



Maria Rute Vilhena Costa

**Pressupostos teóricos e metodológicos  
para a extracção automática  
de unidades terminológicas multilexémicas**

Universidade Nova de Lisboa • Faculdade de Ciências Sociais e Humanas

Lisboa, Agosto 2001



Maria Rute Vilhena Costa

**Pressupostos teóricos e metodológicos  
para a extracção automática  
de unidades terminológicas multilexémicas**



Dissertação para obtenção do grau de Doutor  
em Linguística, especialidade de Lexicologia - Terminologia  
realizada sob a orientação da  
Professora Doutora Maria Teresa Rijo da Fonseca Lino

801'373

Universidade Nova de Lisboa • Faculdade de Ciências Sociais e Humanas

Lisboa, Agosto 2001

55309

*Aos meus pais*

## **Agradecimentos**

*Ao terminar este trabalho não posso deixar de agradecer a todos os que contribuíram para a sua realização, cujo contributo foi essencial, dentro e fora do âmbito académico.*

*À Professora Doutora Maria Teresa Rijo da Fonseca Lino, pela orientação da dissertação, pela oportunidade de discutir ideias obtendo sugestões e esclarecimentos de grande valor, pela total confiança e a liberdade de escolha científica que sempre me concedeu; pela aposta que um dia fez em mim; pelos seus ensinamentos ao longo da minha progressão universitária e pela lição de humanismo que quotidianamente me transmite, o meu muito obrigada.*

*À Professora Doutora Maria Francisca Xavier e à Professora Doutora Maria de Lourdes Crispim agradeço-lhes o apoio incondicional à candidatura a uma Bolsa PRODEP.*

*Ao Professor Doutor Didier Bourigault agradeço ter aceite deslocar-se a Lisboa com o intuito de debater questões científicas de grande importância para o desenvolvimento desta dissertação.*

*Ao Eng.º Pedro Santos agradeço a grande disponibilidade para o apoio ao nível da programação informática.*

*À Drª. Helena Manuelito, pela leitura minuciosa que fez deste trabalho, os meus muito sinceros agradecimentos.*

*À Drª. Raquel Silva agradeço, para além da amizade, todo o apoio, disponibilidade e vontade incansável para ajudar.*

*À Drª Fátima Ferreira agradeço a atenção sempre demonstrada, nomeadamente para colaborar nas obrigações comuns de relações universitárias internacionais.*

*Ao Dr. Paulo Morgado Sousa e ao Dr. José Carlos Ferreira reconheço e agradeço o espírito de solidariedade que sempre demonstraram.*

*À Maria João Patinha agradeço o precioso e contínuo apoio demonstrado em todas as circunstâncias da vida.*

## ÍNDICE

### **Introdução**

- |                                |    |
|--------------------------------|----|
|                                | 5  |
| 1. Objectivos                  | 5  |
| 2. Apresentação da dissertação | 12 |

### **Capítulo I - Corpora**

- |  |    |
|--|----|
|  | 16 |
| 1. Linguísticas de <i>corpora</i>  | 16 |
| 1.1 Definição de « <i>linguística de corpora</i> »                             | 16 |
| 1.2 Arquivo e <i>corpus</i>  | 19 |
| 2. Tipologia de <i>corpora</i>   | 26 |
| 2.1 <i>Corpora</i> de textos integrais e <i>corpora</i> de extractos de textos | 28 |
| 2.2 <i>Corpora</i> comparativos e <i>corpora</i> paralelos                     | 29 |
| 2.3 « <i>Monitor corpora</i> »   | 30 |
| 2.4 <i>Corpora</i> de referência   | 31 |
| 2.5 <i>Corpora</i> de especialidade e <i>corpora</i> especiais                 | 35 |
| 3. <i>Corpora</i> anotados   | 37 |
| 3.1 SGML   | 39 |
| 3.2 TEI  | 45 |
| 3.3 CES  | 48 |
| 3.4 Etiquetagem gramatical   | 50 |

### **Capítulo II - Constituição de *corpus***

- |   |     |
|---|-----|
|   | 57  |
| 1. Comunidade científica, produtora de textos de especialidade      | 57  |
| 2. Texto de especialidade   | 60  |
| 2.1 Texto e discurso  | 60  |
| 2.2 Contextos   | 67  |
| 2.3 Intervenientes  | 72  |
| 3. Tipologia de tipos ou tipologia de géneros?                      | 77  |
| 4. Constituição do <i>corpus</i> em Detecção Remota                 | 90  |
| 4.1. Definição da área de especialidade em análise: Detecção Remota | 90  |
| 4.2. Constituição de tipologias                                     | 96  |
| 4.2.1. Tipologia de tipos de discurso                               | 96  |
| 4.2.2. Tipologia de géneros de discurso                             | 102 |

### **Capítulo III - Denominação: expressões nominais multilexémicas** ..... **107**

1. Denominação	107
2. Expressões nominais multilexémicas	113
2.1. Lexias	113
2.2. Sinapse	117
2.3. Nome composto	119
2.4. Colocação	128
2.5. Frasemas	145
3. Fraseologia	148

### **Capítulo IV - Etiketador gramatical automático: EtiqueLex** ..... **156**

1. Tratamento do <i>corpus</i> informatizado	156
2. Tabela de etiquetas	158
3. Constituição dos dicionários	166
3.1. Dicionário de formas simples flexionadas	166
3.1.1. Desdobramento de formas	166
3.1.2. Variantes ortográficas	168
3.2. Dicionário de locuções	171
3.3. Dicionário de unidades terminológicas multilexémicas da <i>Deteção Remota</i>	175
4. Etiketador gramatical: EtiqueLex	178
5. Tipologia de matrizes terminogénicas	182
5.1. Tipologia elementar	182
5.2. Análise da composição das unidades terminológicas multilexémicas	186
5.3. Tipologia de base	197
5.4. Dicionário de estruturas tipo	200

### **Capítulo V - Extractor automático de unidades terminológicas multilexémicas: ExtractTerm** ..... **203**

1. Concepção do ExtractTerm	203
1.1. Aplicação do dicionário de tipologias	203
1.2. Análise dos resultados	207
2. Regras de desambiguação	216
2.1. Preposição, artigo definido feminino singular e pronomes	219

2.1.1. Regras linguísticas de aprendizagem	221
2.2. Adjectivo e particípio passado pós-nominal.	224
2.2.1. Regras linguísticas de aprendizagem	226
2.3. Nome e adjectivo	236
2.3.1. Regras linguísticas de aprendizagem	238
3. ExtracTerm: utilização das regras de aprendizagem	240
3.1. Análise dos resultados	244
3.2. Teste do ExtracTerm sobre um <i>corpus</i> alargado	247
3.2.1. Análise dos resultados	248
<b>Conclusão</b>	<b>251</b>
<b>Bibliografia</b>	<b>258</b>
<b>Anexos</b>	<b>301</b>
<b>Anexo 1</b> - Dicionário de formas flexionadas	303
<b>Anexo 2</b> - Dicionário de locuções	314
<b>Anexo 3</b> - Dicionário de unidades terminológicas da detecção remota	323
<b>Anexo 4</b> - Texto etiquetado com o EtiqueLex, sem a aplicação de regras de desambiguação	326
<b>Anexo 5</b> - Sequências extraídas com ExtracTerm a partir de texto etiquetado não desambiguado	332
<b>Anexo 6</b> - Texto etiquetado com o ExtracTerm, com a aplicação de regras de desambiguação	339
<b>Anexo 7</b> - Sequências extraídas com ExtracTerm a partir de texto etiquetado desambiguado	345
<b>Anexo 8</b> - Texto etiquetado com ExtracTerm, a partir de novo <i>corpus</i> , com aplicação de regras de desambiguação	351
<b>Anexo 9</b> - Sequências extraídas com ExtracTerm a partir de texto etiquetado desambiguado	358

# Introdução

---

1. Objectivos
2. Apresentação da dissertação

# Introdução

## 1. Objectivos

O tratamento automático da língua produzida em contexto de especialidade é o pano de fundo que subjaz à reflexão teórica e metodológica apresentada nesta dissertação.

O objectivo da nossa investigação consiste, desta forma, na formulação dos pressupostos teóricos da Lexicologia e da Terminologia que estão na base da concepção de um programa informático destinado ao levantamento automático a partir de descrições de análises de carácter linguístico de unidades terminológicas multilexémicas.

Com efeito, desenvolvemos neste trabalho uma investigação que, baseada em critérios exclusivamente linguísticos permite a identificação automática de unidades terminológicas multilexémicas da área de especialidade da Detecção Remota.

No momento da concepção do extractor, tomámos consciência da necessidade da criação de um etiquetador que está na base do primeiro. Deste modo, concebemos dois programas: o EtiqueLex e o ExracTerm. O primeiro atribui etiquetas gramaticais aos lexemas do *corpus*. Funcionando sobre o *corpus* etiquetado, o ExracTerm extrai unidades terminológicas multilexémicas, partindo de regras de reconhecimento e de regras de aprendizagem, concebidas e fundamentadas a partir da observação do *corpus*.

O EtiqueLex é uma componente indispensável do ExtracTerm, sem a qual este último não funciona. Não obstante, o EtiqueLex pode ser utilizado exclusivamente para a etiquetagem, sem que seja necessário proceder-se ao levantamento automático de termos, gozando, assim, de autonomia face ao extractor.

A emergência e o desenvolvimento de uma abordagem lexicológica à Terminologia coloca-nos no centro dos debates actualmente em curso. Neste âmbito, levantam-se questões como as seguintes: o que distingue uma palavra de um termo? Que metodologia(s) adoptar em Terminologia?

Actualmente, a Terminologia é uma área de especialidade no seio da Linguística que estuda o comportamento linguístico das unidades terminológicas, recorrendo aos contextos e, de forma mais abrangente, aos textos em que ocorrem.

Neste debate entram em confronto duas ideias distintas: por um lado, a perspectiva que defende uma abordagem normativa à terminologia, por outro, a que encara a terminologia numa perspectiva descritiva.

A doutrina wüsteriana é apologista da normalização terminológica, porque a univocidade dos termos é uma condição necessária à comunicação entre especialistas, sendo a redução da sinonímia e da ambiguidade um dos principais objectivos.

Foi nos anos trinta que Wüster levou a cabo trabalhos terminológicos dando especial relevo à organização sistemática das terminologias, incrementando a definição de postulados fundamentais para o desenvolvimento de métodos de trabalho. Porém, é na URSS, pela mão de Lotte, responsável do Comité de

Normalização Terminológica do Instituto de Normalização do Conselho dos Ministros da URSS e membro da Academia das Ciências, que a terminologia nasce como ciência (cf. Rondeau, 1983).

Em consequência das preocupações metodológicas de Wüster e das preocupações teóricas de Lotte, é criado um organismo, o ISA - *International Organization for Standardization*, cujo objectivo é o favorecimento do comércio internacional, desenvolvendo a normalização dos produtos e dos processos. Os membros fundadores deste organismo são a França, a Grã-Bretanha, a Alemanha e a URSS. A 2ª Guerra Mundial interrompe as actividades da ISA, mas em 1946 nasce a ISO - *International Organization for Standardization*, sendo criada em 1951 o Comité Técnico 37 (TC 37) que, reunindo pela primeira vez em 1952, tem por função a normalização da Terminologia e dos seus princípios metodológicos.

A metodologia preconizada pelos seguidores de Wüster parte da análise do objecto para chegar à denominação, uma vez que é, essencialmente, através desta que os especialistas veiculam conhecimentos. O conceito está no centro da metodologia onomasiológica, desempenhando a denominação a função de «etiqueta»: «[...] *la démarche terminologique, au contraire de la démarche linguistique, ne part pas d'une appellation pour découvrir l'être ou le groupe d'êtres représenté par cette appellation; elle procède à l'inverse, à partir des entités pour étudier leurs dénominations.*» (Rondeau, 1983:12).

A normalização do termo passa pela normalização do conceito, pressupondo tal abordagem a sua universalidade, uma vez que para a ISO os conceitos não estão ligados a cada uma das línguas individualmente. Tal como

refere Lerat (1995), os linguistas não podem aceitar esta perspectiva, porque: «[...] *les notions sont liées aux conceptualisations, donc aux locuteurs, à leurs cultures et à leurs langues*» (Lerat, 1995:17).

A rápida evolução da técnica, das ciências e da tecnologia fazem com a conceptualização dos factos científicos não seja universal, porque dependente de factores sociais e culturais evidentes. Por esta razão, linguistas como Béjoint e Thoiron rejeitam a doutrina wüsteriana, perspectivando a normalização como um processo artificialmente construído, com vista a atingir o ideal da biunivocidade: um termo para um conceito: «[...] *bref de la langue fabriquée de toutes pièces et contrôlée par la communauté linguistique pour façonner le monde*» (Béjoint; Thoiron, 2000:6).

