciar um conceito de entre conceitos relacionados. Estas questões serão objecto de uma atenção mais detalhada, tendo em conta que a meta de um processo de representação do conhecimento é possibilitar uma eficiente recuperação da informação e consequentemente possibilitar uma precisa comunicação entre especialistas.

### **III. Metodologia**

#### **III.1. Criação do Corpus Especializado Bilingue.**

##### **Suas Características**

Nesta fase, apresentam-se os critérios utilizados para a elaboração de um corpus especializado bilingue (Português e Inglês), subordinado a um subdomínio da tradução médica – Terapia Celular através de Células Estaminais. Apresenta-se, também, a metodologia seguida que é determinante para a exploração eficaz e bem sucedida do corpus. O corpus bilingue pretende ser representativo do universo que queremos estudar, já que “unless the corpus is representative, it is ipso facto unreliable as a means of acquiring knowledge.” (Summers, 1993: 186) Um corpus representativo tem associado a si a noção de equilíbrio, ou seja, um ponto desejável de compensação entre os distintos componentes do corpus, não predominando nenhum elemento em especial, nem um género determinado, nem um autor, nem uma época. Assim, é imperativo, como refere Ball (1994), que o corpus eleito seja escolhido em função da natureza e dos objectivos da nossa investigação. É, por isso, necessário que os textos sejam relevantes para o tipo de estudo que se está a realizar. Neste contexto, fez-se a reunião cuidada e a análise dos textos que integram o corpus e seleccionaram-se as ferramentas electrónicas adequadas para ajudar nessa exploração. Trabalhar-se-á, exclusivamente, na compilação e análise de um corpus bilingue *ad hoc* com o objectivo preciso de elaborar de um glossário sobre Células Estaminais, havendo uma necessidade de localizar e compilar recursos muito adequados, equiparados quanto à temática, ao género e à variedade textual. Contudo, deve acrescentar-se que este trabalho se impõe apenas como uma amostra de um trabalho que se deseja ver ampliado futuramente.

O subdomínio médico da Medicina Celular favorece a rápida criação de corpus. O rápido avanço da investigação e das descobertas nesta área, beneficiou o progressivo aumento de recursos e obras de referência em formato electrónico para os investigadores e profissionais da saúde. Nesta perspectiva, a limitação que por vezes se enfrenta na procura de textos adequados ao objecto de estudo na Internet, não se coloca aqui. O corpus que se apresenta integra um conjunto de textos em inglês e em português. Todavia, não há comparação possível entre os textos encontrados em inglês

sobre *Células Estaminais* e os textos encontrados em português sobre o mesmo tema. O inglês é, como se sabe, a língua que predomina na comunicação científica e, nessa perspectiva, a área biomédica não é uma exceção. Existe um importante volume de textos produzidos na língua inglesa. O mesmo volume não se encontra em português, como é patente, mas é evidente a preocupação crescente para difundir o conhecimento sobre esta temática tão actual. A cura de doenças através da terapia celular é objecto de grande expectativa futura. Portanto, a selecção textual foi feita tendo em conta a área temática e a representatividade do ponto de vista da actualidade científica e social do âmbito delimitado.

Do ponto de vista cronológico, o corpus integra textos elaborados entre 2000 e 2009. Decidiu-se limitar o estudo dos textos publicados a um período de tempo de nove anos. Trata-se de um período de tempo que compreende as últimas descobertas sobre as Células Estaminais. A investigação sobre esta matéria evolui a uma velocidade alucinante, as novas descobertas depressa tornam obsoletas quaisquer conjecturas veiculadas anos antes. A actualidade é um critério fundamental neste trabalho que pretende ser continuado. Nesta perspectiva, criar um corpus dinâmico, incluindo sempre outros textos actuais, é a principal ambição deste trabalho. Teve-se a preocupação de delimitar os subdomínios do objecto de estudo e destacar as inúmeras áreas com ele relacionadas: definição de células estaminais; fontes de células estaminais; tipos de células estaminais quanto à capacidade de diferenciação; localização das células estaminais; aplicações terapêuticas das células estaminais; clonagem terapêutica/reprodutiva etc. A multidisciplinaridade própria do subdomínio seleccionado é um elemento característico do corpus. A estratégia inicial visava seleccionar uma das áreas e explorá-la detalhadamente, deixando as restantes áreas para serem trabalhadas posteriormente. Contudo, mudou-se de estratégia quando se começou a explorar o corpus e se compreendeu que o melhor era contemplar todas as áreas, ainda que limitadamente, com o propósito de extrair a terminologia básica da temática em estudo. Pensa-se que foi uma boa decisão, porque deste modo podemos ter uma visão global da área que estamos a trabalhar. Isto é fundamental, por um lado, para se conhecer o subdomínio objecto de estudo em profundidade. Por outro lado, para estabelecer um corpus de referência, que se poderá utilizar como elemento de contraste na elaboração

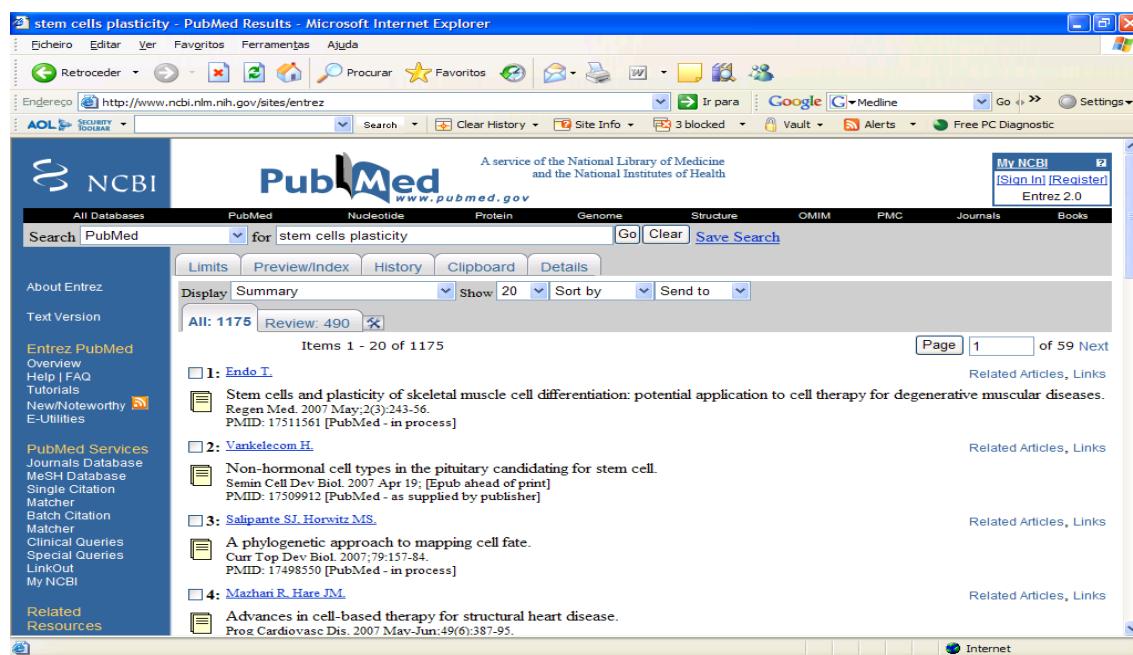
de corpus das diferentes subáreas que constituem este subdomínio. No futuro, não está descartada a hipótese de nos concentrarmos em algumas destas subáreas identificadas.

Assim, os textos que integram o corpus apresentam uma interessante diversidade tanto do ponto de vista temático (derivado da convergência de subáreas diversas), como em relação aos géneros textuais representados: incluem-se artigos procedentes de revistas científicas, artigos jornalísticos, textos de natureza informativa, textos publicitários, uma vez que este subdomínio também envolve grandes interesses financeiros, textos enciclopédicos e textos especulativos sobre a perspectiva de melhoramento da qualidade de vida futura da humanidade. Neste sentido, temos textos produzidos por diferentes tipos de pessoas e instituições envolvidas profissionalmente no âmbito da medicina celular. Incluem-se médicos, profissionais da saúde, jornalistas, empresas de criopreservação de células estaminais, cientistas e investigadores da área em estudo. De acordo com os critérios de qualidade e representatividade, excluem-se páginas pessoais ou publicações que não estejam avaliadas por instituições, empresas e/ou investigadores de reconhecido prestígio.

Nesta primeira fase, em relação à procedência geográfica dos textos em Português, estabeleceu-se o seguinte critério: incluir apenas os documentos escritos em Português de Portugal e excluir todos aqueles escritos em Português do Brasil. Nenhum factor discriminatório esteve na origem desta definição. Muito pelo contrário, o Brasil apresenta um reconhecido e rigoroso trabalho na investigação de células estaminais. Contudo, infelizmente os textos online provenientes do Brasil, apresentam traduções de termos médicos que nem sempre são as mais adequadas. De uma maneira geral, conta mais o aspecto técnico do texto (prejudicado pela má qualidade da tradução), deixando para segundo plano a adequação do ponto de vista linguístico.

A Internet (e os serviços por ela fornecidos) foi, obviamente, a grande aliada. “The worldwide Web (WWW) is the latest and definitely the most influential of the many Internet Services available (...) The WWW was the first graphical user interface for accessing the contents of the Internet” (Austermuhl, 2001:43). O acesso rápido e gratuito a bases de dados como Medline, permitiu o acesso à informação mais actual do que se vai produzindo no âmbito sanitário. A dita base de dados proporciona resumos dos artigos indexados, assim como o acesso a textos completos mediante o pagamento

de uma cota, ou de forma totalmente gratuita. As publicações periódicas em formato electrónico são também uma fonte inestimável de documentos científicos. Muitas destas revistas só permitem o acesso aos resumos e ao índice. Também se encontram na Internet listas actualizadas de revistas electrónicas, como o directório de recursos biomédicos *MedWeb*, Universidade de Emory (E.U), para o inglês, e como é do conhecimento comum PubMed [<http://www.ncbi.nlm.nih.gov/PubMed>] inclui links de artigos completos e milhões de citações da Medline, bem como outros recursos relacionados. Fez-se uma pesquisa com as palavras-chave gerais como “stem cells”, “stem cells plasticity”, “adult stem cells”, “embryonal stem cells” e com termos mais específicos como “haematopoietic, hepatic or muscle stem cells”, “Parkinson’s disease” e “insulin-dependent diabetes”.



O *Index das Revistas Médicas Portuguesas* [<http://www.indexrmp.com>], base de dados da literatura e das teses médicas exclusivamente nacionais é uma grande ajuda no caso do português. Do conjunto de revistas médicas apresentadas, deu-se especial atenção a duas, *Sinapse* e *e.Ciência – A Revista da Ciência, Tecnologia e Inovação em Portugal*, uma vez que contêm artigos de particular interesse da área em estudo.

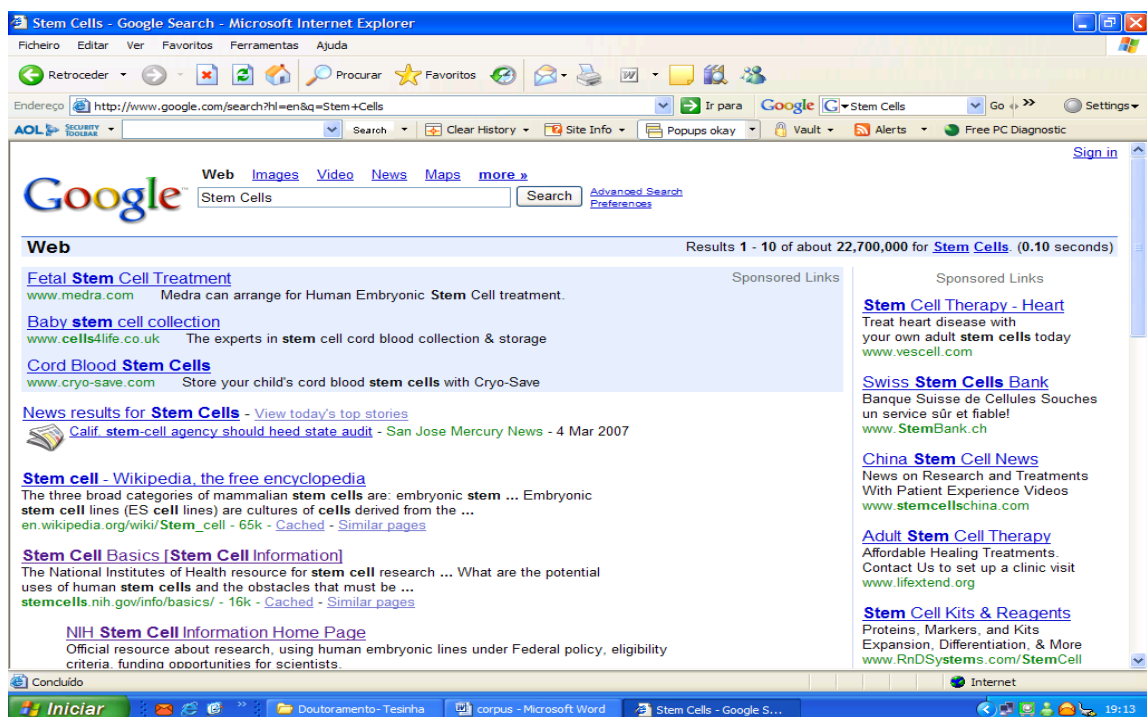


Por seu lado, no caso do inglês, pode consultar-se a base de dados [<http://www.freemedicaljournals.com>], que nos deu acesso a revistas como *The American Journal of Transplantation*, *Human Molecular Genetics* e *Bioscience, Biotechnology and Biochemistry*.



Tanto as bases de dados como as revistas electrónicas de tema biosanitário oferecem um alto grau de qualidade e fiabilidade, tendo em conta que os seus emissores são especializados organismos nacionais ou internacionais, grupos de investigação e departamentos universitários, assim como conselhos editoriais formados por especialistas.

Outro recurso importante para localizar e seleccionar as fontes de informação biosanitárias mais adequadas, para a compilação de um corpus de textos científicos, é apelar aos metapesquisadores especializados no ramo médico: *Sirus* ([www.sirus.com](http://www.sirus.com)), *HealthLinks* (<http://healthlinks.washington.edu/hsl>), Cabe ao compilador realizar a avaliação pertinente dos artigos de qualidade desigual, atendendo a critérios como a autoria e as actualizações. Finalmente, recorreu-se aos motores de pesquisa com o intuito de encontrar artigos completos para engrossar o corpus. Nas palavras de Austermuhl (2001:59), “Indeed, at present no search tool would be able to list all the information contained in all the documents. Yet search engines such as AltaVista, Hotbot, (Google) or Northern Light let the user search through the contents of several hundred million websites at a time”. O *Google*, (<http://www.google.com>), como motor de pesquisa, é muito eficaz, uma vez que utiliza sistemas totalmente automatizados para construir a sua base de dados. Tem um grau elevado de precisão e cobertura, um robot de rastreio hipertextual e um motor de recuperação robusto. Podem fazer-se pesquisas simples, ou avançadas, limitar e ordenar os resultados. Numa primeira busca, introduziram-se, cada uma por sua vez, as seguintes sequências “stem cells/ células estaminais”.



Os textos que se incluem são mistos, ou seja, são textos completos ou apenas resumos e encontram-se em formato electrónico. Incluiu-se os resumos quando se considerou uma fonte vocabular importante, uma vez que aqui importa fazer uma extracção terminológica em inglês e português e, portanto, a recolha do maior número de termos específicos sobre o tema em estudo é fundamental.