Do ponto de vista teórico e metodológico, defendemos uma abordagem descritiva à Terminologia. Numa perspectiva linguística, a automatização requer não só a descrição do comportamento linguístico das denominações em situação real, como também a descrição dos contextos em que tais denominações ocorrem.

A evolução da capacidade de armazenamento da informação em suporte electrónico, bem como o aumento crescente dos recursos linguísticos plurilingues informatizados contribuem para as viragens metodológica e teórica que neste momento atingem claramente a terminologia.

Os textos escritos que compõem os *corpora* são o meio mais efectivo para os especialistas transmitirem e divulgarem o seu conhecimento quer no seio da comunidade científica a que pertencem quer fora dela. Tais textos são um

potencial inegável para o levantamento de terminologias, que reflectem o estado do conhecimento de uma determinada comunidade.

A quantidade de produção científica textual, tal como a facilidade de aceder a textos informatizados, é enorme, levando o terminólogo a reequacionar as suas metodologias. Assim, o texto passa a ser o centro das suas preocupações. O estudo das denominações, com recurso ao contexto linguístico em que estas ocorrem, leva-nos a preconizar a existência de uma terminologia textual: *«Les applications de la terminologie sont le plus souvent des applications textuelles (traduction, indexation, aide à la rédaction); la terminologie doit «venir» des textes pour mieux y «retourner». C'est parce qu'elle n'est jamais déliée du texte qu'on parle de «terminologie textuelle». C'est dans les textes produits ou utilisés par une communauté d'experts, qui se sont exprimées, et donc accessibles, une bonne partie des connaissances partagées de cette communauté, c'est donc par là qu'il faut commencer l'analyse.»* (Bourigault, Slodzian, 1999:30).

Em nosso entender, esta abordagem vem manifestamente contribuir para um desenvolvimento das análise e descrição em linguística, permitindo um desenvolvimento renovado de gramáticas e dicionários.

O levantamento e organização da terminologia é, actualmente, uma necessidade política, económica e industrial ao serviço da inovação e da Sociedade da Informação e do Conhecimento, âmbito em que as aplicações informatizadas são concebidas para fins específicos, sendo a identificação do público alvo uma prioridade. Deste modo, uma denominação tem uma acepção reconhecida e identificada pelos individuos pertencentes a uma comunidade

científica específica, sendo que: «*La tâche de description lexicale est un travail de fixation, de stabilisation, d'homogénéisation d'une signification, dont le résultat est le terme. [...] C'est ainsi qu'on parle de normalisation, non plus au sens que la planification terminologique donne au mot, mais au sens où la communauté d'experts «entérine» des signifiés comme des termes du domaine.*» (Bourigault, D., Slodzian, M. 1999:30).

Apesar de os procedimentos metodológicos propostos e de os programas informáticos desenvolvidos poderem servir qualquer área de especialidade, podemos interrogar-nos sobre a pertinência da escolha dos textos de Detecção Remota como *corpus* de experimentação do nosso trabalho.

Em primeiro lugar, a comunidade científica e os textos produzidos nesta especialidade apresentam uma grande identidade, mesmo quando as fronteiras técnicas e científicas dos discursos da Detecção Remota evidenciam alguma interdisciplinaridade com outros ramos do saber, nomeadamente com a Física, a Matemática, a Geografia e as Ciências da Terra (cf. Capítulo 3). Esta característica faz com que o *corpus* constituído seja seguramente representativo do universo de discursos produzidos.

Em segundo lugar, a Detecção Remota assistiu, em Portugal, nos últimos dez anos, a um desenvolvimento sem precedentes, quer ao nível do ensino quer da investigação. Esse facto traduziu-se na produção exponencial de textos de especialidade, nomeadamente a partir de meados da última década. Assim, foi-nos possível confrontar o *corpus* inicial com um *corpus* «final» de referência, com o objectivo de testar e de enriquecer as regras

criadas - por via do enriquecimento técnico e científico dos próprios discursos -, com vista à subsequente experimentação dos programas informáticos desenvolvidos.

Em terceiro lugar, a comunidade científica da Detecção Remota em Portugal tem vindo a reconhecer a necessidade de harmonização terminológica, no sentido de dar resposta às exigências dos discursos científico-pedagógicos. São um exemplo dessa preocupação, as conclusões do *Workshop* ROT'2000, subordinado ao tema «O Ensino da Detecção Remota em Portugal», que reuniu, em Lisboa, cerca de 70 especialistas em ensino destas matérias. Entre outras conclusões, esta reunião salientou a necessidade de harmonização linguística de termos, por forma a aumentar a qualidade do saber produzido e ministrado nas universidades portuguesas.

Em quarto lugar, a Detecção Remota detém um lugar de destaque entre as Tecnologias de Informação Geográfica para a Sociedade da Informação. A difusão de inovação, que está inevitavelmente associada à Sociedade da Informação, requer qualidade; essa qualidade passa não só pelo enriquecimento dos conteúdos disponibilizados via Internet - é disso exemplo o projecto GEOCID, promovido pela Centro Nacional de Informação Geográfica (CNIG), e o Programa para a Sociedade da Informação (POSI), em desenvolvimento ao abrigo do Quadro Comunitário de Apoio (QCA III) - como também pela utilização de um referencial técnico e científico aperfeiçoado do ponto de vista terminológico.

## 2. Apresentação da dissertação

A presente dissertação desenvolve-se em cinco capítulos. Nos três primeiros expomos os principais pressupostos teóricos que fundamentam e justificam as opções e orientações metodológicas que apresentamos nos dois últimos capítulos.

**Capítulo 1:** Neste capítulo, fazemos uma incursão pelas linguísticas de *corpora*. Definimos o conceito de *corpus*, distinguindo-o do de arquivo, antes de abordarmos os *corpora* informatizados. Em íntima associação com a noção de *corpora* informatizados, damos conta das mais recentes técnicas de anotação, ou seja, das diversas formas de acrescentar informação metalinguística aos lexemas, sublinhando particularmente o papel da etiquetagem gramatical.

**Capítulo 2:** Os *corpora* são constituídos por textos de especialidade produzidos por comunidades científicas. Descrevemos o que entendemos por comunidade científica, antes de justificarmos, do ponto de vista teórico, a selecção dos textos que constituem o *corpus* em análise. Debatemos, de igual forma, os conceitos de texto e de discurso, bem como o papel desempenhado pelos intervenientes activos e passivos na produção de textos. Salientamos a importância dos contextos linguísticos e extra-linguísticos na constituição de tipologias, clarificando, para o efeito, os conceitos de tipo e de género.

**Capítulo 3:** As denominações que nos interessam identificar e analisar são as unidades terminológicas multilexémicas. Neste ponto, debatemos a noção de expressões nominais multilexémicas e fazemos uma sinopse de algumas abordagens linguísticas às expressões nominais, que se reflectem na diversidade de terminologias utilizadas para as denominar.

**Capítulo 4:** Iniciamos, neste capítulo, a exposição da metodologia que nos permite proceder ao levantamento automático de unidades terminológicas multilexémicas.

A primeira fase consiste na elaboração de dicionários: dicionários de formas flexionadas, dicionário de locuções e dicionários de unidades terminológicas multilexémicas da detecção remota. A todas as formas que os constituem são atribuídas etiquetas metalinguísticas. O Etiquelex, etiquetador automático, adiciona etiquetas aos lexemas do *corpus* que consegue identificar.

Com base na observação do *corpus* etiquetado criamos uma tipologia de regras elementares e de regras base, que servem de fundamento ao dicionário de regras de reconhecimento.

**Capítulo 5:** A partir das regras de reconhecimento, o ExtracTerm, programa que extrai automaticamente unidades terminológicas multilexémicas, efectua um primeiro levantamento automático.

Com base nos dados observados, estabelecemos regras linguísticas de desambiguação, que nos permitem elaborar regras de aprendizagem. Com estas regras, o ExtracTerm opera sobre o texto etiquetado, levantando as

ambiguidades, antes de recorrer às regras de reconhecimento, que permitem um levantamento mais rigoroso das unidades terminológicas multilexémicas.

Através da reflexão teórica e metodológica que expomos nesta dissertação, esperamos contribuir para a renovação das teorias e metodologias em Terminologia no âmbito do tratamento automático da língua de especialidade.

# Capítulo I

---

## ***Corpora***

### 1. Linguísticas de *corpora*

#### 1.1 Definição de «*linguística de corpora*»

#### 1.2 Arquivo e *corpus*

### 2. Tipologia de *corpora*

#### 2.1 *Corpora* de textos integrais e *corpora* de extractos de textos

#### 2.2 *Corpora* comparativos e *corpora* paralelos

#### 2.3 «*Monitor corpora*»

#### 2.4 *Corpora* de referência

#### 2.5 *Corpora* de especialidade e *corpora* especiais

### 3. *Corpora* anotados

#### 3.1 SGML

#### 3.2 TEI

#### 3.3 CES

#### 3.4 Etiquetagem gramatical

## Capítulo I - Corpora

### 1. Linguísticas de corpora

#### 1.1. Definição de «linguísticas de corpora»

O corpus, objecto de estudo que está na origem das linguísticas de corpora, é um lugar de observação que permite a descrição de actualizações da língua organizadas em enunciados, discursos ou textos. Na base da constituição destes conjuntos de dados linguísticos estão critérios de selecção sistematizados, que facultam a legítima atribuição do estatuto de *corpus* a tais conjuntos de dados.

O facto de o termo «*corpus linguistics*» não ser utilizado nos textos anteriores a Chomsky, não significa que os corpora não fossem usados e explorados com a finalidade de análise linguística. Assim, recorrer aos corpora como objecto de análise não é um procedimento inovador. Em 1951, Harris considerava já o *corpus* o único objecto legítimo da linguística e designava por linguística estrutural a investigação que operava, *a priori* ou *a posteriori*, com corpora.

Aarts, por sua vez, considera que o conceito de «*corpus linguistics*» não dá conta de uma actividade totalmente nova em linguística: « [...] *if we take corpus linguistics as referring to linguistic research based on observed utterances, we can say that this type of research has a very long history indeed. Only in earlier days it was simply called linguistics of philology*» (Aarts, 1990:13).

Chomsky (1957, 1965) modificou o objecto da linguística, considerando que os *corpora* não poderiam nunca ser entendidos como objectos de análise úteis para o linguista; privilegia, recorrendo à introspecção uma aproximação racionalista ao objecto, em detrimento de uma aproximação empírica: «*Chomsky changed the object of linguistic enquiry from abstract descriptions of language to theories which reflected a psychological reality, cognitively plausible models of language*» (McEnery, Wilson, 1997: 4). Para este autor, os *corpora* dão conta exclusivamente dos actos de *performance*, não revelando os actos de competência, que podem ser unicamente determinados pelo falante. No que concerne ao *corpus*, McEnery e Wilson sintetizam o posicionamento de Chomsky do seguinte modo: «*A corpus is by its very nature a collection of externalised utterances; it is performance data, and such it must of necessity be a poor guide to modelling linguistic competence*» (1997:5).

Nos anos sessenta, as metodologias e as teorias associadas aos *corpora* ganham uma nova dinâmica. Vários autores (Aarts, 1990; Leech, 1997; Habert, 1997) consideram esta década um marco na história recente das linguísticas de *corpora*: «*The year of 1961, which more famously saw then first manned space flight, is the date to which corpus linguistics can look back as the date when the enterprise now known as corpus linguistics (or more precisely computer corpus linguistics) came into being*» (Leech, 1997:1).

É a escola anglo-saxónica que populariza o termo «*corpus linguistics*», que recobre simultaneamente o objecto, bem como as metodologias e as teorias que se constroem a partir dos *corpora*. A própria definição de *corpus* é actualizada tendo por referência o suporte em que este é armazenado, o seu

formato electrónico, incrementando as perspectivas de análise: «*In the past thirty-five years, the term corpus has been increasingly applied to a body of language material which exists in electronic form, and which may be processed by computer for various purposes such as linguistic research and language engineering*» (Leech, 1997:1).

Também Sinclair (1996) propõe uma definição de «*computer corpus*», independente da de *corpus*: «*A computer corpus is a corpus which is encoded in a standardised and homogenous way for openended retrieval tasks. Its constituent pieces of language are documented as to their origins and provenance*» (Sinclair, 1996:6). Tal definição pressupõe, implicitamente, que o *corpus* seja entendido como: «*[...] 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*» (Sinclair, 1996:6).

O *corpus* informatizado pode apresentar-se sob duas formas, isto é, na sua forma bruta («*raw corpus*») ou anotada. Enquanto que o *corpus* bruto é um objecto para testar hipóteses «*[...] the test bed for his hypotheses about the structure of the language, which he has expressed in a formal grammar*» (Aarts, 1990:18), o *corpus* anotado é enriquecido com informação de diversa natureza: morfológica, sintáctica, semântica, prosódica, crítica, etc., e «*[...] serves as a linguistic database for all linguists studying the structure of the language,[...]*» (Aarts, 1990:18).

Pensamos, deste modo, que as linguísticas de *corpora* assumem um duplo estatuto: por um lado, de sub-disciplina no seio da linguística, por outro, funcionam como disciplina auxiliar para todas as restantes disciplinas da

linguística: «*It creates textual databases which have been enriched with detailed morphological and syntactical information and, where possible, with phonological and semantic information. Within the foreseeable future, every linguist will be able to make use of such databases, which means he will also have at immediate numerical data about the use of constructions and sentence patterns, the realisation of grammatical sentences, etc.*» (Aarts, 1990:16).

Para que os resultados obtidos a partir de *corpora* sejam fiáveis, é indispensável que o objecto sobre o qual recaem as nossas hipóteses seja adequadamente definido e delimitado. Com o aumento crescente e variado dos *corpora*, surge a necessidade de reflectirmos adequadamente, por um lado, sobre as características do *corpus* merecedor dessa designação, por outro saber como classificar a diversidade resultante de tal proliferação.

## 1.2 Arquivo e *corpus*

Possuir uma colecção de textos informatizados não é condição suficiente para que possamos considerar estar em presença de um *corpus*; para o constituir é necessário ter em conta um conjunto de pressupostos teóricos e metodológicos considerados de importância fundamental.

Com base nestes princípios, sentimos, numa primeira instância, a necessidade de distinguir *corpus* de arquivo, conceitos que, frequentemente, são usados de forma arbitrária.

Em 1969, Foucault dedica grande parte da sua obra «*Archéologie du savoir*» à definição e à compreensão do conceito de arquivo, abordando para o

efeito os discursos científicos característicos de cada época, centrando a sua reflexão na relação que cada um dos enunciados mantém entre si, com o objectivo de se concentrar na organização interna do conhecimento, secundarizando os conteúdos particulares veiculados pelo mesmo.

Acredita, assim, que o arquivo é «*le domaine des choses dites*». Para chegar a esta concepção de arquivo, Foucault (1969) aborda a questão das formações discursivas, que define como conjuntos de enunciados, historicamente demarcados utilizando o mesmo sistema de regras. Para este autor, a unidade dos enunciados que, segundo ele, se organizam em famílias, tais como a Medicina, a Economia ou a Biologia, é identificável através da descrição das dissemelhanças que caracterizam os mesmos, constituindo sistemas de dispersão inerentes às formações discursivas: «*Une telle analyse n'essaierait pas d'isoler, pour en décrire la structure interne, des îlots de cohérence; elle ne se donnerait pas pour tâche de soupçonner et de porter en pleine lumière les conflits latents; elle étudierait des formes de répartition*» (Foucault, 1969:52). Foucault recorre ao conceito de formação discursiva para contornar noções tão genéricas e abrangentes como as de ciência, ideologia e teoria (cf. Foucault, 1969:53), que considera, no que concerne as suas implicações, serem demasiadamente complexas: «*Dans le cas où on pourrait décrire, entre un certain nombre d'énoncés, un pareil système de dispersion, dans le cas où entre les objets, les types d'énonciation, les concepts, les choix thématiques, on pourrait définir une régularité, [...], on dira, par convention, qu'on a affaire à une formation discursive*» (Foucault, 1969:53). Assim, a formação discursiva compreende o conceito de enunciado, cujo espaço de

comunicação é restrito comparativamente ao espaço de comunicação da ciência, entendida aqui na sua acepção mais ampla.

Desta forma, o conceito de enunciado é fundamental para o entendimento do conceito de arquivo, que Foucault define, não como uma estrutura, mas como uma função : «[...] *qui croise un domaine de structures et d'unités possibles et qui les fait apparaître, avec des contenus concrets, dans le temps et l'espace*» (Foucault, 1969:115). O enunciado é o lugar onde os signos e as regras existem e onde são actualizados.

Em Foucault, o sujeito do enunciado é uma função vazia que, enquanto entidade singular, é desprovido de valor, quando equiparado aos valores de espaço e de tempo, elementos fulcrais na sua teoria para a construção do conceito de arquivo: «[...] *le sujet de l'énoncé est une fonction déterminée, mais qui n'est pas forcément le même d'un auteur à l'autre; dans la mesure où c'est une fonction vide, pouvant être remplie par des individus, jusqu'à un certain point, indifférents, lorsqu'ils viennent à formuler l'énoncé; [...]*» (Foucault, 1969:123).

As formações discursivas que constituem um arquivo são geradoras de sistemas discursivos: «*L' archive c'est la loi de ce qui peut être dit, le système qui régit l'apparition des énoncés comme événements singuliers*» (Foucault, 1969:170). Por sua vez, os últimos dependem de práticas discursivas, isto é: «[...] *un ensemble de règles anonymes, historiques, déterminées dans le temps et l'espace qui ont défini à une époque donnée, et pour une aire sociale, économique, géographique ou linguistique donnée, les conditions d'exercice de la fonction énonciative*» (Foucault, 1969:153). O arquivo é, deste modo, um

lugar aglomerador, que permite o acesso ao sistema geral da formação e da transformação dos enunciados pertencentes a formações discursivas.

A função do arquivo situa-se entre a língua e o *corpus*: «*Entre la langue qui définit le système de construction des phrases possibles, et le corpus qui recueille passivement les paroles prononcées, l'archive définit un niveau particulier: celui d'une pratique qui fait surgir une multiplicité d'énoncés comme autant d'événements réguliers, comme autant de choses offertes au traitement et à la manipulation*» (Foucault, 1969:171). Assim, não existe, em momento algum, identificação entre arquivo e *corpus*, caracterizado o primeiro como um lugar de actualizações e o segundo como um repertório passivo de palavras ditas.

O que motiva Foucault, é o objectivo de compreender e de descrever as leis que regem os discursos de especialidade, para comprovar que a unidade destes advém de irregularidades que estão elas próprias sujeitas a regulação: «*Je voudrais montrer que ces unités forment autant de domaines autonomes, bien qu'ils ne soient pas indépendants, réglés, bien qu'ils soient en perpétuelle transformation, anonymes et sans sujet, bien qu'ils traversent tant d'oeuvres individuelles*» (Foucault, 1969). Como afirma o próprio filósofo, a análise do discurso que preconiza não pretende dar conta da universalidade de um sentido, mas sim de uma ordem do discurso independente do seu objecto material, distanciando-se desta forma do formalismo linguístico (cf. Foucault, 1971:72-73).

Por seu turno, as reflexões explicativas de Maingueneau sobre *corpus* e arquivo afiguram-se-nos menos consolidadas, na medida em que do ponto de

vista teórico, os dois conceitos parecem aproximar-se no discurso deste autor. Não obstante, no que se refere a uma aproximação empírica aos referidos conceitos, a sua perspectiva assume, porventura, uma vertente mais operacional. Na qualidade de linguista, Maingueneau entende o arquivo como sendo constituído por enunciados que se organizam em *corpora*: «*On fera ici distinction entre l'archive, sa surface discursive et les corpus que cette dernière permet de déterminer. La surface discursive correspond à l'ensemble des inscriptions relevant d'une même archive. [...] En fonction de ces objectifs le chercheur peut extraire de multiples corpus de cette surface discursive (un corpus de mots, de phrases de tel ou tel type, etc.) qu'il soumet à manipulations, à un traitement*» (1991:25). O arquivo perflha o valor de instituição, enquanto que o *corpus* está sujeito a manipulação, decorrendo este último do primeiro.

A análise do discurso, domínio da linguística que Maingueneau pretende estabilizar, tem por objecto de estudo textos legitimados; a inclusão de um enunciado no arquivo é o reconhecimento da sua legitimidade: «*L'AD [analyse du discours] s'intéresse en effet surtout aux discours autorisés qui, au-delà de leur fonction immédiate, supposent un rapport aux fondements et aux valeurs*» (Maingueneau, 1991: 23). Neste aspecto, as posições de Foucault e de Maingueneau convergem: tanto um como outro consideram que o arquivo é constituído por enunciados legitimados no seio de instituições: «*[...] [les énoncés sont des] choses qui se transmettent et se conservent, qui ont une valeur, et qu'on cherche à s'approprier; qu'on répète, qu'on reproduit, et qu'on*

*transforme; auxquelles on ménage des circuits préétablis et auxquelles on donne statut dans l'institution»* (Foucault, 1969:157).

Enquanto, para Foucault, o arquivo não é analisável no seu todo, uma vez que, na sua essência, é demasiadamente complexo «*En sa totalité, l'archive n'est pas descriptible; et elle est incontournable en son actualité*» (Foucault, 1969:171), para Maingueneau, o arquivo possui uma superfície discursiva a partir da qual podem ser extraídos *corpora*, correspondendo estes ao conjunto dos enunciados, sendo que para os identificar é necessário recorrer a saberes discursivos e extra-discursivos.

Segundo estes autores, o arquivo é, assim, constituído por um conjunto de enunciados legitimados, que estabelecem relações de diversa ordem, quer entre si quer com o mundo que representam. Uma análise em concreto exige, na perspectiva de Maingueneau, o trabalho a partir de *corpora* extraídos dos arquivos, e segundo a terminologia de Foucault, a extracção de fragmentos desses mesmos arquivos.

Desta forma, os corpora, partes constituintes de um todo - o arquivo - caracterizam-se pela presença ou ausência de traços linguísticos e/ou extra-linguísticos, identificados aquando do estabelecimento dos critérios subjacentes à selecção dos enunciados que os constituem.

A carga filosófica de que está impregnado o conceito de arquivo em Foucault perde-se com a escola anglo-saxónica, que desenvolve uma aproximação mais empírica a esse conceito. Em 1996, Sinclair rejeita taxativamente o conceito de *arquivo*, não o considerando um objecto de estudo pertinente para fins linguísticos: «*Words such as collections and archive refer to*

*sets of texts that do not need to be selected, or do not need to be ordered, or the selection and / or ordering do not need to be on linguistic criteria. They are therefor quite unlike corpora»* (Sinclair, 1996:7). Para este autor, arquivo e corpora são divergentes nos objectivos que justificam as suas existências, considerando que só o segundo pode ser seleccionado e organizado segundo critérios linguísticos, os únicos parâmetros a legitimarem os corpora.

Estabelecemos um paralelismo entre as terminologias de Foucault e de Maingueneau e a de Sinclair. Este último designa por «*monitor corpora*» o que Foucault e Maingueneau designam por arquivo e por *subcorpora* o que Maingueneau designa por *corpora*. Para Sinclair, «*A corpus can be divided in subcorpora. A subcorpus has all the properties of a corpus but happens to be part of a larger corpus*» (Sinclair, 1996:9).

Sinclair privilegia o conceito de «*monitor corpora*», em detrimento do de arquivo: «[...] *a dynamic rather than a static phenomenon, consisting of very large amounts of electronically-held text ... A certain proportion of the data will be stored at any one time, but the bulk will necessarily be discarded after processing. The object will be to 'monitor' such data, from various points of view, in order to record facts about the changing nature of the language*» (Sinclair 1991:21). Este tipo de corpora desempenha, assim, a função de observatório da língua. Para este autor, é primordial a constituição de um «*monitor corpora*» para todas as línguas que tenham um estatuto internacional (cf. Sinclair, 1991:25), devendo a extensão deste, idealmente, igualar a da própria língua: «*It is now possible to create a new kind of corpus, one which has*

*no final extent because, like the language itself, it keeps on developping»* (Sinclair, 1991:25).

Apesar de Sinclair preferir o conceito de «*monitor corpora*» ao de arquivo (cf. Leech, 1991:12), Leech não abandona este último conceito sem primeiramente o distinguir do de *corpus*, apontando algumas iniciativas que surgiram no seio da escola americana, com o propósito de distanciar os referidos conceitos, delimitando as respectivas funções. Assim, imputa à *Association of Computational Linguistics Data Coding Initiative*, nos Estados Unidos, a intenção de querer : «[...] *make the concept of an archive almost comparable in scope to that of a national copyright library*» (Leech, 1991:11), sendo que «[...] *the difference between an archive and a corpus must be that the latter is designed or required for a particular 'representative' function*» (Leech, 1991:11).