Uma outra questão é o tamanho do corpus. Quantidade não significa necessariamente qualidade, mas sabe-se que um corpus representativo necessita ser extenso, pois com a utilização de ferramentas informáticas muito adequadas é possível manejar grandes volumes de texto e obter resultados fiáveis. Contudo, como se sabe os corpus especializados, como é o caso do corpus sobre Células Estaminais, são sempre menos reduzidos que os corpus não especializados. Para além disso, nesta fase, não é uma preocupação compilar corpora muito grandes, de onde se extraem milhares e milhões de palavras, mas sim a compilação de textos sistemáticos, planeados e estruturados, destinados ao propósito de descobrir que vocábulos e significados devem ser incluídos no glossário bilingue por representarem as palavras mais frequentemente usadas no âmbito das Células Estaminais. Concretamente, introduziram-se sete textos na base de dados que se intitulou *Células Estaminais.pt* e seis textos na base de dados

denominada *Stem Cells.com*, originando um corpus com 5209 palavras e outro com 10647 palavras respectivamente. Numa primeira abordagem, parece haver um desequilíbrio entre os dois corpora, no que respeita ao número de palavras. Contudo, após a cuidada exploração e análise dos dois corpora, extraíram-se cerca de 100 candidatos a termo em cada um deles, obtendo-se um resultado final de 87 termos seleccionados, no caso do Português, e 90 termos, no que diz respeito ao Inglês.

Em conclusão, a estratégia prevista é: criar um corpus de domínio específico (Células Estaminais), gerar dentro desse tema quantos subconjuntos forem necessários, pesquisar a Internet e obter três ou quatro URL para cada subconjunto em português e inglês, ir buscar cada um dos ficheiros, limpá-los e juntá-los, processá-los e analisá-los com a ajuda de algumas ferramentas informáticas, nomeadamente o *Wordsmith Tools* e o *Corpógrafo*, obter um corpus temático com o intuito de elaborar um glossário bilingue sobre Células Estaminais.

## IV. Resultados

### IV.1. Descrição dos Resultados

Neste capítulo, compilam-se os textos seleccionados e organiza-se o corpus bilingue sobre Células Estaminais, preparando-o para ser analisado. Recorrer-se-á ao apoio das ferramentas electrónicas de análise linguística para exploração de corpora: O *Corpógrafo* e o *Wordsmith Tools*. O *Corpógrafo* possui um ambiente de trabalho simples para pesquisa de corpora, facilitando a tarefa de compilar e extrair terminologia. O *Wordsmith Tools* é uma ferramenta de análise lexical. Ambos são distribuídos exclusivamente via Internet. Ilustrar-se-á em pormenor os passos que foram efectuados em cada uma das áreas do *Corpógrafo* e as tarefas complementares operadas no *Wordsmith Tools*. Analisar-se-á as vantagens das duas ferramentas e o interesse da sua complementaridade. Comentar-se-á o número de termos que se incluíram nas bases de dados, as relações dos termos mais significativos e algumas ocorrências.

Realizada a tarefa de selecção de textos, o passo seguinte é a construção dos corpora propriamente ditos. A compilação, que implica a união de todos os textos num só arquivo electrónico, e a organização dos nossos corpora pessoais são as tarefas que se impõem após a cuidadosa selecção dos textos que deles farão parte. Para isso, recorrer-se-á fundamentalmente ao apoio do *Corpógrafo*<sup>2</sup>, mostrando o que pode ser feito com ele e o caminho a percorrer para fazê-lo. São várias as razões que concorreram para a escolha desta ferramenta electrónica como base de suporte a esta fase mais prática do trabalho. O *Corpógrafo* é uma ferramenta livre que permite a todo o tipo de utilizadores executarem pesquisas sobre corpora e realizar estudos terminológicos, usando um ambiente simples e intuitivo e que não requer qualquer tipo de instalação. Estas características simplificam consideravelmente a criação e o processamento dos corpora

---

<sup>2</sup> O *Corpógrafo*, [<http://poloclup.linguateca.pt/corpografo>], foi desenvolvido no pólo do Porto da Linguateca (centro de recursos para o processamento computacional da língua portuguesa), localizado na faculdade de letras da Universidade do Porto (FLUP) em estreita relação com as actividades do Mestrado em Terminologia e Tradução, e é orientado por Diana Santos, responsável pelo tratamento computacional das línguas, e Belinda Maia, Professora de Inglês e Tradução da Universidade do Porto. Este projecto iniciou-se em 2003 e o seu acesso processa-se através da Web, fornecendo um ambiente de trabalho para pesquisa e desenvolvimento de corpora especializados e de terminologia, com apoio a tarefas várias de extracção de conhecimento.

que se procura elaborar. Apelar-se-á, assim, a esta plataforma para analisar e estudar o corpus em questão e retirar dele informação e conhecimento. Pretende-se futuramente efectuar a criação de um glossário bilingue em suporte informático para apoiar a actividade tradutológica e facilitar o acesso ao conhecimento. Ao integrar no mesmo ambiente de trabalho várias ferramentas que usualmente se encontram dispersas, o *Corpógrafo* promete ser uma contribuição positiva na elaboração deste trabalho. Pontualmente, recorrer-se-á, também, ao *Wordsmith Tools*<sup>3</sup>, para complementar tarefas executadas.

De seguida, ilustrar-se-á em pormenor os passos que foram efectuados em cada uma das áreas de trabalho do *Corpógrafo*, começando-se por trabalhar com uma área de trabalho denominada de *Gestor*. Tudo o que fizermos aqui condicionará futuras pesquisas, por isso é necessário estar atento para garantir a eficácia desta tarefa. Aqui vamos introduzir todos os textos que fazem parte do corpus, prepará-los, “limpá-los” (retirando tudo o que se considera irrelevante para os nossos propósitos actuais, apesar de estarmos conscientes do interesse de todos estes elementos nos textos científicos: figuras, tabelas, caracteres estranhos, referências bibliográficas, agradecimentos, cabeçalhos, resumos noutras línguas, datas, links), verificar se o correcto fraseamento não foi comprometido, classificá-los e categorizá-los para os tornar pesquisáveis. Estabeleceu-se como primeiro objectivo fazer um levantamento terminológico de todos os textos em português europeu. Seguidamente, far-se-á a mesma operação com os textos em inglês. Descarregou-se os textos da rede e gravou-se os mesmos no computador, de seguida carregou-se os mesmos para a área pessoal do *Corpógrafo*. Uma das vantagens do *Corpógrafo* é o facto de poder receber textos do tipo HTML e Pdf sem ser necessário a sua conversão a texto plano. Clicando em **Fontes** é possível associar a cada texto carregado autores, fontes, instituições, editoras, etc., clicando em **Classificações** é possível classificar o ficheiro semanticamente, introduzindo o domínio ou subdomínio de estudo.

O primeiro texto foi retirado da Wikipédia, pois contém terminologia importante associada à área com a qual se quer trabalhar: as Células Estaminais. Carregado e

---

<sup>3</sup> Wordsmith Tools, [<http://www.lexically.net/wordsmith>], pacote informático de análise linguística criado para exploração de corpora, foi elaborado por Mike Scott em 1997 na Universidade de Liverpool e distribuído pela Oxford University Press.

classificado o texto, é altura de o “limpar”, removendo toda a informação supérflua ou ruidosa, para o tornar pesquisável.

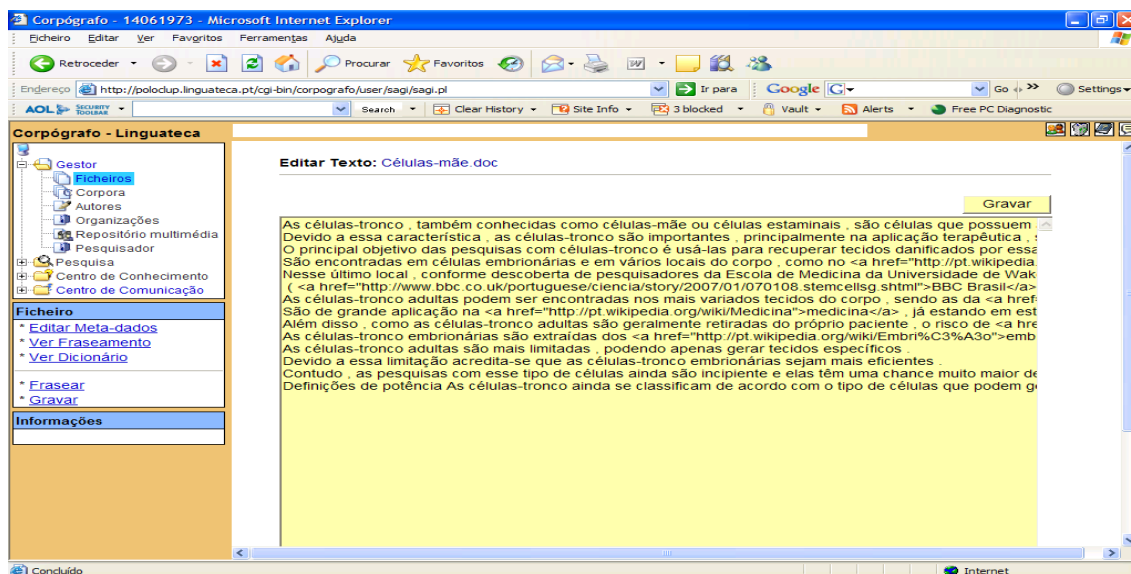


Fig.1 – Tratamento do texto

Em seguida, pode verificar-se se a divisão do texto em frases foi feita correctamente, corrigindo possíveis erros que possam ter ocorrido. Esta divisão frásica permite, também, ter uma visão mais clara do texto, tornando a sua leitura mais fácil.

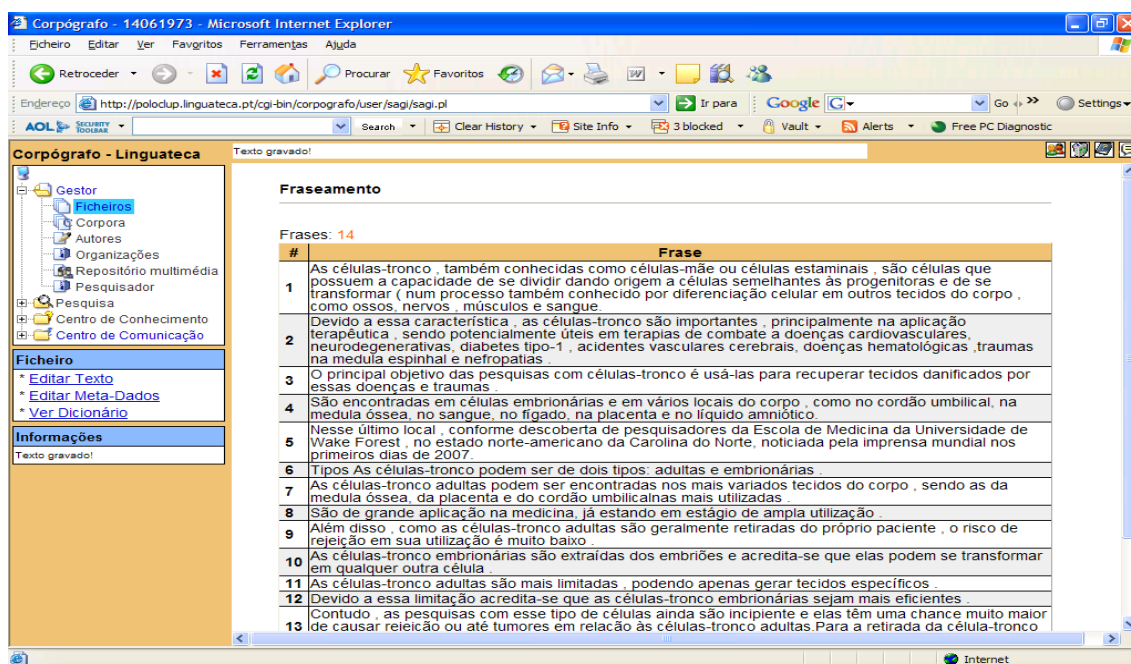


Fig.2 – Ver fraseamento

Para se ter uma panorâmica geral dos átomos que constituem o ficheiro que se encontra a tratar, a opção *Ver Dicionário* é aconselhável.

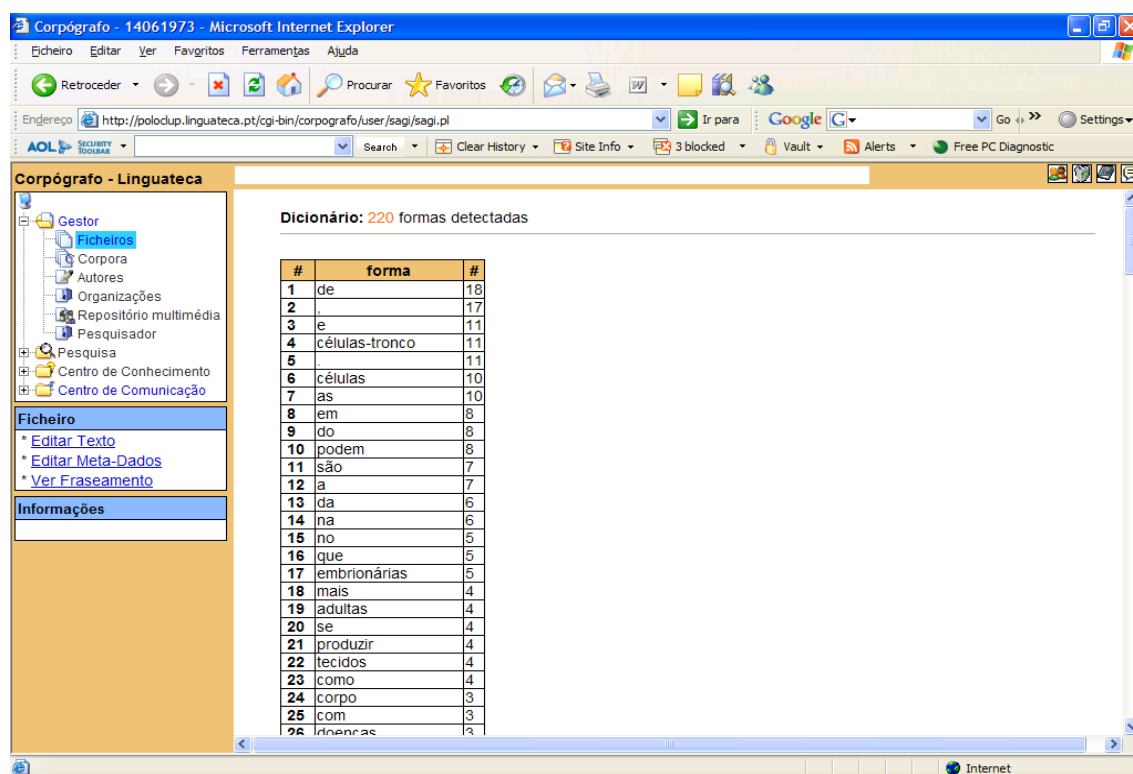


Fig.3 – Ver Dicionário

Outros textos rigorosamente seleccionados em Inglês e Português aguardam para ser introduzidos e, assim, seguir-se-á os passos acima descritos para carregar os textos, “limpá-los” e editá-los. Uma vez cumprido este objectivo, pode observar-se a lista de todos os ficheiros inseridos no Corpógrafo. Em termos concretos, carregaram-se seis textos em Inglês e sete em Português, contendo os textos ingleses 10647 palavras e os textos portugueses 5209, na sua totalidade. À primeira vista parece existir um desequilíbrio entre o Português e o Inglês, mas provar-se-á que tal não vai acontecer.