Como podemos constatar, os conceitos de arquivo e de *corpus* assumem valores diferenciados nas escolas francesa e anglo-saxónica, revelando tais valores posicionamentos filosóficos e teóricos distintos.

A investigação que subjaz a esta dissertação situa-se no âmbito das teorias linguísticas. Desta feita, o objecto sobre o qual efectuamos as análises é um *corpus* informatizado que, numa primeira instância, é desprovido de qualquer acréscimo informativo.

## **2. Tipologia de *corpora***

Não devemos considerar todo e qualquer tipo de *corpora* um objecto válido, *a priori*, para todos os fins da análise linguística. Os critérios de selecção dos

enunciados que compõem o *corpus*, assim como as suas propriedades, têm de estar, necessariamente, em consonância com os objectivos pré-estabelecidos pelo linguista.

Os *corpora*, tal como os enunciados que os constituem, podem ser classificados em tipos distintos. Aludiremos, exclusivamente, à tipologia de *corpora* de textos escritos.

Em 1996, nas «*Preliminary recommendations on Corpus Typology*», Sinclair propõe uma tipologia de *corpus*, que entendemos estar arquitectada sobre pressupostos metodológicos e teóricos que determinam o seu desenho e a sua constituição. Um *corpus* deve ser edificado tendo por base parâmetros que permitam legitimar os resultados obtidos a partir da sua análise. Subjacentes à compilação de um *corpus*, estão respostas a perguntas que Kennedy equaciona do seguinte modo: «*Issues have included whether a corpus should be a static or dynamic sample of a language, how best it can be representative of a language or a genre, how big a corpus should be to be representative or to serve particular purposes, and how big the text samples should be*» (Kennedy, 1998:60).

O facto de os *corpora* poderem ser constituídos por extractos de textos ou por textos integrais, leva Sinclair (1996) a evitar a utilização do termo texto em favor da expressão «*pieces of language*», opção terminológica que indicia, no que concerne ao conceito de texto, um posicionamento teórico não explícito, «*Note that the non-committal word 'pieces of language' is used above, and not 'texts'. This is because of the question of sampling techniques used. If samples are to be all same size, then they cannot all be texts. Most of them will be*

*fragments of texts, arbitrarily detached from their contents*» (Sinclair, 1996:6). Com esta observação, Sinclair faz uma tímida incursão nas questões teóricas relativas ao texto, admitindo que os *corpora* podem ser constituídos por outras realidades que não textos, nunca chegando a explicitar o seu entendimento do mesmo. É ainda com alguma surpresa que verificamos que Sinclair e Ball nem mesmo no documento «*Preliminary recommendations on Text Typology*» (1996) definem o conceito de texto, sendo a supracitada asserção desprovida de valor, uma vez que não revela um posicionamento claro face às teorias do texto.

Convictos da indispensabilidade de assumirmos um posicionamento teórico inequívoco no que se refere ao conceito de texto (cf. ponto 2.2.1), e por entendermos não corresponder a expressão «*pieces of language*» de Sinclair (1996) a um conceito da linguística, recorreremos, neste momento, para mencionar os dados linguísticos que integram os *corpora*, ao conceito de enunciado na acepção de Greimas, isto é, «*[...] toute grandeur pourvue de sens, relevant de la chaîne parlée ou du texte écrit, antérieurement à toute analyse linguistique ou logique*» (Greimas, 1979:123).

Deliberadamente, excluimos da tipologia que apresentamos os *corpora* lexicográficos, isto é, os *corpora* de frequências, os *corpora* de palavras, etc..

## **2.1. Corpora de textos integrais e Corpora de extractos de textos**

Um dos princípios a partir do qual Sinclair constrói a sua tipologia diz respeito aos atributos dos enunciados que constituem os *corpora*. Esses

enunciados podem corresponder a textos integrais ou a amostras de textos, sendo o primeiro tipo de *corpus* designado por «*text corpus*» ou ainda por «*whole text corpus*» e o segundo por «*samples corpus*».

Kennedy, por sua vez, diz-nos que também os «*samples corpus*» podem ser constituídos por textos integrais: «*A sample text corpus may consist of complete texts sampled from a population of complete texts (sometimes called a full-text corpus) or it may consist of samples of a specified size taken from complete texts*» (Kennedy, 1998:21). Pensamos, na verdade, não poder confundir a noção de amostra com a de extracto. Efectivamente, tanto os textos integrais como os extractos de textos podem servir de amostras, sendo que *corpus* de amostragem não se pode opor a *corpus* de textos integrais.

## **2.2. Corpora comparativos e corpora paralelos**

Os *corpora* comparativos e os *corpora* paralelos caracterizam-se por recorrer, no mínimo, a dois sistemas linguísticos distintos.

Os *corpora* comparativos são conjuntos de enunciados bilingues ou multilingues que permitem comparar dois ou mais sistemas linguísticos: «[...] *some multilingual corpora might more truly be described as small collections of individual monolingual corpora in the sense that they use the same or similar sampling procedures and categories for each language but contain completely different texts in those several languages*» (Mc Enery and Wilson, 1996:57).

Por sua vez, os *corpora* paralelos são constituídos por dois conjuntos de enunciados idênticos representados, em geral, por dois sistemas linguísticos

distintos: «*Parallel corpora are corpora which, rather than simply employing the same sampling procedures, actually hold the same texts in more than one language*» (Mc Enery and Wilson, 1996:58). Frequentemente, os *corpora* paralelos são constituídos por textos traduzidos, não se sabendo na maioria dos casos, qual a língua de partida e qual a língua de chegada.

### 2.3. «*Monitor corpora*»

Originariamente, o «*monitor corpora*» era constituído por uma quantidade pré-definida de enunciados. Quando surgia a necessidade de renovar o *corpus*, os enunciados mais desactualizados e antigos eram subtraídos e passavam a fazer parte de arquivos: «*It was originally envisaged that the monitor corpus would remain the same size with 'old' material being relegated to the archives as new material was added [...]*» (Pearson, 1998:45).

Mas de estático, o *corpus* passou a ser considerada uma entidade dinâmica. Na definição deste conceito apresentado no ponto 1.2, o «*monitor corpora*» é concebido como um objecto dinâmico («*dynamic corpora*») que permite a observação e o estudo de aspectos da língua através das suas actualizações ao longo de vários períodos. Um «*monitor corpora*» é dinâmico porque permite a introdução de uma dimensão diacrónica aos *corpora* contemporâneos, que Kennedy explicita do seguinte modo: «*A monitor corpus, more analogous to a moving picture than a snapshot, is so-called because it provides the means to monitor changing patterns of usage over time*» (Kennedy, 1998:61).

Em 1996, Sinclair introduz a noção de «*rate of flow*», que vem reforçar o carácter dinâmico do «*monitor corpus*», já explanado em 1991 (cf. ponto 1.2). A introdução deste conceito tem repercussões ao nível do tamanho do *corpus*, uma vez que o volume de dados aumenta com regularidade oportunista: «*Quantity of text replaces planning of sampling as the main compilation criterion*» (Kennedy, 1998:61).

A metodologia utilizada na recolha e na constituição de dados não obedece ao critério de representatividade, nem de equilíbrio que se refere a um mesmo género ou a um mesmo tipo de texto: «*Instead of setting, say 10 million words as the proper proportion of that genre, the setting could just as easily be 10 millions words a year. Or a month, or a week. The language would flow through the machine, so that at any time there would be a good sample available, comparable to its previous and future state*» (Sinclair, 1996:23).

Com a utilização desta técnica pretende aumentar-se o volume de dados, por forma a permitir a observação de vários estados da língua.

#### **2.4. Corpora de referência**

Num *corpus* de referência, a representatividade e a extensão dos enunciados que o constituem são parâmetros essenciais para que lhe possa ser conferido o estatuto de objecto referencial, cuja finalidade é a de disponibilizar informação compreensiva sobre a língua.

A representatividade do *corpus* traduz-se na diversidade e na homogeneidade dos materiais linguísticos seleccionados, que devem exprimir o

maior número possível de situações em que a língua é usada, estando deste modo representado todos os sujeitos falantes que utilizam um sistema linguístico, num determinado espaço geográfico. Desta diversidade são, porém, excluídos todos os enunciados de especialidade. Assim, *corpus* de referência é definido do seguinte modo: «[...] I recommend for a general corpus that any specialized material is either kept out or stored separately as an ancillary corpus. A general reference corpus is not a collection of material from different specialist areas - technical, dialectical, juvenile, etc.. It is a collection of material which is broadly homogeneous, but which is gathered from a variety of sources so that the individuality of a source is obscured, unless the researcher isolates a particular text» (Sinclair, 1991:17).

Esta definição de *corpus* de referência não nos satisfaz na sua globalidade. A que tipo de material se refere Sinclair, quando sugere que o *corpus* deve ser constituído por «a collection of material which is broadly homogeneous»? Sabemos que os *corpora* de referência são compostos, na sua grande maioria, por textos jornalísticos, estando assim salvaguardado um critério de representatividade, se tivermos em conta o grande público a quem se destinam os jornais.

No entanto, também sabemos que grande parte dos textos jornalísticos incidem em áreas específicas, tais como a Economia, a Crítica, a Arte, a Política, etc.. Quando estes assuntos são explanados em jornais, o grau de especialidade dos conceitos, dos termos e dos discursos é inferior ao grau de especialidade de textos de especialidade, dado o público a quem se dirigem estes últimos ser composto por especialistas.

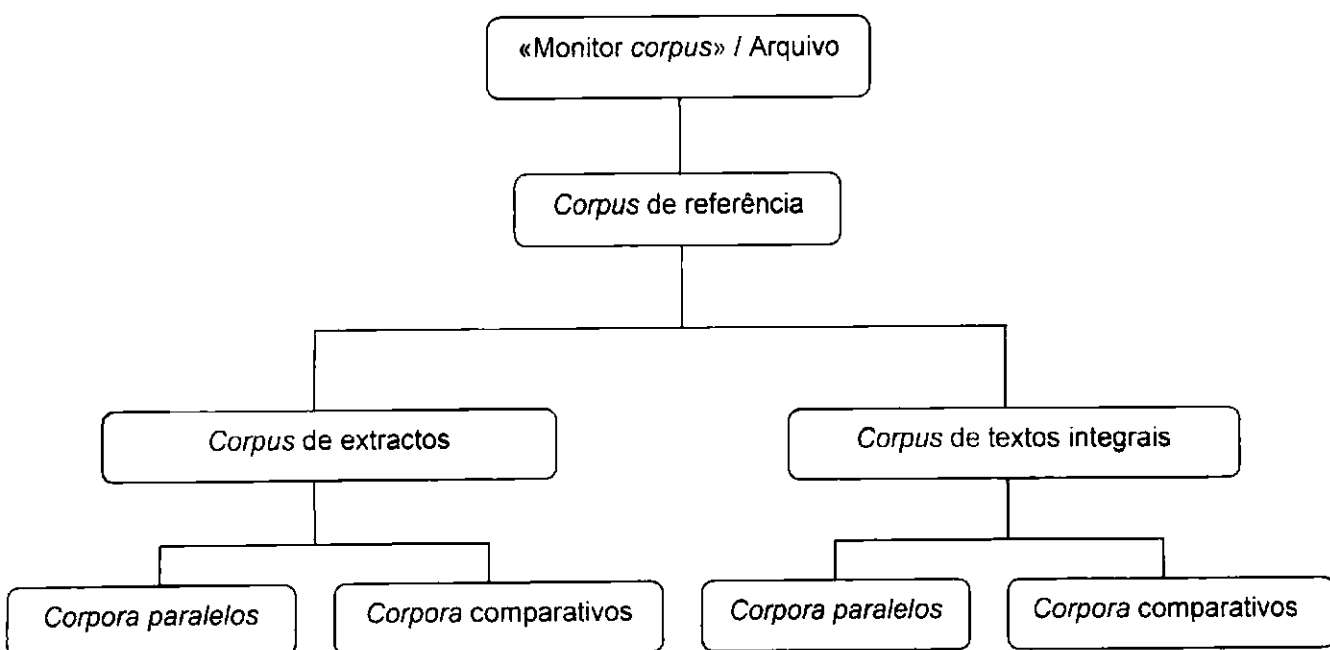
Devido à penetração cada vez maior das áreas de especialidade na vida quotidiana, podemos considerar o texto jornalístico um texto híbrido devido ao recurso intercalar de vocabulário usual com termos. Em 1971, Phal diz-nos a este propósito: «[...] *la langue courante fait des emprunts de plus en plus nombreux aux vocabulaires scientifiques et techniques [...], ce qui traduit le fait que notre vision du monde est de plus en plus influencée par la science et technique*» (Phal,1971:7). A tecnologização da sociedade torna a fronteira entre a língua corrente e a língua de especialidade cada vez menos nítida.