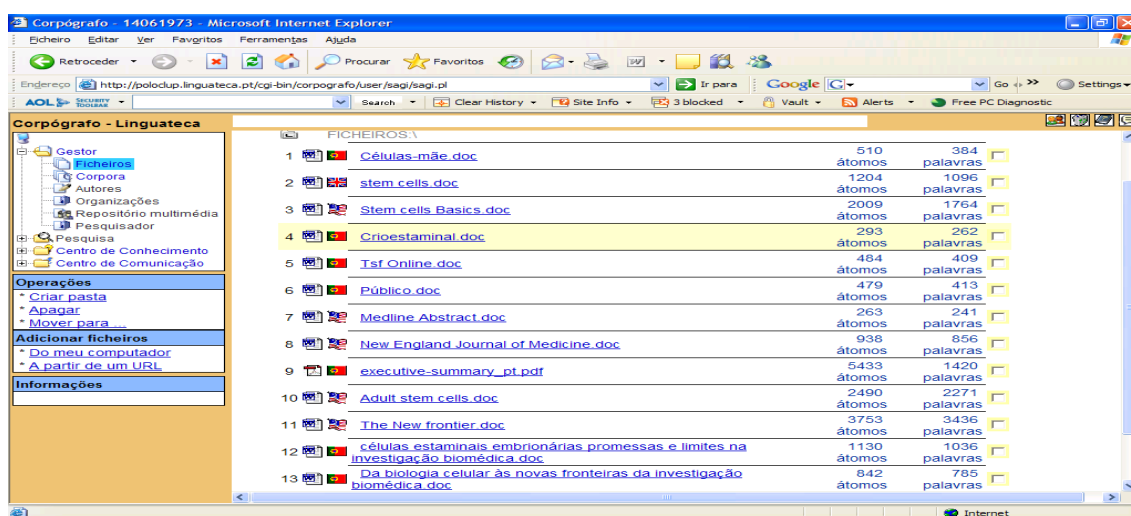


Fig.4 – Lista de Ficheiros

Posteriormente, criou-se duas pastas para inserir estes ficheiros: a primeira a que demos o nome de “Stem Cells.com”, contemplando - como o nome indica - todos os documentos escritos em Inglês, e a outra denominada “Células Estaminais. pt”, para os textos inseridos em Português europeu. Seguidamente, far-se-á a mesma tarefa para a criação de corpora, criando-se um corpus em Inglês, incluindo os ficheiros colocados na pasta “Stem Cells”, e outro em Português, incluindo os ficheiros da pasta “Células Estaminais”. Realizadas todas estas tarefas, os dois corpora estão prontos para ser pesquisados. A operação de pré-processamento de textos é muito importante, já que estes servirão objectivos muito precisos de pesquisa. O nosso objectivo, fazer um levantamento terminológico em Português europeu e Inglês, requer uma operação de fraseamento muito rigorosa para se poder verificar com precisão o contexto de determinadas ocorrências.

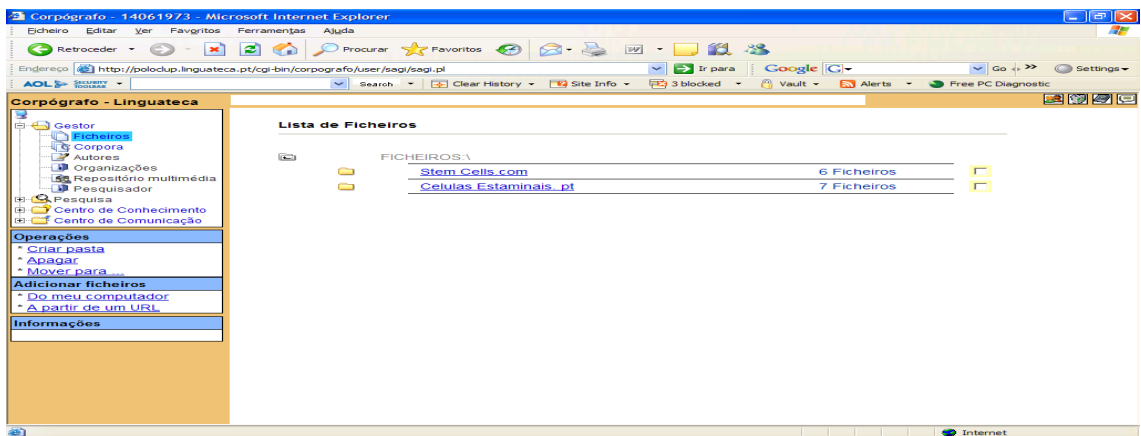


Fig.5 – Lista de Ficheiros

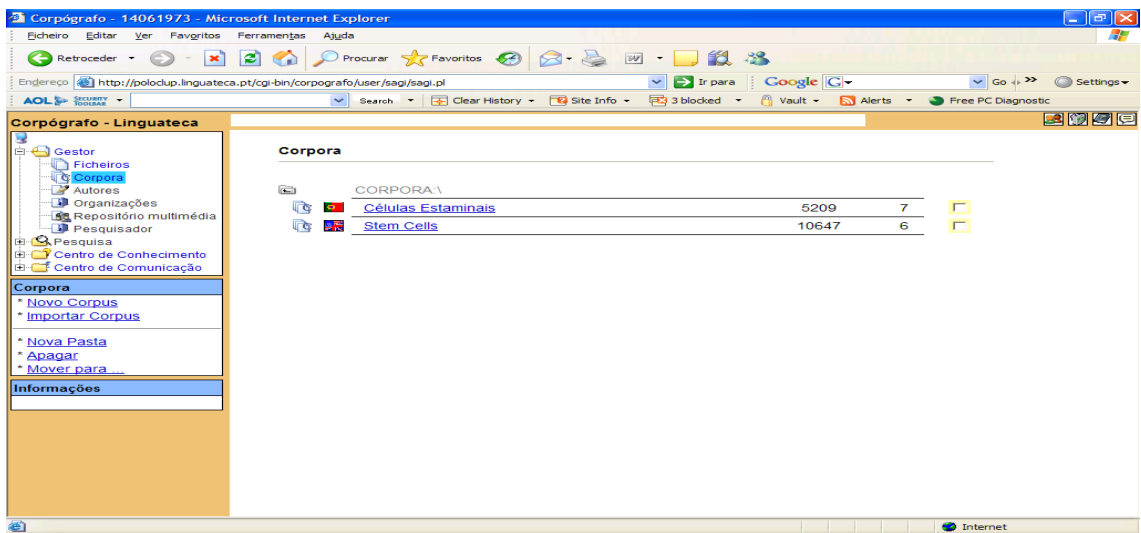


Fig.6 – Lista dos Corpora

Está na altura de se efectuar a primeira aproximação ao léxico, intitulado-se esta fase de pré-selecção de dados. Assim, começar-se-á por fazer um estudo *N-Gramas*, estudo tradicional no âmbito da Linguística Computacional, para termos uma panorâmica geral dos elementos que podemos encontrar nos textos. Este estudo revela seqüências de *N* palavras consecutivas que ocorrem no texto, para detectar a sua frequência, visualizando-se algumas das características que se destacam no texto, por exemplo, construções sistemáticas de uma ou mais palavras que permitam identificar termos ou estilos de escrita.

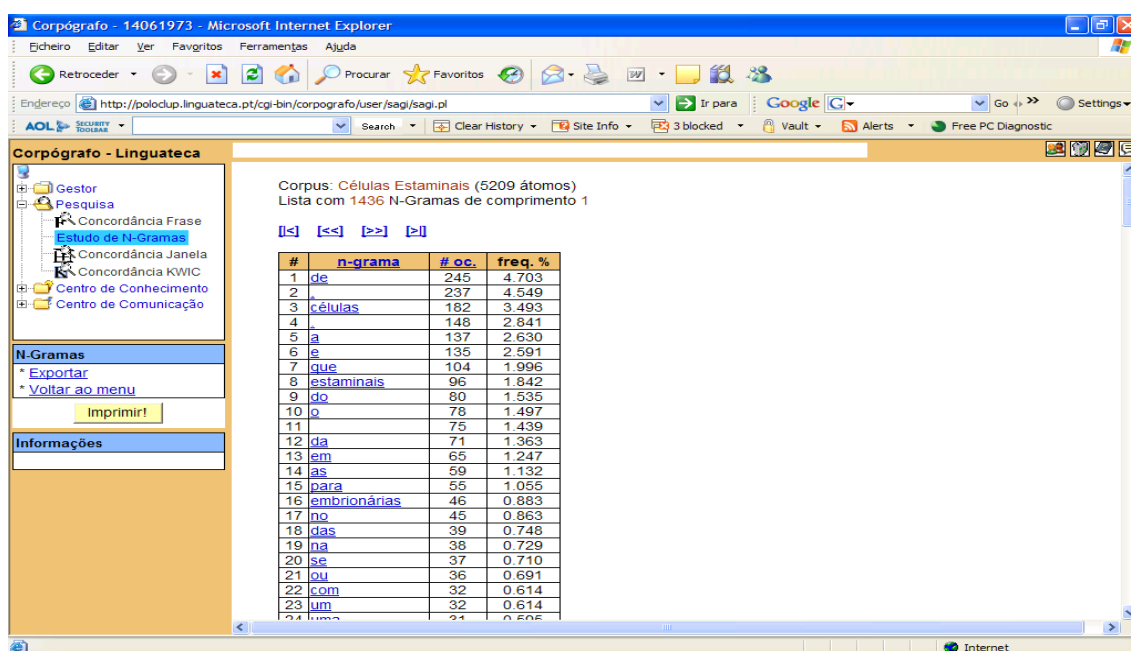


Fig.7 – Resultados do estudo *N-Gramas* (Português)

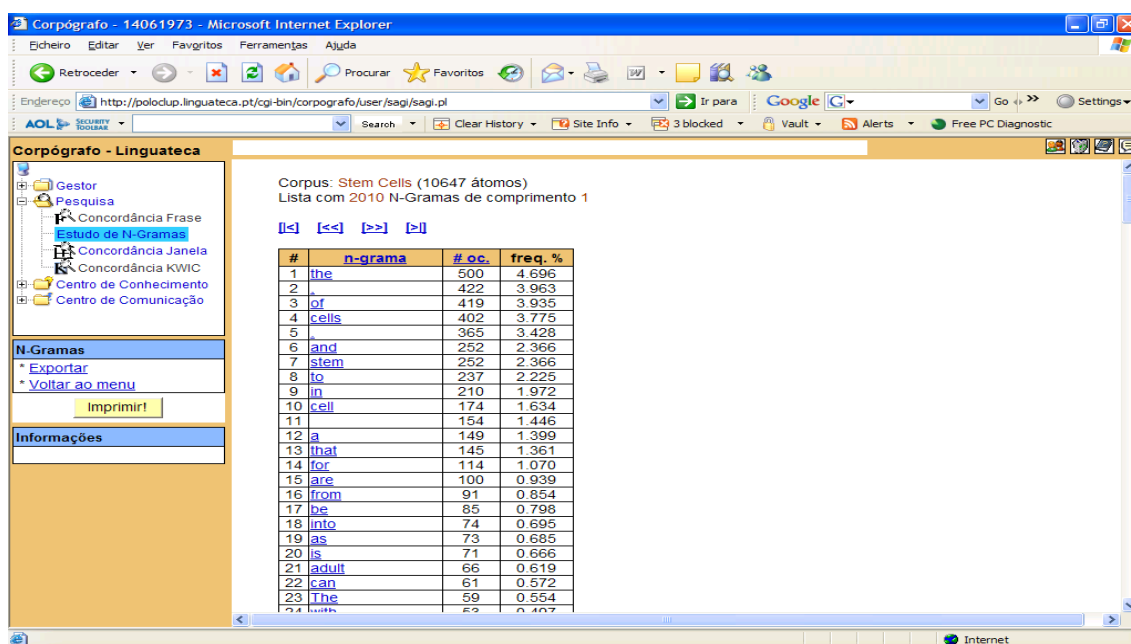


Fig.8 – Resultados do estudo *N-Gramas* (Inglês)

Nas figuras 7 e 8, mostram-se as primeiras 24 palavras do corpus em inglês e em Português. Observe-se que as primeiras palavras com conteúdo semântico e de interesse terminológico aparecem nas posições 3, 8 e 16 (células; estaminais; embrionárias) em português e nas posições 4, 7 e 21 (cells; stem; adult) em inglês. Seguidamente,

recorreu-se ao *Wordsmith Tools* para complementar esta tarefa, uma vez que este programa informático permite eliminar as palavras sem interesse terminológico que aparecem com bastante frequência, criando um ficheiro que contém uma lista de palavras gramaticais, composta por artigos definidos e indefinidos, numerais, possessivos, pronomes pessoais, isto é, um conjunto de palavras de classe fechada que não vão fazer falta. Assim, um dos métodos para estabelecer critérios válidos de identificação de termos consiste em determinar que palavras se excluem como termos. As listas que actuam de modo a filtrar, denominam-se Stopword Lists (filtros léxicos). Como aponta Vargas Sierra, “Resultan muy útiles para evitar que aparezcan en los listados que producen los programas de concordancias palabras sin contenido específico que, además, salen con una elevada frecuencia en los textos y que, por tanto, generan el indeseado ruido.” (2006: 90)

N	Word	Freq	%	Texts	%	Lemmas	Set
2	STEM	293	2,90	1	100,00		
3	EMBRYONIC	49	0,50	1	100,00		
4	TISSUE	38	0,39	1	100,00		
5	HAEMATOPOIETIC	30	0,31	1	100,00		
6	PROGENITOR	18	0,18	1	100,00		
7	EMBRYOS	14	0,14	1	100,00		
8	TRANSPLANTATION	13	0,13	1	100,00		
9	VITRO	13	0,13	1	100,00		
10	DONOR	12	0,12	1	100,00		
11	HEMATOPOIETIC	12	0,12	1	100,00		
12	BLASTOCYST	11	0,11	1	100,00		
13	EMBRYO	8	0,08	1	100,00		
14	MULTIPOTENT	8	0,08	1	100,00		
15	PLURIPOTENT	8	0,08	1	100,00		
16	INJECTED	7	0,07	1	100,00		
17	NERVE	7	0,07	1	100,00		
18	NEURONAL	7	0,07	1	100,00		
19	TRANSDIFFERENTIATION	7	0,07	1	100,00		
20	BLASTOCYSTS	6	0,06	1	100,00		
21	FETAL	6	0,06	1	100,00		
22	HEPATOCTYES	6	0,06	1	100,00		
23	ORGANISMS	6	0,06	1	100,00		
24	PARKINSONS	6	0,06	1	100,00		
25	HEPATIC	5	0,05	1	100,00		
26	MOLECULAR	5	0,05	1	100,00		
27	ENDOTHELIAL	4	0,04	1	100,00		
28	ETHICAL	4	0,04	1	100,00		
29	GENETIC	4	0,04	1	100,00		
30	REGENERATIVE	4	0,04	1	100,00		
31	TRANSPLANTS	4	0,04	1	100,00		

Fig.9 – Wordlist (Inglês)

N	Word	Freq	%	Texts	%	Lemmas	Set
1	CÉLULAS	116	2,78	1	100,00		
2	ESTAMINAIS	61	1,47	1	100,00		
3	INVESTIGAÇÃO	31	0,75	1	100,00		
4	EMBRIÕES	23	0,56	1	100,00		
5	EMBRIONÁRIAS	20	0,48	1	100,00		
6	TECIDOS	20	0,48	1	100,00		
7	CELULAR	15	0,36	1	100,00		
8	TRONCO	12	0,29	1	100,00		
9	TÉCNICAS	11	0,27	1	100,00		
10	VITRO	11	0,27	1	100,00		
11	AMNIOTICO	9	0,22	1	100,00		
12	CELULARES	9	0,22	1	100,00		
13	DIFERENCIAÇÃO	9	0,22	1	100,00		
14	DOENÇAS	9	0,22	1	100,00		
15	EMBRIÃO	9	0,22	1	100,00		
16	LÍQUIDO	9	0,22	1	100,00		
17	BIOLOGIA	8	0,19	1	100,00		
18	BIOQUÍMICA	8	0,19	1	100,00		
19	TERAPÉUTICA	8	0,19	1	100,00		
20	CLONAGEM	7	0,17	1	100,00		
21	COLHEITA	7	0,17	1	100,00		
22	CORDÃO	7	0,17	1	100,00		
23	NASCIMENTO	7	0,17	1	100,00		
24	UMBILICAL	7	0,17	1	100,00		
25	ÉTICAS	6	0,14	1	100,00		
26	MEDULA	6	0,14	1	100,00		
27	REPRODUÇÃO	6	0,14	1	100,00		
28	ABORTAMENTO	5	0,12	1	100,00		
29	GENÉTICA	5	0,12	1	100,00		
30	ÓRGÃOS	5	0,12	1	100,00		
31	AMNIÓTICAS	4	0,10	1	100,00		

Fig.10 – Wordlist (Português)

As figuras 9 e 10, apresentam uma lista dos primeiros trinta e um termos mais frequentes, após a eliminação de palavras gramaticais e outras de uso geral. Note-se a diferença: uma vez “livres de ruído” (Vargas, 2006) pode visualizar-se a lista de palavras de modo mais nítido e claro e a selecção terminológica permite já comprovar que se tratam de artigos que estudam as Células Estaminais.

Este estudo permite já aferir o vocabulário básico no âmbito das Células Estaminas, isto é, permitiu que tivéssemos uma ideia das palavras que fazem parte dos textos do corpus português e do corpus em inglês. É curioso que o mesmo vocabulário ocorre em português e em inglês, o que poderá significar que para início de trabalho se tomaram as decisões certas, pois a lista de palavras é adequada à temática do âmbito em estudo. A título de exemplo, note-se que na primeira posição aparecem em Português e Inglês a palavra-chave *Stem/ Estaminais*, depois aparece na segunda posição em Inglês a palavra *Embryonic* e na quinta em Português *Embrionárias*, na quarta posição em Inglês e na sexta em Português aparecem as palavras *Tissue/ Tecidos*. As unidades linguísticas que compõe os corpora, uma vez que se tratam de léxicos dedicados a um tema concreto, adquirem a categoria de termos. Estipulou-se que os termos devem ser substantivos ou expressões substantivadas. Os termos que serão utilizados podem ser simples (constituídos por uma só palavra) ou compostos (várias palavras ou frase).

É importante atentar aqui na dicotomia termo/conceito. O homem racional não consegue viver num mundo caótico e desordenado, isto é, num mundo onde as coisas que o rodeiam possuam formas sem “nomes” a elas associadas. Justamente devido à consciência que possui, o homem não conseguiria sobreviver num espaço onde é difícil compreender o que o cerca. A tarefa de pensar o espaço exige que tudo seja nomeado e classificado. Os termos surgem da necessidade do homem dar inteligibilidade aos seres, animados e inanimados, do seu espaço. A elaboração dos corpora tem como produto uma base de conhecimento sobre a terminologia de domínio de conhecimento específico: As Células Estaminais.

Efectuar-se-á, agora, uma pesquisa de *Concordâncias*. Começar-se-á por fazer uma *Concordância Janela* para verificarmos as palavras que ocorrem à esquerda ou à direita da palavra *Célula / Cell*. Esta operação resultará muito válida para o nosso trabalho terminológico, já que no vai permitir descobrir, entre outras coisas, que verbos,

adjectivos, advérbios se combinam com os vocábulos seleccionados. Revelar-se-á, ainda, fundamental, porque nos irá permitir caracterizar as Células Estaminais com mais rigor.

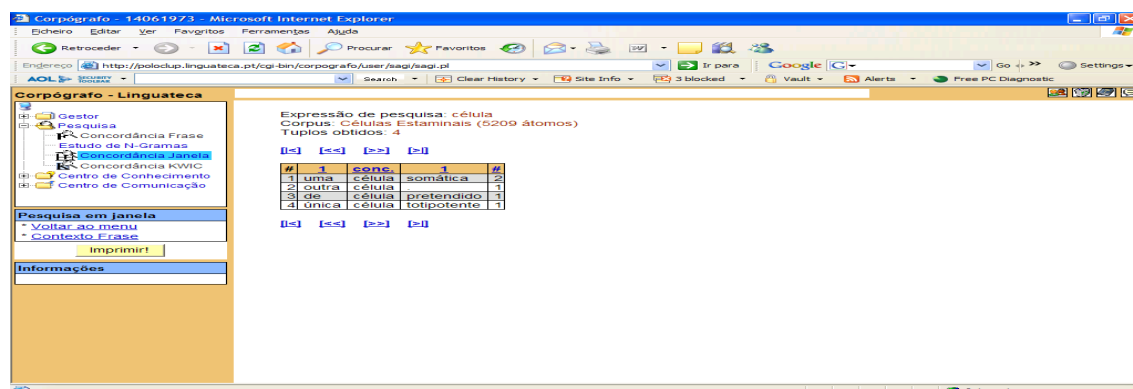


Fig.11 – Resultados de uma pesquisa *Concordância Janela*

No que concerne à pesquisa feita em português com a palavra *célula*, os resultados não foram animadores. Contudo, resultados satisfatórios foram obtidos com a introdução da palavra no plural *estaminais*. Nessa perspectiva, os resultados comunicam-nos que as Células Estaminais podem ser humanas ou não humanas e que, quanto à sua tipologia, podem ser embrionárias, adultas ou somáticas e amnióticas. No que diz respeito à sua localização, as células adultas podem ser hematopoiéticas, mesenquimais, hepáticas, etc.

Expressão de pesquisa: **estaminais**  
 Corpus: **Células Estaminais (5209 átomos)**  
 Tuplos obtidos: 37

#	1	conc.	1	#
1	células	estaminais	embrionárias	30
2	células	estaminais		6
3	células	estaminais	adultas	5
4	células	estaminais	somáticas	5
5	Células	Estaminais	e	4
6	células	estaminais	do	4
7	células	estaminais	humanas	4
8	células	estaminais		3
9	Células	estaminais	são	3
10	células	estaminais	fetais	3
11	células	estaminais	hematopoiéticas	2
12	células	estaminais	da	2
13	células	estaminais	de	2
14	células	estaminais	não	1
15	células	estaminais	mesenquimais	1
16	células	estaminais	neonatais	1
17	células	estaminais	passa	1
18	células	estaminais	com	1
19	células	estaminais	foram	1
20	células	estaminais	no	1
21	células	estaminais	existentes	1
22	células	estaminais	amnióticas	1
23	células	estaminais	embrionárias	1

Fig.12 – Resultados *Concordância Janela*

Em inglês, obtiveram-se resultados muito positivos tanto com a introdução da palavra *cell* como com a introdução da palavra *stem*. Como se pode visualizar, os resultados obtidos resultam muito idênticos daqueles que ocorreram em Português.

Expressão de pesquisa: **cell**  
 Corpus: **Stem Cells (10647 átomos)**  
 Tuplos obtidos: 125

#	1	conc.	1	#
1	stem	cell	research	13
2	inner	cell	mass	6
3	different	cell	types	5
4	stem	cell		5
5	specialized	cell	types	5
6	of	cell	division	3
7	stem	cell	differentiation	3
8	stem	cell	is	3
9	stem	cell	based	3
10	ES	cell	research	2
11	ES	cell	lines	2
12	through	cell	division	2
13	for	cell	differentiation	2
14	stem	cell	population	2
15	specific	cell	types	2
16	multiple	cell	types	2
17	stem	cell	types	2
18	stem	cell	biology	2
19	of	cell	types	2
20	desired	cell	type	2
21	Stem	cell	scientists	2
22	of	cell	membrane	1
23	multiple	cell		1

Fig.13 – Resultados *Concordância Janela*

Corpógrafo - 14061973 - Microsoft Internet Explorer

Endereço: <http://polodup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl>

Expressão de pesquisa: stem  
 Corpus: Stem Cells (10647 átomos)  
 Tuplos obtidos: 94

Concordância Janela

#	1	conc.	1	#
1	Adult	stem	cells	40
2	embryonic	stem	cells	23
3	haematopoietic	stem	cells	15
4	of	stem	cells	13
5		Stem	cells	12
6	adult	stem	cell	10
7	that	stem	cells	8
8	a	stem	cell	7
9	the	stem	cell	6
10	of	stem	cell	5
11	neural	stem	cells	5
12	.	stem	cells	4
13	as	stem	cells	4
14	The	stem	cells	4
15	specific	stem	cells	4
16	Neuronal	stem	cells	4
17	fetal	stem	cells	4
18	embryonic	stem	cell	3
19	hematopoietic	stem	cells	3
20	unspecialized	stem	cells	3
21	these	stem	cells	3
22	for	stem	cells	3
23	in	stem	cell	3

Fig.14 – Resultados *Concordância Janela*

Efectuou-se, também, no *Corpógrafo* e no *Wordsmith Tools*, uma pesquisa de *Concordância Frase* aos dois corpora, usando os termos centrais do âmbito especializado em estudo, ou “termos-mãe” – se quisermos usar a terminologia de Ahmad e Rogers (2001: 742) – Células *Estaminais/ Stem Cells*.

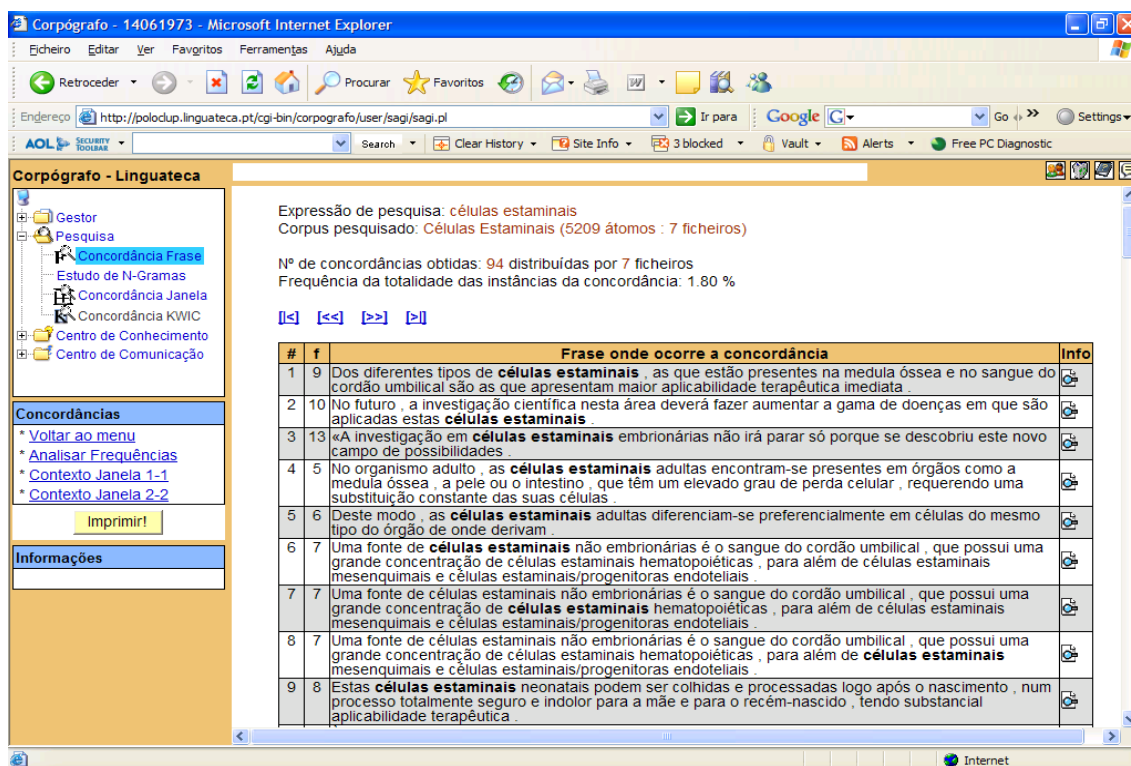


Fig.15 – Resultados da Pesquisa *Concordância Frase*

N	Concordance	et	ag	l	Word #	nt	#	Pos.	ra. #	Pos.	ct. #	Pos.	File	%
1	a partir dos seus botões embrionários células estaminais totipotentes cuja				2,301	74	69%	0	57%	0	57%	0	57% de português.txt	56%
2	curso nas últimas décadas sobre as células estaminais com o objetivo de				2,330	75	24%	0	57%	0	57%	0	57% de português.txt	56%
3	foram alcançados com a utilização de células estaminais adultas da medula				2,375	76	52%	0	59%	0	59%	0	59% de português.txt	57%
4	corso nas últimas décadas sobre as células estaminais com o objetivo de				1,751	61	24%	0	43%	0	43%	0	43% de português.txt	43%
5	Post - coisa que é comum nas células estaminais embrionárias. Para				1,072	40	92%	0	26%	0	26%	0	26% de português.txt	27%
6	embrionárias. Para recolher estas células estaminais do líquido amniótico				1,078	41	24%	0	27%	0	27%	0	27% de português.txt	27%
7	a partir dos seus botões embrionários células estaminais totipotentes cuja				1,722	60	69%	0	42%	0	42%	0	42% de português.txt	42%
8	as sociedades contemporâneas. Células estaminais são células que se				2,551	82	8%	0	63%	0	63%	0	63% de português.txt	62%
9	das suas células. Deste modo as células estaminais adultas				2,687	87	26%	0	66%	0	66%	0	66% de português.txt	65%
10	órgão de onde derivam. Uma fonte de células estaminais não embrionárias é o				2,705	88	15%	0	67%	0	67%	0	67% de português.txt	66%
11	possui uma grande concentração de células estaminais hematopoiéticas para				2,721	88	64%	0	67%	0	67%	0	67% de português.txt	66%
12	No organismo adulto as células estaminais adultas encontram-se				2,653	86	17%	0	65%	0	65%	0	65% de português.txt	64%
13	renovar e dividir indefinidamente. As células estaminais com potencial mais				2,575	83	16%	0	64%	0	64%	0	64% de português.txt	62%
14	com potencial mais abrangente são as células estaminais embrionárias,				2,583	83	58%	0	64%	0	64%	0	64% de português.txt	63%
15	do organismo avança o potencial das células estaminais passa a ser mais				2,629	85	41%	0	65%	0	65%	0	65% de português.txt	64%
16	a descoberta da potencialidade das células estaminais existentes no líquido				458	18	62%	0	11%	0	11%	0	11% de português.txt	12%
17	descobriu uma nova fonte de células estaminais no líquido amniótico				492	19	68%	0	12%	0	12%	0	12% de português.txt	13%
18	Segundo o artigo, aquelas células estaminais foram utilizadas para				507	20	16%	0	13%	0	13%	0	13% de português.txt	13%
19	da Sociedade Portuguesa de Células Estaminais e Terapia Celular,				446	18	30%	0	11%	0	11%	0	11% de português.txt	11%
20	conhecidas como células-mãe ou células estaminais, são células que				6	0	13%	0	0%	0	0%	0	0% de português.txt	0%
21	um único tipo celular maduro. As células estaminais amnióticas são mais				379	16	9%	0	9%	0	9%	0	9% de português.txt	10%
22	são mais seguras do que as células estaminais embrionárias, pois				388	16	38%	0	10%	0	10%	0	10% de português.txt	10%
23	na área da medicina regenerativa e das células estaminais». O presidente da				568	21	97%	0	14%	0	14%	0	14% de português.txt	15%
24	Foi encontrada uma nova fonte de células estaminais fetais, com				801	31	16%	0	20%	0	20%	0	20% de português.txt	20%
25	nesta população celular verdadeiras células estaminais? A resposta é sim",				893	33	93%	0	22%	0	22%	0	22% de português.txt	22%
26	do líquido amniótico é na verdade células estaminais do feto, que ele				920	35	32%	0	23%	0	23%	0	23% de português.txt	23%
27	embrionárias. «A investigação em células estaminais embrionárias não irá				730	28	26%	0	18%	0	18%	0	18% de português.txt	18%
28	embora não tanta como as células estaminais embrionárias. No				632	23	87%	0	16%	0	16%	0	16% de português.txt	16%
29	para criar teratomas. Ou seja, as células estaminais embrionárias têm				687	26	18%	0	17%	0	17%	0	17% de português.txt	17%
30	não é alternativa à investigação em células estaminais embrionárias. «A				724	27	88%	0	18%	0	18%	0	18% de português.txt	18%
31	finalidade. A obtenção deste tipo de células estaminais para este propósito				3,714	98	30%	0	91%	0	91%	0	91% de português.txt	91%

Fig.16 – Resultados da Pesquisa *Concordância (Wordsmith Tools)*

Corpógrafo - 14061973 - Microsoft Internet Explorer

Endereço: <http://polodup.linguatca.pt/cgi-bin/corpografo/user/sagi/sagi.pl>

Expressão de pesquisa: **stem cells**  
 Corpus pesquisado: **Stem Cells (10647 átomos : 6 ficheiros)**

Nº de concordâncias obtidas: **189** distribuídas por 6 ficheiros  
 Frequência da totalidade das instâncias da concordância: **1.77 %**

Concordâncias

- \* Voltar ao menu
- \* Analisar Frequências
- \* Contexto Janela 1-1
- \* Contexto Janela 2-2

Imprimir!