Sinclair não se debruça suficientemente sobre o conceito de área de especialidade e por isso não tem em conta os vários níveis de língua usados em contextos especializados distintos. Porém, tal investida seria fundamental para se decidir que tipo de materiais deve constar de um *corpus*, para que ele seja considerado de referência.

Aceitar a validade da definição de Sinclair, levar-nos-ia a considerar que o *Corpus de Referência do Português Contemporâneo* (CRPC) se baseia em pressupostos errados de selecção de materiais: «*The CRPC [...] is an electronically based linguistic corpus containing at the present 92 million words by sampling from several types of written speech (literary, newspaper, technical, scientific, didactic, economics, decisions of the supreme court of justice parliament) and oral speech formal and informal*» (Bacelar, 2000:1603). Tal como podemos constatar, este *corpus* é constituído por textos de especialidade que, de acordo com Sinclair, não devem fazer parte de um *corpus* de referência.

Se relativamente à definição do conceito de *corpus* de referência (cf. Sinclair, 1991, 1996) e à sua constituição (cf. Bacelar, 2000) não ficámos satisfatoriamente esclarecidos, ficámo-lo quanto à função atribuída a um *corpus* de referência: «*It aims to be large enough to represent all the relevant varieties of the language, and the characteristic vocabulary, so that it can be used as a basis for reliable grammars, dictionaries, thesauruses and other reference material*» (Sinclair, 1996:21). É a partir do *corpus* de referência que se extrai a informação pertinente para a constituição de materiais de referência, tais como os dicionários e as gramáticas, entre outros.

Finalmente, queremos distanciar-nos de Pearson, quando esta afirma: «[...] *a general reference corpus appears to be the superordinate in the hierarchy, [...]*» (Pearson, 1998:44). Se partirmos do princípio que o conceito de «*monitor corpora*» pode ser equiparado ao de arquivo, proporcionando uma abordagem diacrónica aos *corpora* contemporâneos, então o «*monitor corpora*» funciona como hiperónimo de todos os outros tipos de *corpora*, uma vez que é mais abrangente e lato do que qualquer outro:



As relações hierárquicas existentes entre os vários tipos de *corpora* de língua corrente, assim como as suas dependências, estão patentes nesta representação, estando dela excluídos os *corpora* de especialidade.

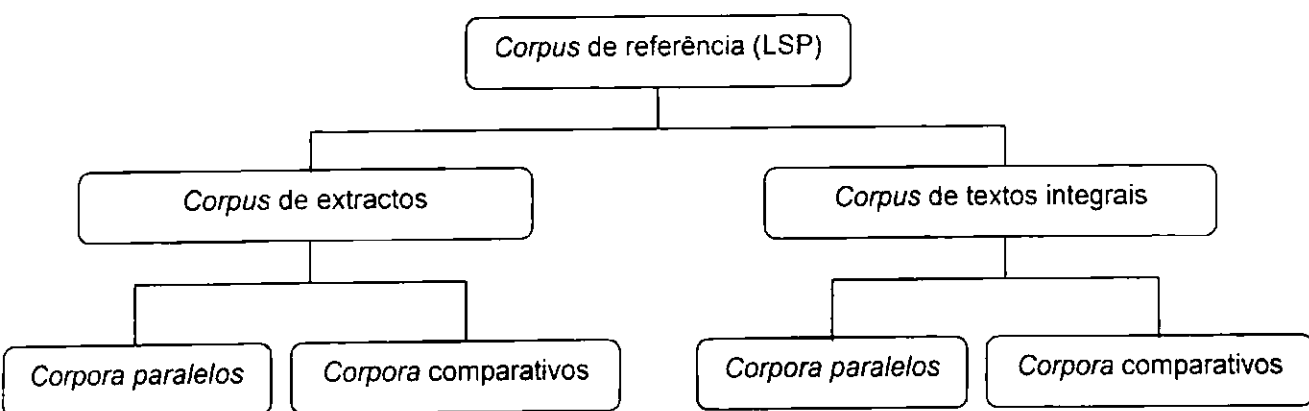
## 2.5. *Corpora de especialidade e corpora especiais*

Sinclair utiliza os termos «*sublanguage*» e «*special corpora*» para designar todos os enunciados que são produzidos em situações sociais específicas, devendo além dos critérios internos, recorrer-se também a critérios externos para definir tais conceitos que se opõem, pela sua natureza, ao de *corpus* ao de referência: «*A sublanguage is at the other end of the linguistic spectrum from a reference corpus*» (Sinclair, 1996:20). Assim, deduzimos que todo o enunciado que, devido às suas características, não pode integrar o *corpus* de referência, faz parte integrante de *corpora* especiais. Baseando-se em Sinclair, Pearson apresenta-nos alguns exemplos de corpora especiais: «*Examples of special corpora given by Sinclair are corpora of the language of children, the language of geriatrics, the language of non native-speakers and the language of very specialized areas of communications*» (Pearson, 1998:46). Geralmente, quando tais conjuntos de dados são abordados do ponto de vista linguístico, cria-se a expectativa de encontrar desvios à norma. É a constatação desses desvios que impede de os incluir no *corpus* de referência, em razão de o resultado da extração de um enunciado deste tipo não constituir um *subcorpus*, porque: «*A subcorpus has all the properties of a corpus but happens to be part of a larger corpus*» (Sinclair, 1996:9).

Não podemos, no entanto, aceitar que enunciados resultantes de situações de comunicação entre especialistas sejam incluídos num *corpus* especial, ou sejam entendidos como uma sublíngua, já que de forma alguma podemos considerar que nestes enunciados existem desvios à norma. O que encontramos neste tipo de enunciados são termos utilizados com acepções diferentes e construções morfosintáticas que, idealmente, reflectem comunicações monorreferenciais e não ambíguas, inerentes às situações de comunicação especializada: «*Les corpus spécialisés réunissent des données linguistiques relatives à une dimension particulière: un domaine, un thème, une situation de communication*» (Habert et alii, 1998:37)

Consideramos que os discursos proferidos em situações de especialidade constituem *corpora* de especialidade, que requerem aproximações metodológicas e teóricas particulares.

Assim, o conjunto dos enunciados produzidos por especialistas em Detecção Remota constituem um *corpus* de especialidade. Se o conjunto de enunciados de especialidade for representativo dos enunciados produzidos pela classe profissional em causa e se a quantidade de enunciados recolhidos for significativa, então assumimos estar perante um *corpus* de referência de especialidade:



A representatividade dos enunciados que compõem um *corpus* é uma questão central da constituição dos *corpora* de especialidade, uma vez que a diversidade dos textos no seio de uma área de especialidade é imensa: «[...] *l'étroitesse du sujet n'empêche pas sa diversité interne, et le problème ne consiste là encore à concevoir un corpus équilibré, que l'on puisse considérer comme échantillon de l'ensemble que l'on veut analyser*» (Habert et alii, 1998:37).

A noção de representatividade em *corpora* especializados não pressupõe a noção de quantidade, dado que a produção de textos numa área de especialidade, numa língua determinada, pode ser diminuta, assumindo o tamanho do *corpus* um valor relativo.

### **3. Corpora anotados**

Com o intuito de ampliar as possibilidades das análises textuais, surgem os *corpora* anotados que pressupõem uma selecção rigorosa dos textos, aos quais se adiciona informação em conformidade com os objectivos de um grupo de investigadores ou com as necessidades gerais de um sector da comunidade.

A informação acrescentada ao *corpus* no seu estado bruto («*raw corpus*») pode ser da ordem da:

- i) linguística: informação morfológica, sintáctica, pragmática, semântica, etc.,

- ii) representação física e lógica: título, subtítulos, maiúsculas, sublinhados, parágrafos, etc..

Em ambos os casos estamos perante a introdução de informação metalinguística, representada através de um sistema de códigos que obedece a pressupostos teóricos e metodológicos, assim como a finalidades bem distintas.

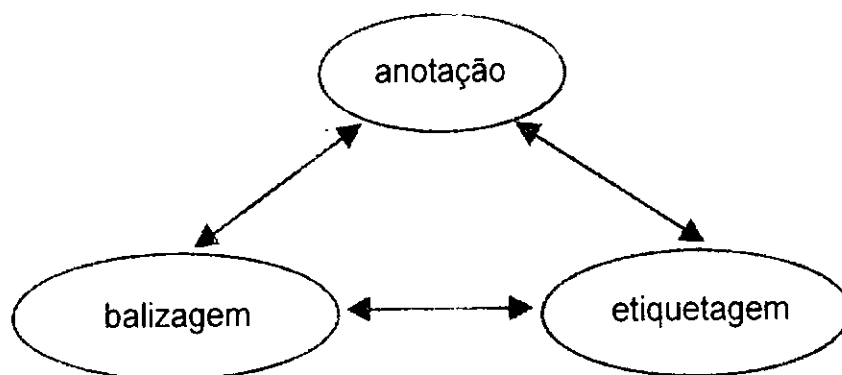
Para Leech, a anotação é entendida como: «[...] *the practice of adding interpretative, linguistic information to an electronic corpus of spoken and / or written language data*» (Leech, 1997:2), que não se confunde com os conceitos de «*mark-up*» ou de «*encoding*».

Assim, anotar um *corpus* significa associar informação linguística a segmentos de texto, recorrendo para o efeito a um conjunto de símbolos, as etiquetas, por forma a poder identificá-los, com vista ao seu tratamento automático. Esta operação é designada de etiquetagem, constituindo o produto final um *corpus* etiquetado.

Anotar um *corpus* pode também ter por finalidade a introdução de um sistema de códigos, cujo objectivo é o de delimitar, por exemplo, um itálico, um título, um parágrafo, um sublinhado, etc.: «*This record can be represented electronically in the computer by special codes and mark-up, and one-to-one mapping between these and symbols can be maintained: the original orthographic document is simply replaced by an unambiguous representation in the forma of an electronic document*» (Leech, 1997:3). Os códigos que permitem, de forma não ambígua, dar conta da representação de um texto

designam-se por balizas, sendo esta operação designada de balizagem e o produto final de *corpus* balizado.

Assim, anotar significa balizar ou etiquetar segmentos de textos no seio de um *corpus*. Deste modo, consideramos a anotação o hiperónimo de balizagem e de etiquetagem, sendo ambos co-hipónimos:



Em língua portuguesa, o termo anotação é utilizado ora como sinónimo de etiquetagem, ora como sinónimo de balizagem.

No contexto desta dissertação utilizaremos o termo anotação como um genérico, para designar qualquer tipo de adição de informação ao texto na sua forma bruta.

### 3.1. SGML

SGML - *Standard Generalized Markup Language*<sup>1</sup> é uma norma ISO (*International Organization for Standardization*) que foi criada a fim de se atingir uma maior flexibilidade na utilização, na reutilização e na permuta de

---

<sup>1</sup> International Organization for Standardization, ISO 8879: Information processing - Text and office systems - Standard Generalized Markup Language (SGML), ([Geneva]: ISO, 1986).

documentos. O objectivo desta norma é a de facultar: «[...] *an unambiguous format for text interchange and a markup language that would be sufficiently rich to permit any (future) processing*» (Herwijnen, 1994:22), permitindo uma codificação universal de algumas das componentes de um texto.

SGML é uma metalinguagem que inclui a) uma sintaxe abstracta, cuja finalidade é a de permitir a balizagem descritiva dos elementos presentes num documento; b) uma sintaxe concreta de referência, que serve de base à sintaxe abstracta e c) declarações que exprimem uma sintaxe de balizagem aplicada.

A sintaxe abstracta dá conta dos esquemas de codificação de aplicações que podem ser empregues pelos utilizadores e que estabelece o esquema conceptual da metodologia SGML. A sintaxe concreta é, por sua vez, uma realização da sintaxe abstracta, que é utilizada para codificar um documento, fornecendo um lote comum de códigos de caracteres e de delimitadores que devem ser utilizados nas aplicações.

Os códigos de caracteres estão definidos na norma ISO 646, enquanto que os delimitadores fazem parte da sintaxe concreta de referência usada pela norma SGML:

<b>Delimitadores</b>	<b>Utilização</b>
<	Abertura de baliza de início
</	Abertura de baliza de fim
<!	Abertura de declaração de balizagem
<?	Abertura de instrução de tratamento
>	Fecho da baliza
--	Início e fim de comentário
#	Indicador de nome reservado

Os delimitadores são utilizados para construir etiquetas <MEMO> que marcam os elementos, isto é, as partes lógicas do documento (capítulo, secção, título, parágrafo, etc.).