Informações

#	f	Frase onde ocorre a concordância	Info
1	49	Studies in Drosophila gemarium have identified the signals dpp and adherins junctions that prevent gemarium <b>stem cells</b> from differentiating .	
2	47	An alternative theory is that <b>stem cells</b> remain undifferentiated from environmental cues in their particular niche .	
3	42	To ensure self-renewal , <b>stem cells</b> undergo two types of cell division .	
4	40	Adult <b>stem cells</b> can be isolated from a tissue sample obtained from an adult .	
5	36	Many adult <b>stem cells</b> may be better classified as progenitor cells , due to their limited capacity for cellular differentiation .	
6	38	The use of adult <b>stem cells</b> in research and therapy is not as controversial as embryonic stem cells , because the production of adult stem cells does not require the destruction of an embryo .	
7	38	The use of adult stem cells in research and therapy is not as controversial as embryonic <b>stem cells</b> , because the production of adult stem cells does not require the destruction of an embryo .	
8	38	The use of adult stem cells in research and therapy is not as controversial as embryonic stem cells , because the production of adult <b>stem cells</b> does not require the destruction of an embryo .	
9	34	Also known as somatic <b>stem cells</b> , they can be found in children , as well as adults .	
10	33	Adult <b>stem cells</b> are undifferentiated cells found throughout the body that divide to replenish dying cells and regenerate damaged tissues .	
11	29	The pluripotency of ES cells has been rigorously demonstrated in vitro and in vivo , thus they can be indeed classified as <b>stem cells</b> .	
12	30	Because of their unique combined abilities of unlimited expansion and pluripotency , embryonic <b>stem</b>	

Fig.17 – Resultados da Pesquisa *Concordância Frase*

Concord

File Edit View Compute Settings Windows Help

N	Concordance	Set	Tag	Word #	t. #	os. #	. #	os. #	. #	os. #	File	%
1	daughter cells both endowed with <b>stem cell</b> properties. Asymmetric			4,263	162	8%	0	4%	0	4%	pia de inglês.txt	46%
2	funding has been provided for adult stem cell research. Adult stem cells can			4,208	158	0%	0	4%	0	4%	pia de inglês.txt	45%
3	person or animal is still alive. When a stem cell divides, each new cell has the			4,504	175	0%	0	7%	0	7%	pia de inglês.txt	48%
4	on the other hand, produces only one stem cell and a progenitor cell with			4,275	163	5%	0	5%	0	5%	pia de inglês.txt	46%
5	have been derived from embryonic stem cell research. This is not surprising			4,038	150	7%	0	2%	0	2%	pia de inglês.txt	43%
6	into different cell types) of the stem cell. Totipotent stem cells are			3,740	135	4%	0	9%	0	9%	pia de inglês.txt	40%
7	embryo. In contrast with the embryonic stem cell research, more US government			4,196	158	3%	0	4%	0	4%	pia de inglês.txt	45%
8	as well as adults. A great deal of adult stem cell research has focused on			4,106	154	9%	0	3%	0	3%	pia de inglês.txt	44%
9	cell has the potential to either remain a stem cell or become another type of cell			4,517	175	0%	0	7%	0	7%	pia de inglês.txt	48%
10	in a mature organism that cause a stem cell population to proliferate and			5,818	232	1%	0	1%	0	1%	pia de inglês.txt	61%
11	it to perform specialized functions. A stem cell cannot work with its neighbors			5,573	222	5%	0	8%	0	8%	pia de inglês.txt	59%
12	Therefore, many questions about stem cell differentiation remain. For			5,956	239	7%	0	2%	0	2%	pia de inglês.txt	63%
13	inside and outside cells that trigger stem cell differentiation. The internal			5,896	236	9%	0	1%	0	1%	pia de inglês.txt	62%
14	scientists now use the term somatic stem cell instead of adult stem cell.			5,095	198	0%	0	3%	0	3%	pia de inglês.txt	54%
15	and tear, injury, or disease. An adult stem cell is an undifferentiated cell found			5,032	196	1%	0	2%	0	2%	pia de inglês.txt	54%
16	One of the fundamental properties of a stem cell is that it does not have any			5,554	221	5%	0	8%	0	8%	pia de inglês.txt	59%
17	term somatic stem cell instead of adult stem cell. Unlike embryonic stem cells,			5,100	198	3%	0	3%	0	3%	pia de inglês.txt	54%
18	experiments to show that a single stem cell can differentiate into more than			1,366	43	5%	0	4%	0	4%	pia de inglês.txt	15%
19	few as 30 highly purified cells, multiple stem cell types could be present for			1,336	42	8%	0	4%	0	4%	pia de inglês.txt	14%
20	evidence to support the concept of stem cell transdifferentiation. These and			1,918	63	3%	0	0%	0	0%	pia de inglês.txt	21%
21	muscle contains a potent myogenic stem cell population with the ability to			1,459	47	4%	0	5%	0	5%	pia de inglês.txt	16%
22	system it has been possible to define a stem cell as a cell with the capacity to			227	8	6%	0	2%	0	2%	pia de inglês.txt	2%
23	residing in particular tissues. Adult stem cell populations have been most			186	7	0%	0	2%	0	2%	pia de inglês.txt	2%
24	The most striking suggestion of stem cell plasticity was published in			684	23	6%	0	7%	0	7%	pia de inglês.txt	7%
25	lineage within the tissue in which the stem cell resides. The ability of the			252	8	5%	0	3%	0	3%	pia de inglês.txt	3%
26	with enzymes. Some human embryonic stem-cell lines are maintained solely by			2,927	102	0%	0	0%	0	0%	pia de inglês.txt	32%
27	through research involving embryonic stem-cell lines that such factors may be			3,345	116	7%	0	5%	0	5%	pia de inglês.txt	36%
28	lines express the accepted embryonic stem-cell markers and demonstrate			3,219	112	9%	0	4%	0	4%	pia de inglês.txt	35%
29	candidates. The rigorous definition of a stem cell requires that it possesses two			3,864	126	3%	0	7%	0	7%	pia de inglês.txt	38%
30	cell types. Research in the human stem cell field grew out of findings by			3,400	118	5%	0	5%	0	5%	pia de inglês.txt	37%
31	not known how many human embryonic stem-cell lines exist in laboratories			3,011	105	9%	0	1%	0	1%	pia de inglês.txt	33%

concordance collocates plot patterns clusters filenames source text notes

68 Set Symmetric division gives rise to two identical daughter cells both endowed with stem cell properties. Asymmetric division, on the other hand, produces only one

Fig.18 – Resultados da Pesquisa *Concordância (Wordsmith Tools)*

Estas listas de concordâncias, ao permitirem observar os termos no seu contexto, autorizam a extracção de mais termos compostos e expressões que se geram a partir dos termos centrais. É o caso de combinações léxicas como *totipotent stem cells*; *progenitor stem cells*; *undifferentiated cells* e *células estaminais progenitoras endoteliais*; *células estaminais neonatais*. Na pesquisa de concordância efectuada no *Wordsmith Tools*, pode ler-se na entrada número catorze: “*Scientists now use the term somatic stem cell instead of adult stem cell*”, o que nos permite aferir que o termo somáticas e sinónimo de adultas.

Chegou o momento de se trabalhar no **Centro de Conhecimento do Corpógrafo**, onde se pode encontrar as ferramentas necessárias para criar e editar bases de dados terminológicas, pesquisar e gerir relações semânticas entre os termos. Esta fase é fundamental uma vez que é o nosso objectivo extrair toda a terminologia relevante dos dois corpora. Criou-se então duas Bases de Dados, uma em português “Células Estaminais”, e a outra em inglês “Stem Cells”. Clicando na opção pesquisar termos foram apresentadas duas tabelas, cada uma na sua vez, com os candidatos terminológicos. Seleccionou-se 108 candidatos a termos em português e 114 em inglês. Após a extracção de terminologia e o seu armazenamento na Base de Dados, este é o momento apropriado de organizar e gerir o conhecimento produzido.

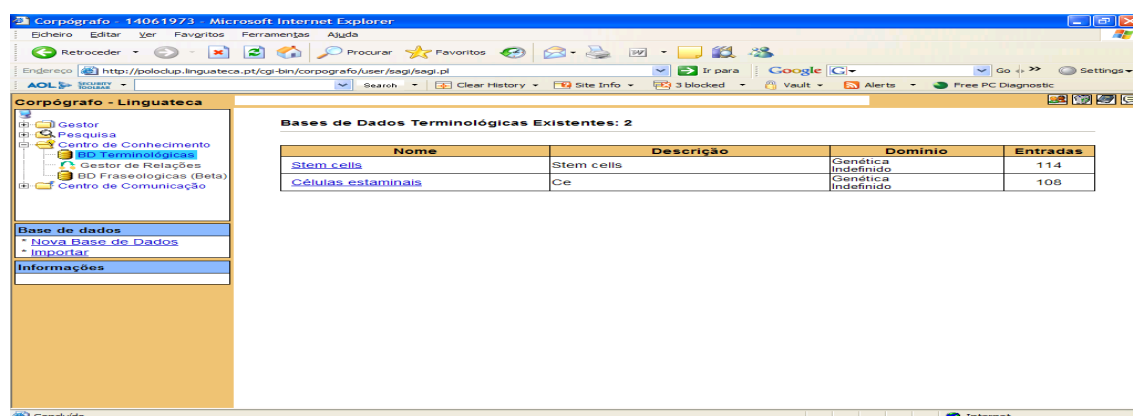


Fig.19 - Bases de Dados existentes

A opção **Listar e Editar Termos** dá-nos acesso à lista de todos os termos inseridos em cada uma das línguas por ordem alfabética. Também podemos beneficiar da opção de

ver todos os termos inseridos nas duas línguas. Pode, ainda, introduzir-se a expressão de pesquisa “célula” e temos acesso a todos os termos que começam por essa palavra.

Termo	Ícone	Código	Coluna 4	Coluna 5	Coluna 6	Coluna 7	Coluna 8	Coluna 9	Coluna 10
7 células cancerosas	🚫	? : ? : ?	1	1	0	0	0	0	0
8 células diferenciadas	🚫	? : ? : ?	1	1	0	0	0	0	0
9 células embrionárias	🚫	? : ? : ?	0	1	0	0	0	0	0
10 células embrionárias humanas	🚫	? : ? : ?	1	1	0	0	0	0	0
11 células estaminais	🚫	? : ? : ?	1	1	0	0	0	0	0
12 células estaminais adultas	🚫	? : ? : ?	1	1	0	0	0	0	0
13 células estaminais amnióticas	🚫	? : ? : ?	0	1	0	0	0	0	0
14 células estaminais embrionárias	🚫	? : ? : ?	1	1	0	0	0	0	0
15 células estaminais embrionárias humanas	🚫	? : ? : ?	1	1	0	0	0	0	0
16 células estaminais humanas	🚫	? : ? : ?	1	1	0	0	0	0	0
17 células estaminais mesenquimais	🚫	? : ? : ?	0	1	0	0	0	0	0
18 células estaminais somáticas	🚫	? : ? : ?	1	1	0	0	0	0	0
19 células estaminais/progenitoras endoteliais	🚫	? : ? : ?	0	1	0	0	0	0	0
20 células fetais do líquido amniótico	🚫	? : ? : ?	0	1	0	0	0	0	0
21 células germinais embrionárias humanas	🚫	? : ? : ?	1	1	0	0	0	0	0
22 células hepáticas	🚫	? : ? : ?	0	1	0	0	0	0	0
23 células nervosas	🚫	? : ? : ?	1	1	0	0	0	0	0
24 células produtoras de insulina	🚫	? : ? : ?	1	1	0	0	0	0	0
25 células progenitoras	🚫	? : ? : ?	0	1	0	0	0	0	0
26 células semelhantes às	🚫	? : ? : ?	0	1	0	0	0	0	0

Fig.20 – Lista de termos existentes

Para se editar os termos há que preencher a ficha do termo, introduzindo os seus dados gerais, editar informação morfológica, pesquisar definições do termo no corpus, pesquisar contextos para o termo no corpus, pesquisar relações semânticas para o termo na base de dados, associar termos relacionados numa mesma língua, associar objectos multimédia ao termo e consultar estatísticas para um termo no corpus. Esta tarefa permite seleccionar termos apropriados no universo dos termos candidatos e excluir aqueles que se apresentam como irrelevantes para esta pesquisa concreta. Nesta ordem de pensamento, elegeram-se 90 termos em Inglês e 87 em Português, o que resulta equilibrado.