Um documento que tenha sido balizado com SGML está preparado para permutar informações ao nível da forma e da estrutura, separando, através da utilização de balizas, a representação física da representação lógica, propondo convenções gerais para os vários tipos de textos. (cf. Habert, 1997:153),

A norma SGML constitui, assim, um meio para descrever documentos, independentemente dos tipos de computadores, dos sistemas operativos e dos programas de paginação.

Um documento SGML está dividido em três partes lógicas: a declaração SGML, a definição de tipo de documento (DTD) e a instância do documento.

A declaração SGML define a sintaxe concreta de referência e especifica toda a informação dependente da máquina em que se trabalha. Esta declaração antecede a DTD e corresponde à primeira parte do documento SGML : «*The SGML declaration defines which characters are used in a document instance, which syntax the DTD is written in, which SGML features are used, and so on*» (Herwijnen, 1994:269). Caso o utilizador não mencione explicitamente uma declaração, o sistema assume uma declaração por defeito.

Exemplifiquemos:

```
<! SGML "ISO 8879 : 1986"
```

```
CHARSET
```

```
BASESET      "ISO 646 : 1983 // CHARSET  
International Reference Version (IRV) // ESC 2/5 4/0  
DESCSET      0 9 UNUSED -- the first 9 characters starting are not used --  
              9 2 9 -- the next two, 9 and 10, map onto 9 and 10 of  
              the baseset --
```

	11	2	UNUSED	-- 11 and 12 are not used--
	13	1	13	-- 13 is mapped onto 13 on the baseset --
	14	18	UNUSED	-- the next 18, starting with 14, are unused --
	32	95	32	-- the next 95, starting with 32, are mapped onto 32 - 126 of the baseset --
	127	1	UNUSED	-- the character 127 is unused --
CAPACITY	PUBLIC "ISO 8879 : 1986/ / CAPACITY REFERENCE/ / EN"			
SCOPE	DOCUMENT			
SYNTAX	PUBLIC "ISO 8879: 1986/ / SYNTAX REFERENCE/ / EN"			

			FEATURES					
MINIMIZE	DATATAG	NO	OMITTAG	YES	RANK	NO	SHARTTAG	YES
LINK	SIMPLE	NO	IMPLICIT	NO	EXPLICIT	NO		
OTHER	CONCUR	NO	SUBDOC	NO	FORMAL	NO		

APPINFO NONE>

In Herwijnem (1994:122)

As componentes da declaração SGML estão explicitadas na norma e, a título de exemplo, apresentamos um quadro explicativo das palavras chave e dos parâmetros que podem constar de uma declaração:

Keyword	Parameter	Explanation
CHARSET	document character st	start of description of document's character set
BASESET	public identifier	Document base character set
DESCSET	0 9 UNUSED	document's character set, in terms of the base set
CAPACITY	public identifier	limits not to be exceeded by document
SCOPE	DOCUMENT	syntax applies to DTD as well as document
SYNTAX	public identifier	syntax used by document
FEATURES	Start of features	describes features used (MINIMIZE, LINK, etc.)
APPINFO	NONE	application specific information

Herwijnem (1994), *Practical SGML*, Boston/Dordrecht/London, Kluwer Academic, pp. 122-123

Todas as regras que definem os tipos de elementos próprios a um documento estão contidas numa DTD. Assim, a DTD define as regras de

balizagem de um documento ou de uma classe de documentos, permitindo a descrição da sua estrutura lógica hierarquizada.

Segundo Habert, balizar um documento significa indicar as subdivisões e as suas relações que compõem a estrutura de um documento. Tal identificação passa por dois momentos distintos, mas consecutivos. Num primeiro momento, define-se uma DTD, que pode ser equiparada à escrita de uma gramática de classes de documentos e num segundo momento, introduz-se: «[...] *des balises choisies dans le document relevant de cette DTD, en respectant les règles éditées pour leur combinaison*» (Habert, 1997:154).

São consideradas componentes da DTD SGML os elementos, as entidades e os atributos.

Os elementos definidos pela DTD «[...] *peuvent apparaître dans un type de document précis et leur associe un nom, qui figure dans les balises correspondantes. Elle précise le contenu des éléments (s'ils peuvent ou non être décomposés dans des éléments plus petits*» (Habert et alii, 1998:64). As entidades, por sua vez, são formas abreviadas de cadeias de caracteres que ocorrem com frequência, reenvios ou servem ainda para incluir outras DTD.

Os atributos, finalmente, permitem associar informação complementar ao documento balizado.

A estrutura da DTD é bivalente, determinando, num primeiro momento, os jogos de caracteres utilizados, fixando as opções permitidas; num segundo momento, definindo de forma efectiva um documento.

Apresentamos uma DTD definida pelo autores do TEI (*Text Encoding Initiative*) e retomada como exemplo por Habert (1998:65):

```

<! -- <!DOCTYPE tei.2 SYSTEM "tei2.dtd" [ - ->

<! -- TEI Lite DTD

:

<! -- Text Encoding Initiative: Guidelines for Electronic      - ->
<! -- Text Encoding and Interchange. Document TEI P3, 1994   - ->
<! -- Copyright (c) 1994 ACH, ACL, ALLC. Permisson to copy   - ->
<! -- in any form is granted, provided this notice is        - ->
<! -- included in all copies.                                 - ->

:

<!ELEMENT listBibl -- ((head)?, (bibl | biblFull)+, (trailer)?)>

:

<!ELEMENT bibl - 0 (#PCDATA | ident| code | kw | abbr | address |
date | name | num | rs | time | add | corr | del | gap | orig | reg |
sic | unclear | emph | foreign | gloss | hi | mentioned | soCalled |
term | tilte | ptr | ref | xptr | xref | anchor | s | seg | gi |
formula | author | biblScope | edition | editor | extent | idno|
imprint | note| publisher | pubPlace | respStmt) *>

:

<!ELEMENT biblFull - 0 (titleStmt, (editionStmt)?, (extent)?,
publicationStmt, (seriesStmt)?, (notesStmt)?, (sourceDesc)*)>

:

<!ELEMENT title - 0 (#PCDATA | ... | text)*>

:

<!ATTLIST title
:
    lang IDREF #IMPLIED
    rend CDATA #IMPLIED
    level (a | m | j | s | u) #IMPLIED
    type CDATA #IMPLIED
    TEIform CDATA "title">

:

<!ELEMENT imprint - 0 (pubPlace | publisher | date | biblScope)*>

:

```

Toda a informação contida na DTD é definida pelo balizador, que conjuga a sintaxe SGML com o conhecimento que tem de um documento, construindo uma gramática do documento estudado.

É a instância que contém os dados balizados incluindo as referências ao DTD, caso este não esteja presente no texto. A utilização do termo instância deve-se ao facto de ele remeter para uma realização concreta de uma classe de documentos que tem a estrutura definida por uma DTD: «*The document instance is the part of SGML document that contains the marked up textual data*» (Herwijnen, 1994:273).

O SGML é uma metalinguagem que permite definir uma gramática normalizada da representação física e lógica dos diferentes tipos de documentos.

### **3.2. TEI**

O esquema de balizagem definido pelo TEI - *Text Encoding Initiative* - é uma aplicação do SGML, que foi desenhado para ser aplicado a qualquer tipo de texto, abstraindo da língua em que foi escrito, da data em que foi produzido ou do género a que pertence.

O TEI é um projecto que resulta do financiamento de três grandes associações - *Association for Computational Linguistics* (ACL), a *Association for Literary and Linguistic Computing* (ALLC) e a *Association for Computers and the Humanities* (ACH), a fim de «[...] *provide standardised implementation for machine-readable text interchange*» (McEnery, Tony e Wilson, Andrew, 1997:27).

A originalidade do TEI consiste em propor conjuntos de directrizes para o uso adequado da norma SGML, aplicado à codificação de textos (cf. Sperberg-Mc Queen and Burnard, 1994).

No TEI, cada texto ou documento é constituído por um cabeçalho («*header*») e por um corpo, isto é, o texto propriamente dito. Do cabeçalho consta informação acerca do texto : autor, título, etc.. Do texto, faz parte informação que permite uma interpretação não ambígua do mesmo.

Para a anotação do cabeçalho e do texto de cada documento recorre-se às balizas («*tags*») e às entidades referenciais («*entity references*»), sendo o texto, assumidamente, constituído por elementos («*elements*») que são anotados com as balizas ou os delimitadores que fazem parte da sintaxe concreta de referência proposta na norma SGML. Apresentamos um exemplo da utilização das balizas SGML que dá conta da extensão de um parágrafo:

```
<p>  
Isto foi a essência.  
</p>
```

Assim, as balizas <...> e </...> indicam, respectivamente, a abertura e o fecho de uma baliza, sendo que a letra minúscula *p* remete para a noção de parágrafo.

Contrariamente às balizas, as entidades referenciais são delimitadas pelos caracteres & e ;. Assim sendo, uma entidade referencial é uma forma abreviada («*shorthand form*») para codificar informação específica dentro do texto. Sempre que um elemento é acompanhado do carácter & ou ; seguido de uma forma abreviada significa que pode ser substituído por uma informação mais ampla e detalhada que se encontra fora do texto, designada por «*feature*

*system declaration*» (FSD), que é considerado um tipo de documento auxiliar no seio do TEI.

O exemplo dado por McEnery, Tony e Wilson, Andrew (1997:28) ilustra bem este procedimento. Quando, no seio de um texto, surge a forma *polished* anotada com *vvd*, *polished&vvd;*, significa que o primeiro v está para verbo, o segundo v para verbo de base lexical e o d para participio passado. A entidade referencial *&vvd;* pode referir-se a uma FSD que contém a informação extensa referente ao significado da forma abreviada:

```
<fs id=vvd type=word-form>
  <f name=verb-clas><sym value=verb>
  <f name=base><sym value=lexical>
  <f name=verb-form><sym value=past>
</fs>
```

A forma abreviada pode a qualquer momento ser substituída pela sua descrição.

No entanto, é a DTD a noção fulcral do TEI: «*This is a formal representation which tells the user or a computer program what elements the text contains, how these elements are combined, and also contains a set of entity declarations, for example representations of non-standard characters*» (McEnery, Tony e Wilson, Andrew (1997:28). O TEI contém normas pré-definidas para balizar tipos de documentos, tais como poemas, artigos de dicionários, fichas terminológicas, etc..

Na criação de uma DTD, cabe ao utilizador decidir o grau de complexidade que quer conferir à anotação de um documento, em consoância

com o objectivo que o move: «*It is important to remember that every document type definition is an interpretation of a text. There is no single DTD which encompasses any kind of absolute truth about a text, [...]*» (Sperberg C.M. e Burnard, Lou, 1994:19).

É o utilizador que interpreta um documento, recorrendo ao TEI para fixar a sua interpretação.

### 3.3. CES

O CES (*Corpus Encoding Standard*) surge no seguimento do TEI em consequência dos projectos europeus EAGLES e MULTEXT. Tal como o TEI, o CES ainda não é uma norma, mas sim uma proposta de norma que tem por finalidade «*[...] fournir des propositions d'encodage minimal pour les corpus électroniques, à la fois pour faciliter leur intégration dans une base de données textuelles et pour leur associer les informations structurales et linguistiques nécessaires*» (Habert, 1998:100).

Tal como vimos no ponto anterior, o TEI visa a balizagem de todo o tipo de documentação: *corpus*, artigos lexicográficos, fichas bibliográficas, etc.. O CES vem especializar-se e dedicar-se exclusivamente à anotação de *corpora*, retendo somente uma parte do TEI, que especifica e prolonga.

O objectivo explícito do CES é a reutilização dos *corpora*, incidindo o seu campo de acção no tratamento automático dos dados, recorrendo para tal a três níveis de anotação. O primeiro nível diz respeito à anotação dos dados estruturais de base do *corpus*, o segundo nível de anotação ocorre nos

objectos linguísticos e o terceiro nos objectos textuais com um estatuto linguístico menos nítido (cf. Habert et alii, 1998:101-102).

Ainda de acordo com Habert: «*Chaque niveau d'annotation doit être validé par l'utilisation d'un parseur SGML muni de la DTD correspondante*» (Habert et alii, 1998:102).

O CES corresponde a um proposta de norma mais completa e complexa que visa a balizagem de *corpora*, permitindo associar-lhe etiquetagens metalinguísticas, visando o levantamento automático de informação.

O CES distingue-se do TEI por aplicar-se exclusivamente a *corpora*, permitindo diversos níveis de anotação, correspondendo cada nível de anotação a uma DTD diferente: «*Là où la TEI insère dans le même document les balises correspondant aux différents niveaux d'annotation envisagés, CES répartit ces balises dans des documents distincts, relevant de DTD différentes*» (Habert et alii, 1998:101). A anotação, é assim, feita por camadas distintas e não através da interacção das diferentes anotações, podendo coexistir várias anotações do mesmo nível: «*On peut ainsi utiliser différents étiqueteurs morphosyntaxiques*» (Habert et alii, 1998:101).

A associação das diversas formas de anotação - balizagem e etiquetagem - permite aumentar em qualidade e em quantidade o tratamento automático da língua natural (TALN), obtendo resultados cada vez mais performativos, necessários para o desenvolvimento da língua.

### 3.4. Etiquetagem gramatical («part-of-speech annotation»)

As etiquetas metalinguísticas são adicionadas aos *corpora* brutos, por forma a tornar explícita a informação sobre a língua que está presente de modo implícito no corpus, permitindo um tratamento mais rápido e eficiente da informação e da análise linguística. A este propósito, McEnery e Wilson apresentam o seguinte exemplo: «*For example, the part-of-speech information 'third person singular present tense verb' is always present implicitly in the forma loves, but it is only retrieved in normal reading by recourse to our preexisting knowledge of the grammar of English*» (McEnery; Wilson, 1997:24). A etiquetagem de um *corpus* explicita a interpretação levada a cabo por um anotador ou por uma equipa de investigação.

Para que um *corpus* seja reutilizado, multifuncional e facilmente manejável pelos utilizadores, é essencial terem-se presentes algumas linhas orientadoras no momento de proceder à sua etiquetagem. Leech (1997) menciona algumas máximas que devem guiar o anotador.

Assim, o *corpus* etiquetado deve permitir uma grande flexibilidade aos utilizadores, podendo estes a qualquer momento subtrair as etiquetas ao *corpus*, facultando um regresso inequívoco ao *corpus* bruto. A bateria de etiquetas, tal como as orientações que presidem à etiquetagem, devem ser disponibilizadas ao utilizador, devendo este ficar esclarecido sobre como e quem procedeu à etiquetagem. Geralmente, este tipo de informação está contido num ficheiro, em apêndice ao *corpus*.

A etiquetagem não sendo um processo infalível, é um instrumento de grande utilidade. Mas para que um *corpus* etiquetado seja reutilizável pelo maior número possível de utilizadores, é fundamental que a etiquetagem usada obedeça a critérios tão consensuais quanto possível e que se baseie em princípios teóricos neutros. No entanto, nenhum esquema de etiquetagem tem *a priori* o condão de ser considerado uma norma, uma vez que a norma surge de uma prática consensual.

O tratamento automático ou semi-automático da língua natural pressupõe a etiquetagem de elementos linguísticos mediante uma bateria de etiquetas concebidas em função do nível linguístico e do grau de minúcia com que se pretende tratar o *corpus*. As etiquetagens podem ser atribuídas aos segmentos de forma automática, utilizando para o efeito um etiquetador («*tagger*») ou podem ainda ser atribuídas de forma manual.

O *corpus* pode ser etiquetado totalmente ou parcialmente, obedecendo aos critérios pré-estabelecidos pelo linguista.

Actualmente, os níveis linguísticos tratados com sucesso são os níveis: ortográfico, fonético, prosódico, gramatical, sintáctico, semântico, discursivo, pragmático e estilístico.

A etiquetagem sintáctica corresponde à prática de adicionar informação sintáctica a um *corpus*, ao incorporar indicadores de informação sintáctica no texto.

As etiquetagens gramatical e sintáctica são, frequentemente, tratadas em conjunto, já que ambas dão conta das características gramaticais de um texto, sendo a etiquetagem gramatical considerada um pré-estágio necessário

à etiquetagem sintáctica. Leech considera estes dois níveis de tratamento simbióticos: «*In fact, there are strong argument that these are nor really distinct levels at all: gramatical tagging is merely a specification of the leaves (or pre-terminal nodes) of the phrase-structure (PS) tree which is favoured model for syntactic annotation*» (Leech, 1997:19).

O facto de uma etiquetagem sintáctica pressupor uma etiquetagem gramatical confere a esta última um valor acrescentado no seio do TALN.

Atribuir etiquetas gramaticais a um *corpus* significa associar a cada unidade lexical a sua categoria morfossintáctica, utilizando para o efeito um conjunto de etiquetas pré-definidas. Embora não exista nenhuma norma que institua e defina o conjunto de etiquetas a utilizar, tem-se por referência as baterias de etiquetas («*tagset*») utilizadas nas etiquetagens dos grandes *corpora*, tais como a *British National Corpus* (BNC), o *Lancaster-Oslo/Bergen Corpus* (LOB *Corpus*) e a *Surface Underlying Structure Analysis of Natural English* (SUSANNE) para a língua inglesa; para a língua francesa, o *Menelas* e o *Mitterand 1*, entre outros.

O resultado obtido após a etiquetagem gramatical de um *corpus* pode ser apresentado num formato horizontal ou vertical. Apresentamos um extracto retirado do LOB *corpus*<sup>2</sup> no seu formato horizontal, em que a cada unidade lexical é acrescentada uma etiqueta morfossintáctica:

A01	2	^ * * stop_VB electing_VBG life_NN peers_NNS * * * _.
A01	3	^ by_IN Trevor_NP Williams_NP _.
A01	4	^ a_AT move_NN to_TO stop_VB \0Mr_NPT Gaitskell_NP from_IN
A01	4	nominating_VBG any_DTI more_AP labour_NN
A01	5	life_NN peers_NNS is_BEZ to_TO be_BE made_VBN at_IN a_AT meeting_NN
A01	5	of_IN labour_NN \0MPs_NPTS tomorrow_NR _.
A01	6	^ \0Mr_NPT Micheal_NP Foot_NP has_HVZ put_VBN down_RP a_AT

<sup>2</sup> <http://www.hit.uib.no/icame/lob-eks.html>

A01 6 resolution\_NN on\_IN the\_ATI subject\_NN and\_CC  
 A01 7 he\_PP3A is\_BEZ to\_TO be\_BE backed\_VBN by\_IN \OMr\_NPT Will\_NP  
 A01 7 Griffiths\_NP\_' \OMP\_NPT for\_IN Manchester\_NP  
 A01 8 Exchange\_NP\_'.

Apresentamos o mesmo extracto na sua forma vertical, em que cada unidade lexical é apresentada em linhas separadas, acompanhada da sua etiqueta, assim como de alguma informação adicional:

```

A012001 ----
A012002 *'          *'          H
A012010 VB          stop          H
A012020 VBG        electing      H
A012030 NN          life          H
A012040 NNS        peers         H
A012041 **
A012042 .          .          H   @
A013001 ----
A013010 IN          by          H
A013020 NP          Trevor        H
A013030 NP          Williams      H
A013031 .          .          H   @
A014001 ----
A014010 AT          a            P
A014020 NN          move
A014030 TO          to
A014040 VB          stop
A014050 NPT        \OMr          \O
A014060 NP          Gaitskell
A014070 IN          froma
A014080 VBG        nominating
A014090 DTI        any
A014100 AP          more
A014110 NN          labour        N
A015010 NN          life          N
A015020 NNS        peers         N
A015030 BEZ        is
A015040 TO          to
A015050 BE          be
A015070 IN          at
A015080 AT          a
A015090 NN          meeting
A015100 IN          of
A015110 NN          labour        N
A015120 NPTS       \OMPs         \O
A015140 NR          tomorrow
A015141 .          .
  
```

Cada uma das etiquetas que acompanham as unidades lexicais são etiquetas que constituem baterias de etiquetas, que permitem identificar em contexto cada uma das formas constituintes do *corpus*.

Assim, a título de exemplo, no LOB *corpus*, a classe de palavras 'nome' apresenta as seguintes classificações:

N	=	<i>noun</i>
NP	=	<i>'true proper name'</i>
NPS	=	<i>plural proper name</i>
NPL(S)	=	<i>locative nouns with word-initial capital</i>
NPT(S)	=	<i>titular names with word-initial capital</i>
NNP(S)	=	<i>common noun habitually written with a word-capital</i>
JNP	=	<i>adjective habitually written with a word-initial capital</i>

Outro jogo de etiquetas é o utilizado pelo *corpus* francês *Mitterand 1*, em que a frase:

[...] *moi, je suis de la France - je ne dis pas: je suis la France* [...] <sup>3</sup>

é etiquetado da seguinte forma:

moi, je,5  
" " " " "p"  
je, je,5  
suis, être, 11  
de, de, 81  
la, le, 7  
France, France, 22  
- , , p  
je, je, 5  
ne, ne, 6  
dis, dire, 11  
pas, pas, 6  
: , , p  
je, je, 5  
suis, être, 11  
la, le, 7  
France, France, 22

O texto etiquetado é constituído por conjuntos de três elementos, em que o primeiro corresponde à unidade lexical tal como surge no *corpus*, o segundo corresponde ao lema e o terceiro corresponde à sua categoria, que está representada por um número.

Os poucos exemplos de baterias de etiquetas que apresentámos, deixam antever uma grande dispersão e disparidade quer na criação e na utilização de etiquetas, quer, mais particularmente, das realidades linguísticas que estas cobrem: «*On peut ainsi par exemple avoir des étiquettes pour les articles, et considérer les possessifs (mon, ton, son, etc.) comme faisant partie des adjectifs, ou bien inclure les uns et les autres dans une catégorie "déterminants"*» (Véronis J. , Khouri L., 1995: 236).

As divergências teóricas subjacentes à etiquetagem das unidades lexicais e dos segmentos linguísticos constituem um dos maiores entraves ao consenso entre os especialistas. A identificação e a atribuição das etiquetas adequadas às entidades linguísticas, assim como o estabelecimento de regras linguísticas localizadas que permitam, num primeira fase, um levantamento automático dos termos e a desambiguação automática, numa segunda, representam, para nós, o grande desafio do tratamento automático da língua natural.

---

<sup>3</sup> Exemplo retirado de HABERT (1997:22)

## Capítulo II

---

### Constituição de *corpus*

1. Comunidade científica, produtora de textos de especialidade
2. Texto de especialidade
  - 2.1 Texto e discurso
  - 2.2 Contextos
  - 2.3 Intervenientes
3. Tipologia de tipos ou tipologia de géneros?
4. Constituição do *corpus* em Detecção Remota
  - 4.1. Definição da área de especialidade em análise: Detecção Remota
  - 4.2. Constituição de tipologias
    - 4.2.1. Tipologia de tipos de discurso
    - 4.2.2. Tipologia de géneros de discurso

## Capítulo II - Constituição de *corpus*

### 1. A comunidade científica, produtora de textos de especialidade

A comunidade científica compõe-se de indivíduos que possuem conhecimentos específicos para exercer, ensinar, discursar e escrever sobre uma especialidade científica.

Kuhn reflectiu longamente sobre a questão da pertença a uma comunidade científica e aponta como razão principal para definir essa pertença a existência de um paradigma comum aos seus membros: «*Un paradigme est ce que les membres d'une communauté scientifique possèdent en comum, et réciproquement, une communauté scientifique se composent d'hommes qui se réfèrent au même paradigme*» (Kuhn, 1970:240).

Assim, ser membro integrante de uma comunidade científica implica, na perspectiva kuhniana, ser portador de um conjunto de conhecimentos baseados em investigações sólidas (cf. Kuhn, 1970:30). Para que um membro de uma comunidade científica seja aceite e reconhecido como especialista, é necessário que active mecanismos discursivos (escritos e/ou orais) que lhe permitam partilhar o seu saber com os membros da comunidade a que pertence. Os grupos receptores da comunicação especializada são constituídos por indivíduos de uma comunidade científica que, teoricamente, detém um elevado grau de conhecimento científico.

No entanto, estamos cientes de que a comunicação científica também pode ser dirigida a indivíduos exteriores a ela. Se quiséssemos introduzir uma

noção de escala, diríamos que, no que respeita ao conhecimento, tais indivíduos encontram-se em oposição simétrica aos membros de uma comunidade científica, correspondendo, desta forma, ao grupo com o menor grau de conhecimento científico.