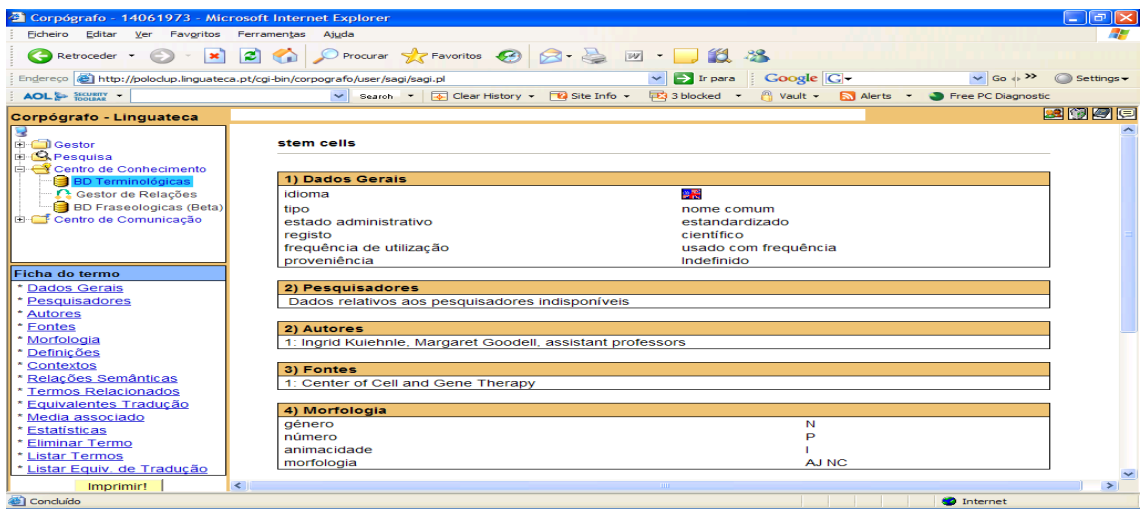


Fig.21 – Ficha do Termo

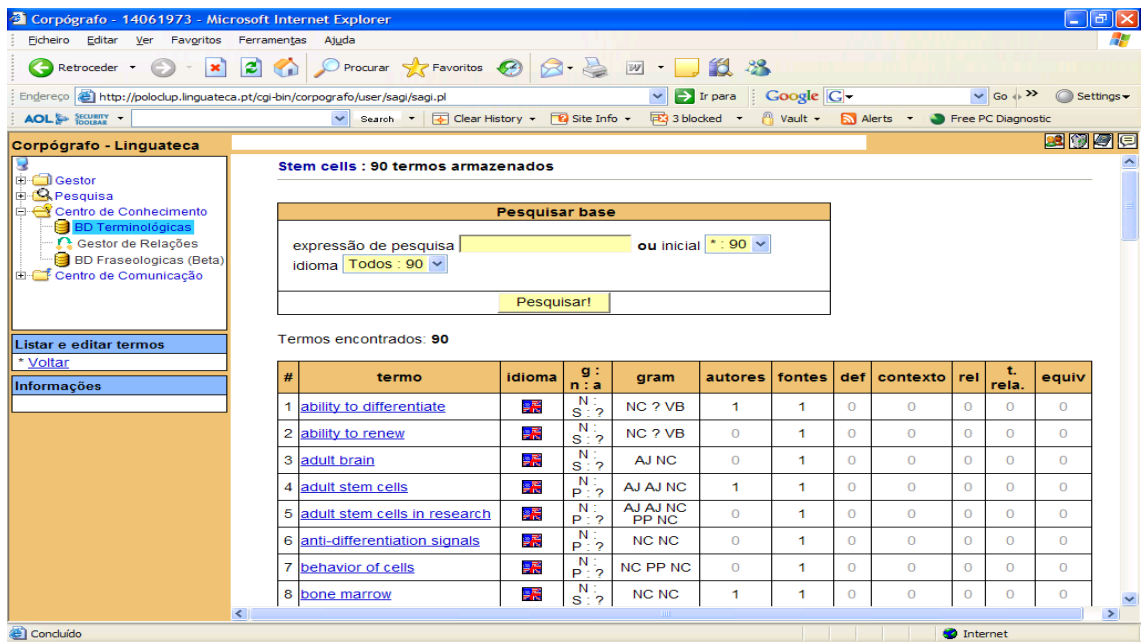


Fig.22 - Lista dos termos inseridos na base de dados

Corpógrafo - 14061973 - Microsoft Internet Explorer

Endereço: http://polodup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

**Corpógrafo - Linguateca**

**Células estaminais : 86 termos armazenados**

**Pesquisar base**

expressão de pesquisa  ou inicial \* : 86

idioma Todos : 86

**Pesquisar!**

Termos encontrados: 86

#	termo	idioma	g : n : a	gram	autores	fontes	def	contexto	rel	t. rela.	equiv
1	<a href="#">aplicação clínica</a>	PT	F : S : ?	NC AJ	1	1	0	0	0	0	0
2	<a href="#">aplicação terapêutica</a>	PT	F : S : ?	NC AJ	0	1	0	0	0	0	0
3	<a href="#">aplicações médicas específicas</a>	PT	N : P : ?	NC AJ AJ	1	1	0	0	0	0	0
4	<a href="#">área da medicina regenerativa</a>	PT	F : S : ?	NC PP NC AJ	0	1	0	0	0	0	0
5	<a href="#">auto-transplante de células estaminais</a>	PT	M : P : ?	NC PP NC AJ	1	1	0	0	0	0	0
6	<a href="#">biologia celular</a>	PT	F : S : ?	NC AJ	1	1	0	0	0	0	0
7	<a href="#">botão embrionário</a>	PT	M : P : ?	NC AJ	1	1	0	0	0	0	0
8	<a href="#">capacidade de diferenciação</a>	PT	F : S : ?	NC PP NC AJ	1	1	0	0	0	0	0

Fig.23 - Lista dos termos inseridos na base de dados

Corpógrafo - 14061973 - Microsoft Internet Explorer

Endereço: http://polodup.linguateca.pt/cgi-bin/corpografo/user/sagi/sagi.pl

**Corpógrafo - Linguateca**

**Células estaminais**

**Estatísticas sobre o corpus "Células Estaminais" (7 ficheiros - 5209 átomos)**

#	Termo	idioma	# Termos (\$+P)	OPM	# - % Ficheiros	?
1	<a href="#">aplicação clínica</a>	PT	3 (3+0)	99.99	2 - 28.6%	↓
2	<a href="#">aplicação terapêutica</a>	PT	1 (1+0)	99.99	1 - 14.3%	↓
3	<a href="#">aplicações médicas específicas</a>	PT	1 (0+1)	99.99	1 - 14.3%	↓
4	<a href="#">área da medicina regenerativa</a>	PT	1 (1+0)	99.99	1 - 14.3%	↓
5	<a href="#">auto-transplante de células estaminais</a>	PT	1 (1+0)	99.99	1 - 14.3%	↓
6	<a href="#">biologia celular</a>	PT	3 (3+0)	99.99	1 - 14.3%	↓
7	<a href="#">botão embrionário</a>	PT	4 (3+1)	99.99	2 - 28.6%	↓
8	<a href="#">capacidade de diferenciação</a>	PT	2 (2+0)	99.99	2 - 28.6%	↓
9	<a href="#">capacidade de substituir células</a>	PT	1 (1+0)	99.99	1 - 14.3%	↓
10	<a href="#">caso do transplante de células diferenciadas derivadas de células estaminais</a>	PT	1 (1+0)	99.99	1 - 14.3%	↓
11	<a href="#">células adultas</a>	PT	2 (0+2)	99.99	1 - 14.3%	↓
12	<a href="#">células amnióticas</a>	PT	3 (0+3)	99.99	1 - 14.3%	↓
13	<a href="#">células cancerosas</a>	PT	1 (0+1)	99.99	1 - 14.3%	↓
14	<a href="#">células diferenciadas</a>	PT	7 (0+7)	99.99	2 - 28.6%	↓
15	<a href="#">células embrionárias humanas</a>	PT	4 (0+4)	99.99	1 - 14.3%	↓
16	<a href="#">células estaminais adultas</a>	PT	5 (0+5)	99.99	3 - 42.9%	↓
17	<a href="#">células estaminais amnióticas</a>	PT	1 (0+1)	99.99	1 - 14.3%	↓
18	<a href="#">células estaminais embrionárias humanas</a>	PT	9 (0+9)	99.99	2 - 28.6%	↓

Fig.24 - Cálculo de Estatísticas (corpus português)

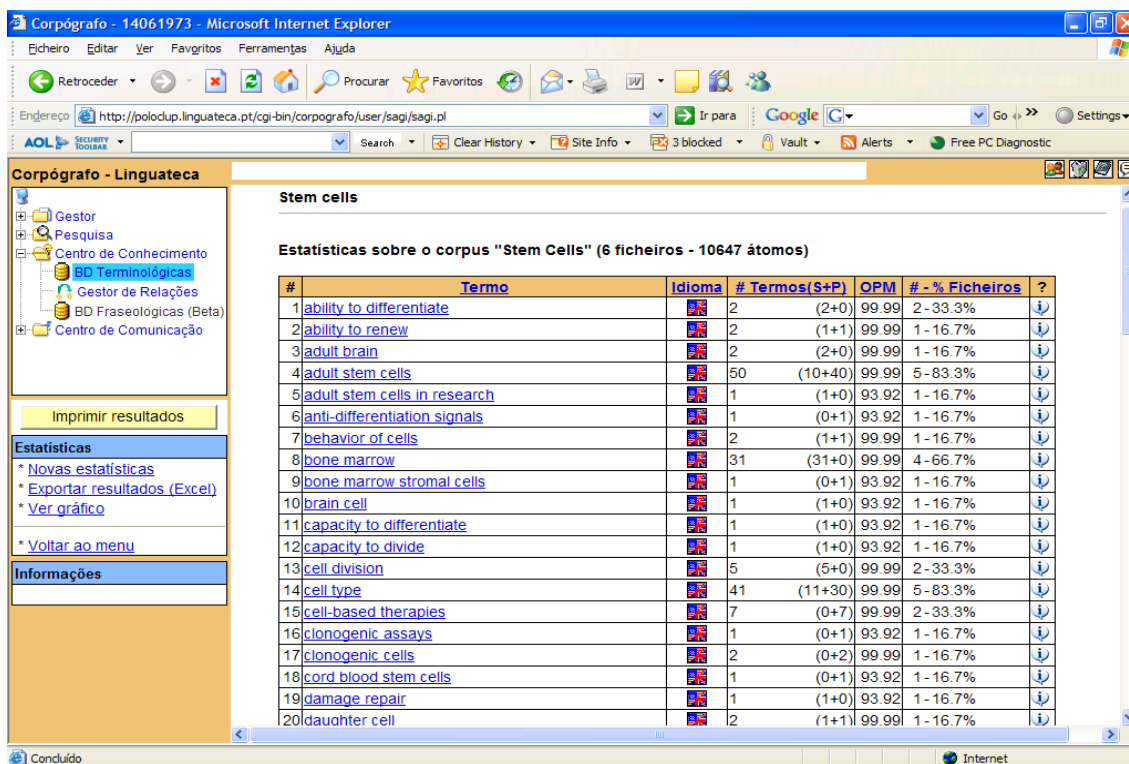


Fig.25 - Cálculo de Estatísticas (corpus inglês)

O Corpógrafo permite pesquisar candidatos à definição de um termo, desde que este já tenha sido introduzido. Esta tarefa é importante, pois é através dela que vamos reunir elementos para elaborar os textos informativos do glossário bilingue. Conceder-nos-á igualmente a oportunidade recolher as características essenciais das Células Estaminais, com o intuito de dominar cientificamente as questões que rodeiam o assunto em questão. Especificamente pesquisar-se-á o corpus português à procura de definições do termo *células estaminais* e far-se-á o mesmo com o corpus inglês para o termo *stem cells*. As definições são pesquisadas com base em padrões lexicais que indicam a presença de uma definição.

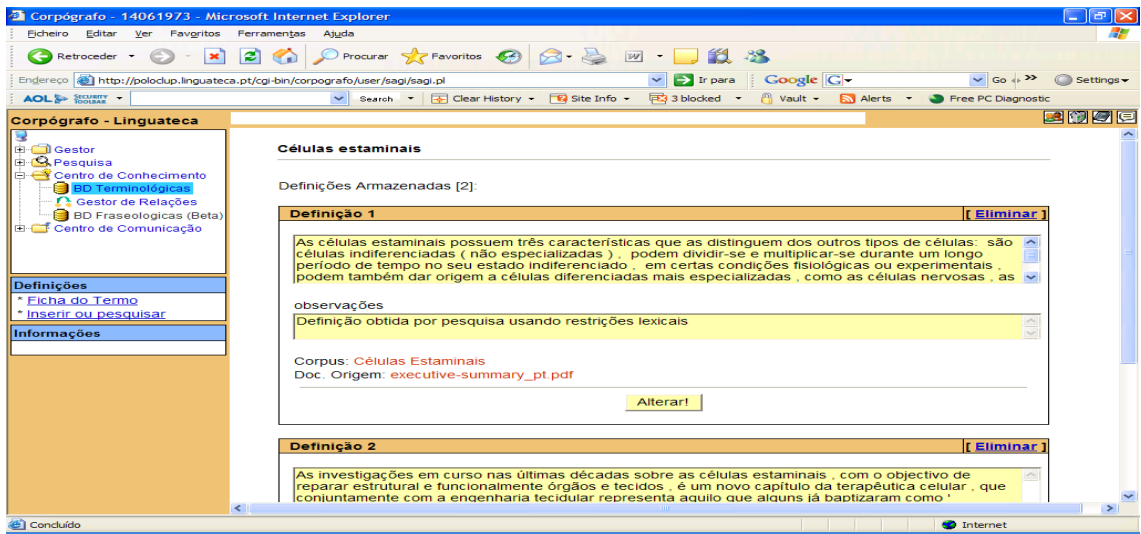


Fig.26 - Resultado de uma pesquisa de Definições, usando restrições lexicais

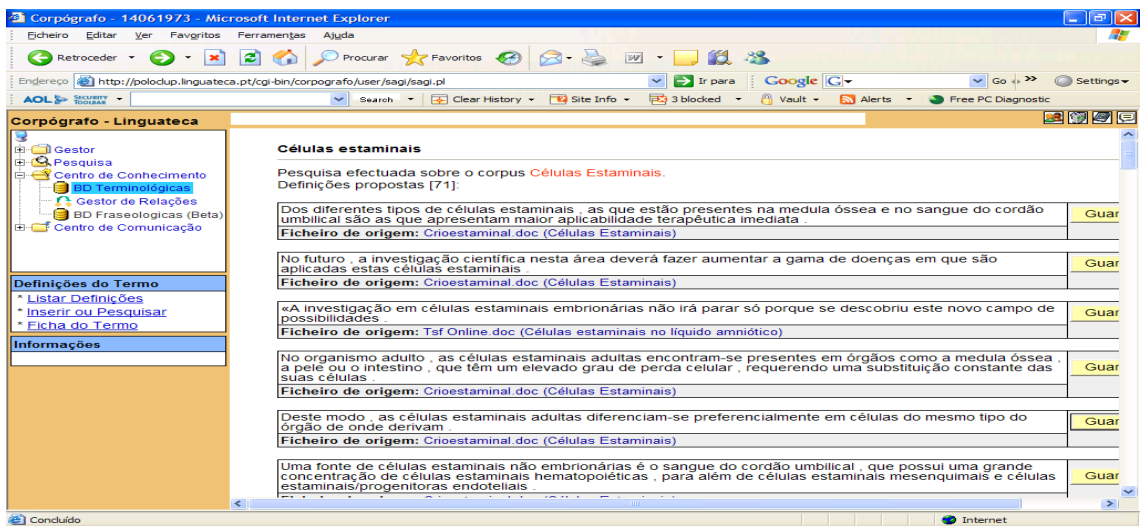


Fig.27 - Resultado de uma pesquisa simples de Definições

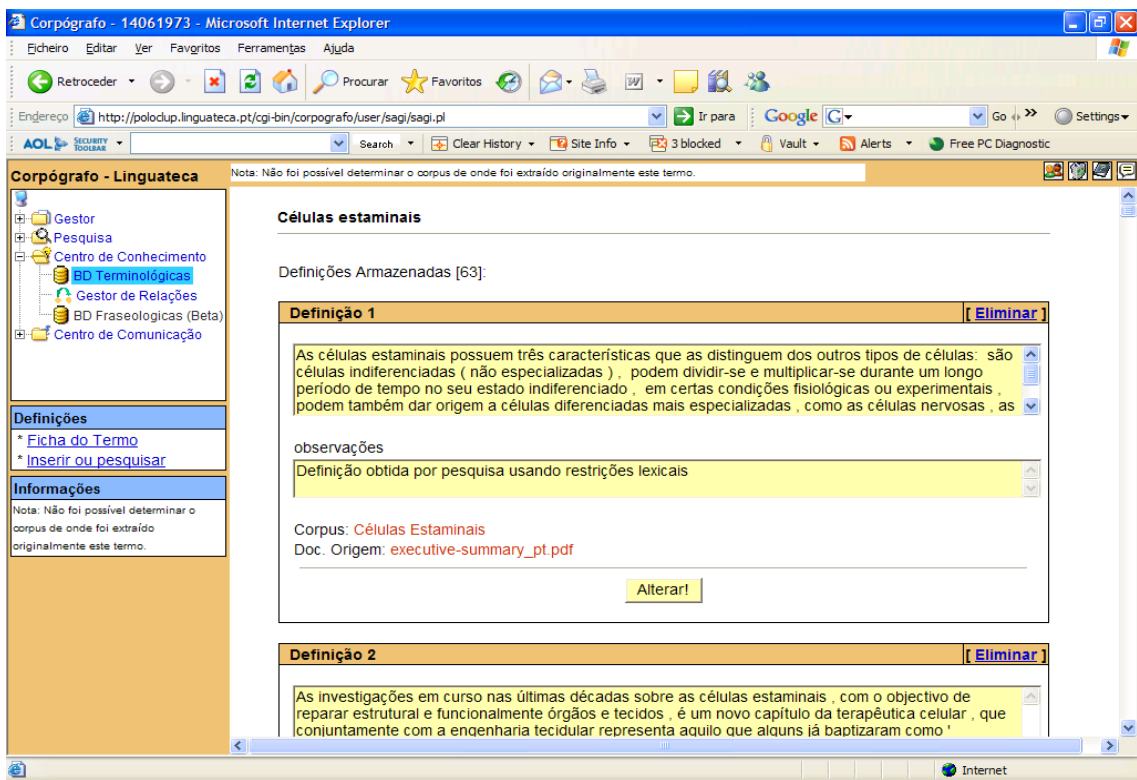


Fig.28 – Resultado de uma pesquisa de Definições já seleccionadas

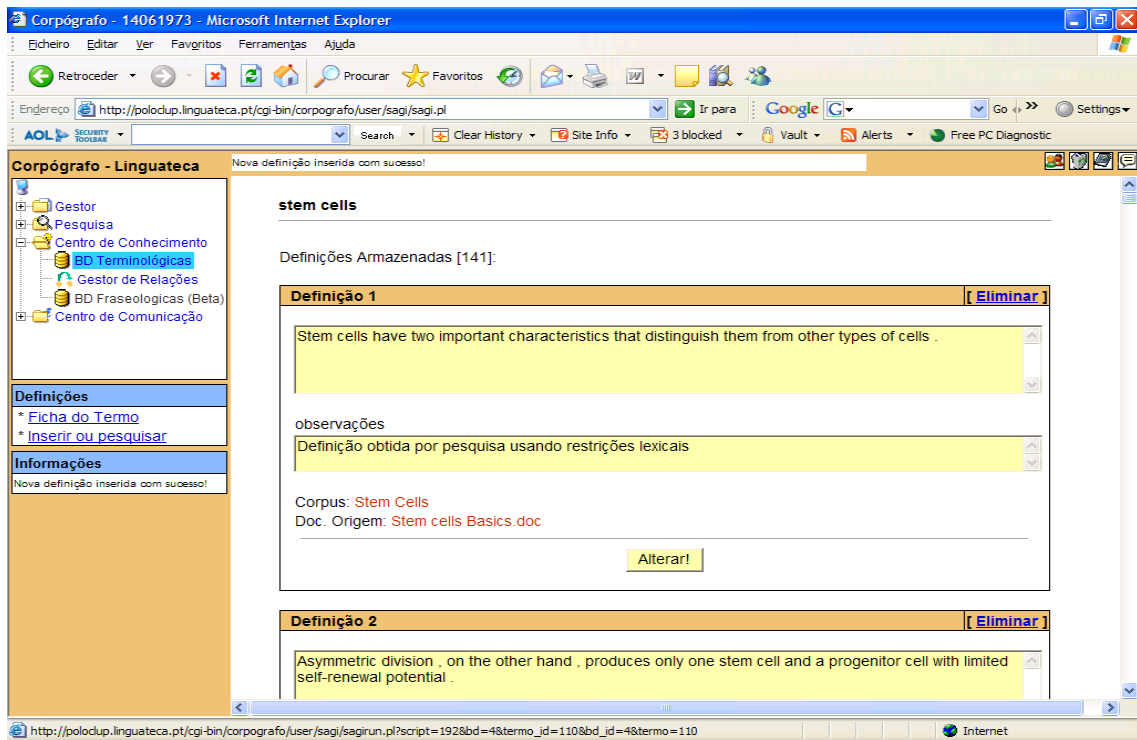


Fig.29 - Resultado final de uma pesquisa de Definições com e sem restrições lexicais

Desta pesquisa surgiram dados relevantes. As Células Estaminais são células muito especiais que se caracterizam por serem indiferenciadas (não especializadas) e por terem capacidade ilimitada de auto-renovação. Estas características, que as distinguem das outras células, fazem com que estas células possam ser programadas para desempenhar tarefas específicas no organismo. Posteriormente, é possível pesquisar e armazenar contextos para os termos introduzidos na base de dados, a fim de se observar o comportamento dos termos em contexto.

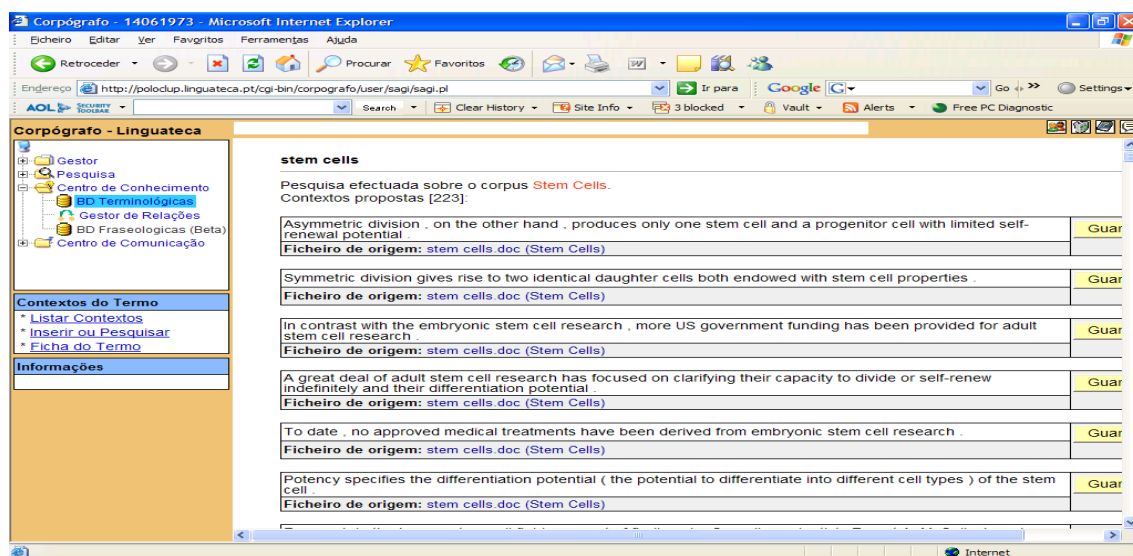


Fig.30 - Resultado de uma pesquisa de Contextos

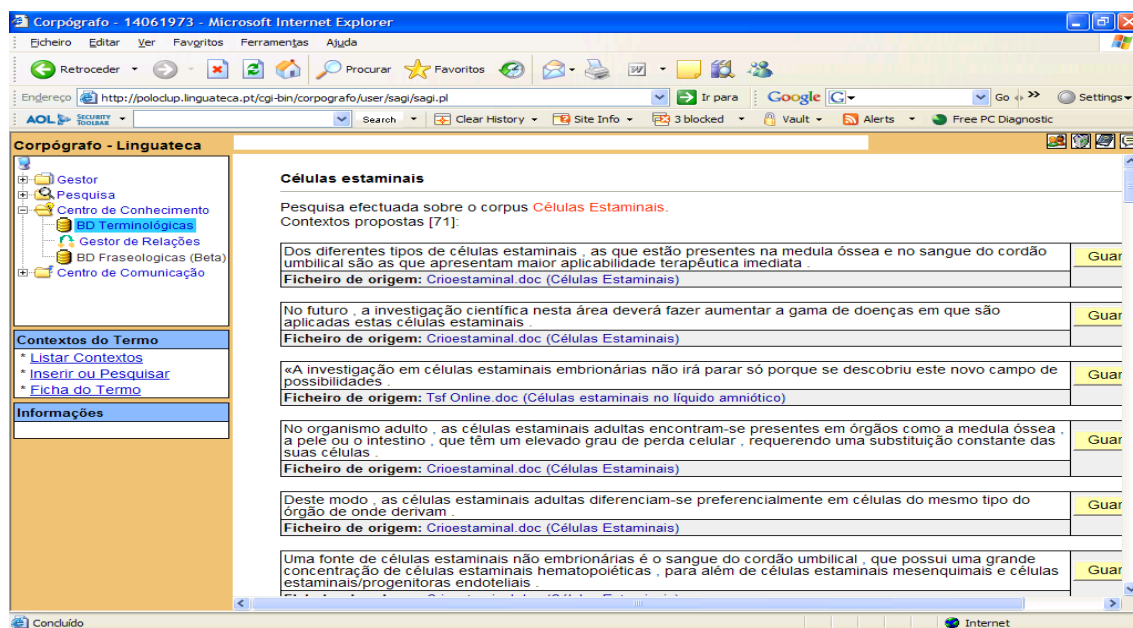


Fig.31 – Resultado de uma pesquisa de Contextos

A pesquisa de Relações Semânticas, usando restrições lexicais, revelou-se infrutífera para todos os termos armazenados tanto em Português como em Inglês.

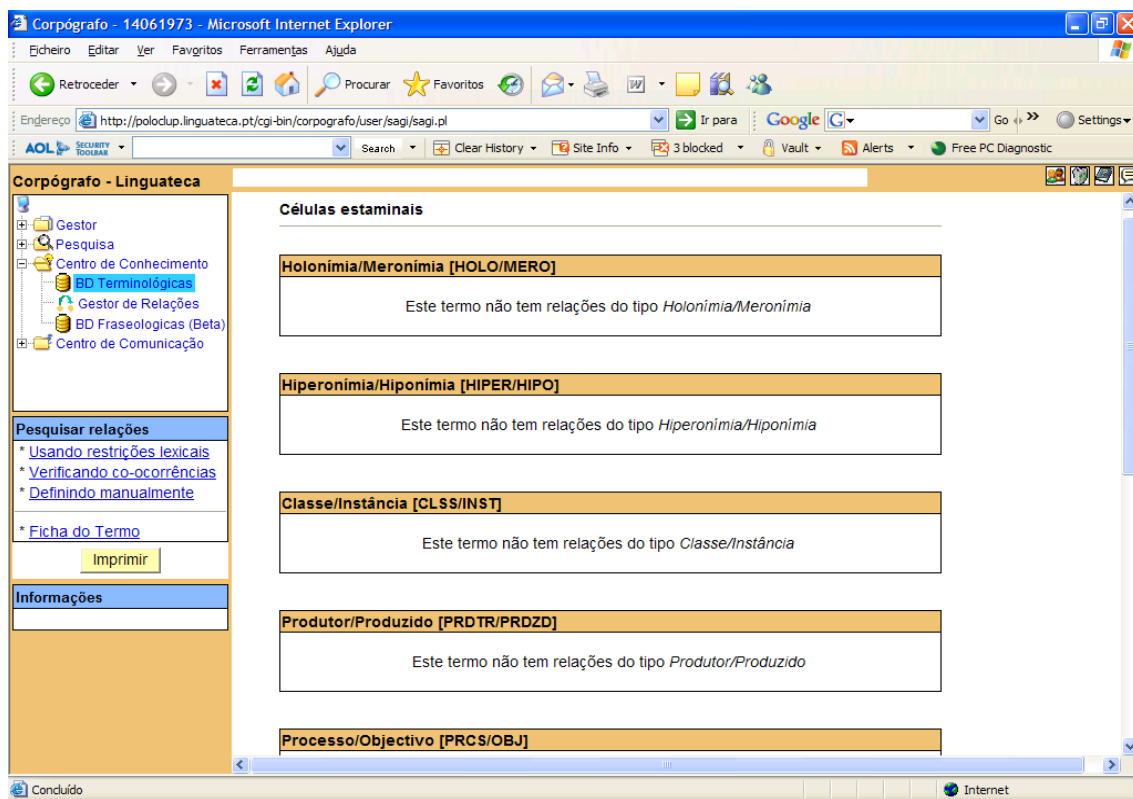


Fig.32 – Resultado de uma pesquisa de Relações Semânticas

Optou-se, então, por pesquisar as co-ocorrências dos termos, fazendo uma pesquisa sem restrições lexicais, para verificar se esse termo e os que co-ocorre estão ligados por alguma das relações semânticas previstas pelo Corpógrafo.

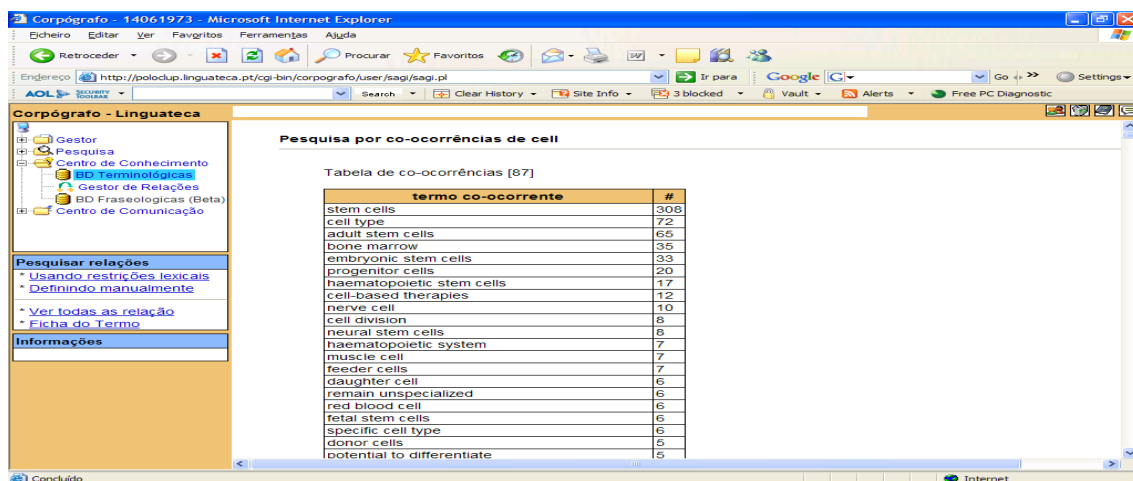


Fig.33 – Resultado de uma pesquisa de co-ocorrências

Por agora, finalizou-se a tarefa de extracção terminológica no *Corpógrafo*. É significativo o número de termos armazenados, tanto para Inglês como para Português. Esta tarefa permite-nos iniciar a elaboração do nosso glossário bilingue, com a ajuda das pesquisas de definições e de contextos, bem como da tarefa manual de estabelecimento de equivalentes de tradução. Contudo, este processo não tem verdadeiramente um fim, uma vez que as bases de dados podem ser constantemente enriquecidas com novos termos, à medida que novos textos vão sendo introduzidos.

## Conclusão

Este trabalho foi consagrado à produção de um glossário bilingue sobre Células Estaminais. Para isso, foi necessário extrair terminologia dos corpora elaborados para o efeito, recorrendo-se à ajuda das ferramentas electrónicas *Corpógrafo* e *Wordsmith Tools*. As ferramentas electrónicas seleccionadas, o *Corpógrafo* e o *Wordsmith Tools*, centram-se em noções muito simples e exploráveis por qualquer terminólogo que não tem que ser necessariamente um especialista em informática. Percorrendo, de forma metódica, os instrumentos incorporados no *Corpógrafo* e pontualmente no *Wordsmith Tools*, executou-se, para servir o propósito de extracção terminológica, uma série de tarefas para a exploração sistemática e semi-automática de um corpus especializado que se podem resumir: introdução e edição de textos, verificação do fraseamento, observação dos índices de frequência dos vocábulos que ocorrem nos textos, pesquisa de concordâncias, consideração de forma simples ou combinada dos termos que as listas nos proporcionam, criação de bases de dados terminológicas, edição de termos, pesquisa de definições e contextos, validação e o processamento de dados. As fases da metodologia seguida no *Corpógrafo* foram quase sempre sequenciais na identificação, análise e tratamento dos termos em evidência, tendo sido, no entanto, necessário voltar a fases anteriores para clarificar as informações linguísticas e os dados obtidos anteriormente. Após a extracção terminológica, elaborou-se o glossário do ponto de vista do conteúdo. Seguidamente, criou-se um índice hiperlinkado na própria página. Isto é, o glossário foi elaborado com hiperligações no interior do documento, usando como âncoras os títulos. Constituiu-se cada letra um título. Criou-se uma hiperligação de retorno ao menu principal, para facilitar a navegação. Assim, o menu principal é, por assim dizer, a carta de marear de todo o documento, sendo fácil de lá partir e lá chegar.

Existe, ainda, a necessidade de ampliar os textos explicativos que foram elaborados com a ajuda das pesquisas de contextos e definições do *Corpógrafo* – irá notar-se que existem letras e assuntos sem conteúdo, mas que de futuro se encherão de matéria, de substância. Deve, também, esclarecer-se que os títulos que aparecem no menu inicial reflectem a leitura atenta dos textos introduzidos no corpus bilingue. Isto é,

da análise textual percebeu-se uma organização de conteúdos dentro do âmbito das Células Estaminais. Assim, pensa-se que é de importância primária a definição das Células Estaminais, esclarecendo-se as razões pelas quais estas se distinguem das outras células, posteriormente é necessário falar dos tipos de células estaminais informando-se que estas podem ser adultas ou embrionárias e onde se podem localizar, de seguida apontar-se-á algumas razões pelas quais é tão importante investir na investigação desta área, mencionando-se a possibilidade destas células virem a curar doenças crónicas e degenerativas para as quais a medicina ainda não consegue dar uma resposta positiva, finalmente esclarecer-se-á que a investigação e o estudo do assunto em questão permitiu descobrir a técnica da clonagem, levantando a clonagem reprodutiva inúmeras questões éticas.

É, também, necessário trabalhar informaticamente o trabalho investido no glossário. O pesquisador de palavras ainda não funciona. Ainda se está a pensar no modo como se vai introduzir um motor de busca que direcione o usuário de imediato para o termo procurado ou que simplesmente o informe da inexistência de resultados, quando são colocados termos não contemplados pelo glossário. Sempre que possível, teve-se a preocupação de introduzir imagens para melhor ilustrar o termo introduzido. Pensou-se que as imagens tornariam o glossário mais cativante do ponto de vista estético, para que este não ficasse confinado apenas à densidade textual e, também porque, como se sabe, há pessoas que são muito visuais na apropriação de conhecimento. As imagens provêm de fontes variadas e livres na Internet, como a Wikipédia, a Medline e outros sites médicos sobre genética, clonagem, saúde pública, aos quais fomos direccionados através do motor de busca Google. É importante acrescentar que não existe qualquer diferença, para além da linguística, entre a versão portuguesa e a versão inglesa, sendo os termos seleccionados de uma língua os equivalentes de tradução da outra dentro do âmbito especializado em estudo.

Pensa-se que o glossário bilingue é representativo da boa estruturação do trabalho, pois permitiu obter resultados úteis para todos aqueles interessados na área em estudo. O trabalho realizado, põe em manifesto a importância do subdomínio seleccionado, ainda que estejamos conscientes de que a tarefa não está completa. Pode concluir-se que os corpora electrónicos e as ferramentas utilizadas para a sua exploração foram de grande utilidade, permitindo que o nosso projecto terminológico fosse bem

sucedido. Contudo, por muito útil e prático que os corpora desta natureza possam ser, não estão isentos de dificuldades. A primeira das quais se prende com a acessibilidade das fontes de documentação na língua portuguesa. Por razões por demais evidentes, é mais difícil localizar e elaborar corpus em Português do que em Inglês.

Um tradutor acreditado, para além de uma preparação rigorosa em diversos campos do saber, para além da capacidade de estar familiarizado com o tema e reconhecer o léxico próprio de um campo de saber especializado, de possuir bons conhecimentos das características dos textos e das convenções que os regem, bem como das relações de intertextualidade que estabelecem com outros textos e para além de possuir a capacidade de produzir textos adequados às diferentes situações comunicativas, tendo em conta os padrões das línguas e das culturas com as quais se trabalha, terá, também, que contar com o seu instinto pessoal, para que as suas decisões estabelecidas *ad hoc*, que se definem e delimitam durante o processo de análise e tradução, validem cada caso concreto, cada objectivo, cada situação. Pois, como aponta a Professora Elena Sánchez Trigo, no artigo “Aproximación traductológica al análisis de corpus para el estudio de las convenciones textuales”, “cada texto es único, ya que es el resultado de una situación comunicativa determinada e irrepetible”. (2003: 113) Neste sentido, não só são fundamentais os aspectos linguísticos, a análise das convenções textuais, como também não se pode descurar os aspectos extralinguísticos, a situação comunicativa e a intenção do autor que se esconde por trás da trama textual. Aqui, o instinto pessoal do tradutor impõe-se como fundamental, uma vez que nenhum trabalho obedece a leis fixas como se de uma normativa se tratasse. É antes um processo de negociação entre vários factores já enumerados, a argumentação que explica uma decisão em detrimento de outra. Após o estudo que se efectuou conclui-se que não existem princípios predeterminados e de aplicação universal.

Na óptica de Gideon Toury (1995), as traduções cumprem a sua função comunicativa se transmitirem a informação que o produtor deseja expressar e se influenciarem de algum modo o comportamento dos receptores. Os corpora que se apresentam são ainda reduzidos, uma pálida imagem da ampliação que se ambiciona fazer no futuro próximo. Os resultados que se podem obter da análise de um número reduzido de textos são sempre representativos de uma porção limitada da realidade. Nessa perspectiva, as conclusões serão sempre parciais e restritas. Contudo, o facto da

análise ser parcial não implica que o estudo efectuado seja anti-científico. É antes um processo de contínua actualização e, portanto, histórico porque não é absoluto e é progressivo e alterável.

Neste sentido, termina-se este trabalho como se começou, reiterando-se a importância de se manter uma mente aberta em tudo e, também, em estudos de tradução. Concorda-se que a primeira unidade mínima de sentido na tradução são os termos, de extrema importância sobretudo em domínios de saber muito específicos, como é o caso das Células Estaminais. Pensa-se que o vocabulário específico de cada área é por demais importante e deve ser objecto da nossa preocupação inicial. A validação desses termos é decisiva e crucial para a obtenção de um produto sério. Para proceder à validação do sistema conceptual e dos conceitos inerentes, bem como dos termos que os designam é fundamental que o tradutor solicite a colaboração de um especialista da área, tendo sempre presente a distinção entre termo e conceito, não esquecendo que o primeiro habita no domínio linguístico e o segundo pertence ao domínio extralinguístico. Cada termo deve ser validado por um número ímpar de especialistas e os resultados obtidos devem ser geridos pelo terminólogo. Os termos rejeitados no processo de validação devem ser excluídos e os termos sobre os quais foram levantadas dúvidas devem ser submetidos a nova validação.

O nosso trabalho está elaborado nesse sentido, como um ponto de partida de abordagem do TO. Depois de dominada a terminologia, é altura de conquistar sentido a sentido, analisar segmento a segmento, estabelecendo as devidas equivalências entre as línguas implicadas, o que não é mais do que o conceito de “manageable chunks” de Toury (1986: 83) – a porção ou quantidade de TO que o tradutor pode processar independentemente da sua categoria sintáctica, extensão (que pode variar) e do tipo de tradução (se mais ou menos literal).

## Referências Bibliográficas

As referências bibliográficas organizaram-se, para facilitar a sua consulta, em dois capítulos. Em primeiro lugar, recolheram-se aquelas que se citaram ao longo do presente trabalho de investigação. Em segundo lugar, reuniram-se as referências dos textos que constituem o corpus compilado, classificadas por línguas (inglês e português).

### I. Referências Citadas

AARTS, Jan (Ed.). 1990. *Theory and Practice in Corpus Linguistics*, Amsterdam/Atlanta: Rodopi.

AIJMER, Karin (Ed.). 1991. *English Corpus Linguistics*, London: Longman.

AUSTERMUHL, Frank. 2001. *Electronic Tools for Translators*, UK: T.J.International Ltd.

BAKER, Mona (1993): "Corpus Linguistics and Translation Studies: Implications and Applications" In Mona Baker, Gill Francis, and Elena Tognini-Bonelli (Eds), *Text and Technology: in Honour of John Sinclair*, Amsterdam and Philadelphia, John Benjamins, pp. 233-250.

CABRÉ, M. Teresa. 1993. *La Terminología: teoría, metodología, aplicaciones*. Traducción castellana de Carles Tebé. Barcelona: Editorial Antártida/Empúries.

CABRÉ, M. Teresa. 1999. *La Terminología: representación y comunicación*, Barcelona: Institut Universitari de Lingüística Aplicada.

CHESTERMAN, Andrew. 1997. *Memes of Translation: the spread of ideas in translation theory*, Amsterdam: John Benjamins, cop.

CORPAS, Pastor, Gl. 2002. "Traducir com corpus: de la teoria a la práctica". In J. Garcia Palácios y M<sup>a</sup> T. Fuentes (eds.), *Texto; terminologia y Traducción*, Salamanca: Almar.

- CORPAS, Pastor, Gl. 2004. “Localización de recursos y compilación de corpus via Internet: Aplicaciones para la didáctica de la traducción”. In C. Gonzalo Garcia y V. Garcia Yebra, *Manual de Documentación y Terminología para la Traducción Especializada*, Madrid: Arco/ libros.
- COSTA, Rute. 2001. “O Termo como Veículo de Especificidades Conceptuais e Semânticas”. In *Polifonia*, Lisboa: Edições Colibri, p.p. 199-204.
- COSTA, Rute. 2005. “Plurality of theoretical approaches to terminology”. In *Proceedings of LSP 2005: New Trends in Specialized Discourse*, Bergamo: Springer-Verlag.
- DELABASTITA, D. 1996. *Worldplay and Translation: special issue dedicated to the memory of André Lefevere (1945-1996)*, Manchester: St. Jerome.
- FRIES, Tottie and Schneider (Eds.). 1994. *Creating and Using English Language Corpora*, Amsterdam/ Atlanta: Rodopi.
- GAUDIN, François. 1993. *Socioterminologie: des problemes semantiques aux pratiques institutionnelles*. Rouen, France: Publications de L' Université de Rouen.
- GUESPIN, Gaudin. 2000. *Initiation à la Lexicologie Française*, Bruxelles: De Boeck & Larcier s.a.
- HERNÁNDEZ, M. Chantal Pérez. 2000. *Explotación de los Corpora Textuales Informatizados para la Creación de Bases de Datos Terminológicas Basadas en el Conocimiento*. Universidad de Málaga. (Em linha) Consultado a 20 de Dezembro de 2009. Disponível em <<http://elies.rediris.es/elies18/>>
- HOLMES, J. S. 1988 [1972]. “The Name and Nature of Translation Studies”. In R. Van den Broeck (ed.), *Translated!: Papers on Literary Translation and Translation Studies*, Amsterdam: Rodopi, p.p. 67-80.
- JOHANSSON, S. 1991. *English Computer Corpora*, New York: Monton de Gruyter.
- JOHANSSON, S. e S. Okesefjell. 1998. *Corpora and Cross-Linguistics Research. Theory, Method and Case Studies*, Amsterdam/ Atlanta: Rodopi.
- KENNEDY, Graeme. 1998: *An Introduction to Corpus Linguistics*, London: Longman.

- LAVIOSA, S. 1998. "The Corpus-Based Approach: A New Paradigm in Translation Studies", *Meta*, 43(4), *Special Issue*. (<http://www.erudit.org/revue/meta/1998/v43/n4/index.html>) [Consulta: 15.02.10].
- LEFEVERE, André. 1998. *Constructing Cultures: Essays on Literacy Translation*, London: Longman.
- LEHMANN, Martin-Berthet. 2000. *Introduction à la Lexicologie – Sémantique et Morphologie*, Paris: Éditions Nathan/HER.
- LERAT, Pierre. 1995. *Les Langues Spécialisées*, Paris: Presses Universitaires de France.
- MOSSOP, Brian. 2001. *Editing and Revising for Translators*, UK: T.J.International Ltd.
- RENOUF, Antoinette (Ed.). 1998. *Explorations in Corpus Linguistics*, Clevedon: Multilingual Matters, cop.
- SÁNCHEZ TRIGO, E. 2001. *Teoría de la Traducción. Convergencias y divergencias*, Vigo: Universidad de Vigo, SECO, M. et alii, *Diccionario del Español actual*, 2 vols. Madrid, Aguilar Lexicografía, 1999.
- SECO, M. et alii. 1999. *Diccionario del Español actual*, 2 vols. Madrid, Aguilar Lexicografía.
- SOUTER, Clive (Ed.). 1993. *Corpus-Based Computational Linguistics*, Amsterdam/Atlanta: Rodopi.
- TOURY, Gideon. 1995. *Descriptive Translation Studies and Beyond*, Amsterdam: John Benjamins, cop.
- TYMOCZKO, M. 1998. "Computerized corpora and the future of Translation Studies", *Meta*, 43(4), *Special Issue*. (<http://www.erudit.org/revue/meta/1998/v43/n4/index.html>) [Consulta: 15.02.10].
- ZANETTIN, F. 1998. "Bilingual Corpora and the Training of Translators" in S. Laviosa, *The Corpus-based Approach*, *Meta*, 43(4), *Special Issue*. (<http://www.erudit.org/revue/meta/1998/v43/n4/index.html>) [Consulta: 15.02.10].

## **II. Fontes Textuais do Corpus**

### **II.1. Inglês**

Pessina A, Gribaldo L. The key role of adult stem cells: therapeutic perspectives.

*Highwire Press* 2006; 22(11): 2287-300 [<http://highwire.stanford.edu/cgi/medline/pmid;17076989>][Medline Abstract]

Petersen Bryon E., Terada Naohiro. Stem Cells: A Journey into a New Frontier. *JASN*

*Science Watch* 2001; 12:1773-1780 [<http://jasn.asnjournals.org/cgi/content/full/12/8/1773>] [Abstract/Free Full Text].

Kuehnle Ingrid, Goodell Margaret A. The therapeutic potential of stem cells from

adults. *BMJ* 2002; 325:372-376 [<http://www.bmj.com/cgi/content/full/325/7360/372>] [Abstract/Free Full Text].

Stem Cell Basics. In Stem Cell Information. Bethesda, MD: National Institutes Of

Health, U.S., 2006 [<http://stemcells.nih.gov/info/basics/basics3.asp>]

Stem Cell Science and Medical Updates. Stem Cell Research. Maryland,U.S., 2006; 1-

877-842-3442 [<http://www.stemcellresearchfoundation.org>]

Stem Cells. Wikipedia Foundation, Inc.,U.S.,

2007[[http://en.wikipedia.org/wiki/Stem\\_cells](http://en.wikipedia.org/wiki/Stem_cells)]

### **II.2. Português**

Criopreservação de Células Estaminais do Sangue do Cordão Umbilical. Crioestaminal.

Coimbra, 2006 [[http://www.crioestaminal.pt/imagens/brochura\\_2006.pdf](http://www.crioestaminal.pt/imagens/brochura_2006.pdf)]

Células Estaminais. Wikipédia Foundation, Inc., 2007 [<http://pt.wikipedia.org/wiki/C%C3%A9lula-tronco>]

Relatório sobre a Investigação das Células Estaminais Embrionárias Humanas.

*Comissão das Comunidades Europeias*. 2003 [[http://ec.europa.eu/research/conferences/2003/bioethics/pdf/executive-summary\\_pt.pdf](http://ec.europa.eu/research/conferences/2003/bioethics/pdf/executive-summary_pt.pdf)]

[Abstract/Free Full Text].

Células Estaminais do Sangue do Cordão Umbilical. *Cytothera. Grupo Medinfar*. 2006  
[\[http://www.medinfar.pt/medinfaronline/pt/grupo/biotecnologia/cytotherababy/celulasestaminais/\]](http://www.medinfar.pt/medinfaronline/pt/grupo/biotecnologia/cytotherababy/celulasestaminais/).

Terapias com Células Estaminais. *Viver Melhor em Boa Forma*. 2002007  
[\[http://www.crioestaminal.pt/imagens/Viver%20Melhor%20em%20Boa%20Forma\\_122006.pdf\]](http://www.crioestaminal.pt/imagens/Viver%20Melhor%20em%20Boa%20Forma_122006.pdf).

Nova Fonte de Células Estaminais. TsF Online. 2007 [\[http://www.tsf.pt/online/ciencia/interior.asp?id\\_artigo=TSF176690\]](http://www.tsf.pt/online/ciencia/interior.asp?id_artigo=TSF176690).

*Cientista Português Desenvolve Técnica de Clonagem Mais Simples. Ciência Viva*. Agência Nacional Para a Cultura Científica e Tecnológica. 2004  
[\[http://www.cienciaviva.pt/imprensa/index.asp?acao=showartigo&id\\_media\\_artigo=319\]](http://www.cienciaviva.pt/imprensa/index.asp?acao=showartigo&id_media_artigo=319).