Os mecanismos discursivos que o especialista activa nestas duas situações de comunicação são distintos. No primeiro caso, estamos perante um discurso científico especializado; no segundo caso, estamos perante um discurso vulgarizado. O grau de complexidade conceptual representado nos textos resultantes dos respectivos discursos vai do mais complexo, o texto especializado, ao menos complexo, o texto vulgarizado.

O especialista produz textos com diferentes graus de dificuldade, ao nível do conhecimento, para veicular e difundir o seu saber, sendo, no entanto, através do texto de especialidade que contribui de forma mais efectiva para a evolução do conhecimento no seio da comunidade científica a que pertence. O texto de especialidade é uma das formas privilegiadas ao qual o especialista recorre para transmitir e aceder ao conhecimento.

A comunidade científica reúne indivíduos que, nas mais diversas situações de comunicação, podem assumir tanto o papel de produtor como o de receptor de discursos e/ou textos. Na situação em que existe um debate de ideias, os papéis invertem-se, alternadamente. Para que a comunicação seja possível, é imprescindível que os intervenientes tenham os mesmos referentes, que tenham conhecimentos e leituras que lhes permitam, a cada momento do debate, actualizar o seu saber e contribuir, a longo prazo, para o progresso das

ciências e para a transformação das representações sociais (Bronckart, 1996:111).

Os membros da comunidade auto-regulam os seus discursos, sendo eles os únicos verdadeiros juizes da sua produção científica: «[...] *les membres d'un groupe scientifique se considèrent et sont considérés par les autres, comme les seuls responsables de la poursuite d'un ensemble d'objectifs qui leur sont communs et qui englobent la formation de leurs sucesseurs*» (Kuhn, 1970:241).

Ao especialista, produtor e receptor de discursos e de textos, é reconhecida legitimidade para debater paradigmas e referências. Estabeleceu-se um contrato, tacitamente aceite, entre os intervenientes que se propõem debater ideias, uma vez que possuem a mesma memória científica: «[...] *la connaissance scientifique est intrinséquement la propriété d'un groupe [...]*» (Kuhn, 1970:284). O saber que um indivíduo detém sobre uma ciência consiste, na essência, no conhecimento e na capacidade de criação de discursos e de textos que poderão servir de referência a outros ou a futuros membros da comunidade científica. Dominar as referências, que são comuns a uma mesma comunidade, significa pertencer-lhe de pleno direito.

O especialista é, conseqüentemente, o produtor e o receptor dos textos de especialidade que são produzidos e consumidos por e para uma comunidade de especialistas, que perfaz uma comunidade de comunicação, também ela restrita. Assim, fazer parte de uma comunidade científica implica, entre outras perícias, dominar um discurso próprio a um conjunto de indivíduos que têm em comum saberes específicos sobre uma área do conhecimento.

O texto escrito, expressão de um saber, poderá transformar-se em conhecimento, sendo o consenso entre os especialistas relativamente à verdade explanada no documento, uma condição *sine qua non*. Maingueneau refere-se à importância do texto escrito, enquanto produto, para conferir e consolidar a identidade de uma comunidade restrita: «[...] *la production écrite implique cruciallement un groupe, une communauté associée à ses rites* [...]» (Maingueneau, 1992:117). A noção de autoridade, a existência de rituais e a importância da memória são fulcrais para que a relação entre os membros que constituem uma comunidade científica seja coesa, sendo os textos, através das relações sociais, conceptuais e linguísticas que neles estão expressas, um elo de ligação entre os seus membros.

É, efectivamente, o texto escrito que permite a um grupo de indivíduos perpetuar a sua memória científica, porque «[...] *l'écrit conserve le discours et en fait une archive disponible pour la mémoire individuelle et collective*» (Ricoeur, 1986:139).

## **2. Texto de especialidade**

### **2.1 Texto e discurso**

O texto escrito, produto estável resultante de uma actividade intelectual e profissional, provindo de uma comunidade de comunicação restrita, é o objecto de estudo desta dissertação.

Mas abordar o texto, entendido como um objecto material e concreto, acarreta necessariamente uma reflexão teórica sobre o texto, entendido como

um objecto em abstracto (cf. Adam, 1999:40). Intimamente ligado ao conceito de texto, surge o conceito de discurso, sendo que, nos planos teórico e metodológico, nem sempre se conseguem distinguir cabalmente um do outro.

Para Maingueneau discurso e texto têm em comum o facto de designarem produções verbais. Na sua obra intitulada *Analyser les textes de communication* (1998), o autor define longamente o conceito de discurso e aborda de forma sumária o conceito de texto. Também Rastier evidencia a primazia do conceito de discurso sobre o de texto, reconhecendo que este último tem vindo a ser substituído por uma panóplia de novos conceitos com características interdisciplinares, centrados nos contextos extralinguísticos inerentes ao acto verbal: «[...] *le concept de texte a été supplanté par ceux de discours, d'interaction verbale, etc.*» (Rastier, 1998:106).

A razão pela qual o conceito de discurso é abundantemente utilizado nas ciências da linguagem deve-se ao facto de isso representar «[...] *le symptôme d'une modification dans notre façon de concevoir le langage*» (Maingueneau, 1998:38). Maingueneau defende que tal modificação se operou devido à confluência de diversas correntes das ciências da linguagem que compõem a pragmática, que se consubstanciam numa forma particular de apreensão da comunicação verbal. Aceitar tal constatação implica perspectivar o discurso em termos das propostas teóricas e metodológicas inerentes à pragmática. Greimas considera a pragmática da linguagem, na perspectiva da escola americana, como «[...] *un des aspects de la dimension cognitive, car elle concerne en fait la compétence cognitive des sujets communicants, telle qu'on peut la reconnaître [...] à l'intérieur des discours-énoncés [...]*» (1979:288).

Desta forma, os intervenientes na comunicação, isto é, o autor e o destinatário, fazem parte integrante do processo da comunicação, uma vez que são indivíduos com competências específicas.

Enquanto Greimas introduz aqui o conceito de enunciado, que equipara a discurso, Maingueneau equipara tal conceito ao de texto. Para este autor, o enunciado e o texto são co-ocorrentes de discurso, admitindo, no contexto da sua obra, que «[...] «*énoncé*» possède donc une valeur à peu près équivalente à celle de «*texte*»» (1998:42), enquanto o enunciado for entendido como «[...] *une séquence verbale qui forme une unité de communication complète relevant d'un genre de discours déterminé*» (1998:42).

Na obra supracitada, o texto merece pouca atenção em termos definitórios. O texto é incluído no âmbito do estudo da linguística textual, devendo ser apreendido como um todo coerente, caracterizado por ser uma produção oral ou escrita, estruturado de forma a perdurar fora do contexto em que foi enunciado.

Em *Les termes clés de l'analyse du discours*, Maingueneau opõe discurso a texto, com base na seguinte afirmação: «*En parlant de discours, on articule l'énoncé sur une situation d'énonciation singulière; en parlant de texte, on met l'accent sur ce qui lui donne son unité, qui en fait une totalité et non une simple suite de phrase*» (Maingueneau, 1996:82). Depreende-se desta afirmação que tanto o discurso como o texto são enunciados; o que os distingue, na realidade, é o ponto de vista sob o qual são abordados.

O discurso deve ser apreendido como um enunciado associado às condições em que foi produzido, enquanto o texto deve ser sentido, numa

primeira instância, como um todo coerente, sendo tal coerência um dos parâmetros que permite definir o conceito de texto.

O texto é um documento que faz parte do todo, o arquivo, na acepção de Foucault, e pode ser analisado segundo duas realidades distintas: as realidades inerentes à sua produção e as realidades inerentes à sua leitura, isto é, o momento em que o texto está a ser actualizado à luz dos conhecimentos e das realidades históricas e sociodiscursivas do leitor. Os discursos são, maioritariamente, apreendidos sob forma de textos, recorrendo o linguista ao conceito de texto ou de discurso em função da análise que pretende efectuar. O texto é uma das formas privilegiadas para veicular um discurso.

Distinguir discurso e texto, com base na essência deste último conceito, isto é, «[...] *productions verbales orales ou écrites qui sont structurées de manière à durer, à être répétées, à circuler loin de leur contexte original*» (Maingueneau, 1998:43), parece-nos insuficiente para decidir, por exemplo, se estamos a trabalhar o texto jurídico ou o discurso jurídico, um texto científico ou um discurso científico.

A possível ambiguidade resultante da opção por um ou por outro conceito, texto ou discurso, deve-se, também, ao facto de o suporte físico sobre o qual se trabalha ser o mesmo. O texto escrito pode ser abordado do ponto de vista discursivo e/ou textual, nem sempre sendo consciente e explícita a passagem de um a outro nível de análise.

Bronckart considera que os textos podem ser apreendidos a diversos níveis: o texto enquanto entidade genérica ou geral e o texto enquanto entidade

singular ou empírica (1996:73-79). O texto, na qualidade de entidade genérica, é definido do seguinte modo: «[...] *chaque texte est en relation d'interdépendance avec les propriétés du contexte dans lequel il est produit; chaque texte exhibe un mode déterminé d'organisation de son contenu référentiel; chaque texte est composé de phrases articulées les unes aux autres selon des règles compositionnelles plus ou moins strictes; chaque texte enfin met en oeuvre des mécanismes de textualisation et de prose en charge énonciative destinés à lui assurer sa cohérence interne*» (Bronckart, 1995:74).

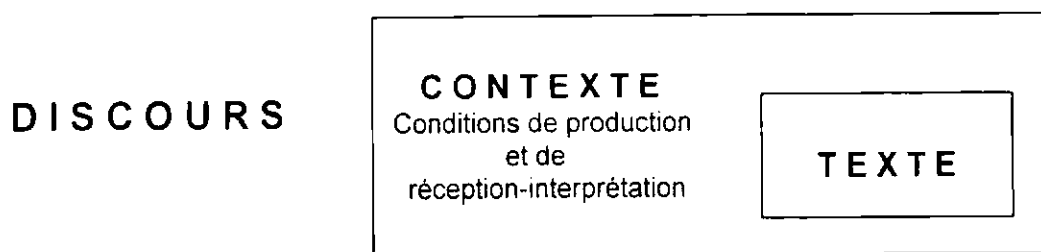
O texto, enquanto entidade singular, é definido como um objecto concreto que, embora apresentando semelhanças com outros textos, possui características específicas que o distinguem de todos os outros: «*Chaque texte singulier exhibe, en d'autres termes, des caractéristiques individuelles et constitue de ce fait un objet toujours unique*» (Bronckart, 1996:79).

Adam, por sua vez, tenta distinguir, do ponto de vista teórico e metodológico, texto de discurso, revendo algumas considerações que teceu, nomeadamente, nos *Éléments de linguistique textuelle* (1990). Nesta obra, ele apresenta uma fórmula para distinguir o discurso de texto, que sintetiza na seguinte expressão:

<b>Discours</b>	=	<b>Texte + Conditions de production</b>
<b>Texte</b>	=	<b>Discours - Conditions de production</b>

Adam (1990), *Éléments de linguistique textuelle*, Bruxelles-Liège, Mardaga, p.23.

De acordo com esta formulação, o texto é um objecto desprovido das condições discursivas em que foi produzido. Na sua obra mais recente, Adam afirma, actualmente, discordar da fórmula que tinha proposto, uma vez não se poder descontextualizar o texto das práticas discursivas: *«Cette formule ne doit pas opérer la décontextualisation que je préconisais alors. Il s'agit d'une formule d'inclusion du texte dans le champ plus vaste de pratiques discursives [...]»* (Adam, 1999:39). Por conseguinte, apresenta uma nova fórmula, em substituição da anterior, que obedece à seguinte expressão:



Adam (1999), *Linguistique textuelle. Des genres de discours aux textes*, Paris, Nathan, p. 39.

Decorrente desta remodelação, Adam propõe definições renovadas de texto e de discurso, em que texto assume duas acepções distintas. Na primeira acepção, o texto é entendido como um objecto abstracto que é *«[...] l'objet d'une théorie générale des agencements d'unités [...] au sein d'un tout de rang de complexité linguistique plus ou moins élevé.[...]»*; na segunda acepção, o texto é entendido como *«[...] objet[s] concret[s], matériel[s], empirique[s]. Chaque texte se présente comme un énoncé complet, le résultat toujours singulier d'un acte d'énonciation. C'est pas excellence, l'unité de l'interaction textuelle [...]»* (Adam, 1999:40).

Por seu turno, falar em discurso, acarreta implicações a dois níveis, que Adam explicita do seguinte modo: *«[...] ouvrir le texte, d'une part, sur une*

*situation d'énonciation-interaction toujours singulière et, d'autre part, sur l'interdiscursivité dans laquelle chaque texte est pris - en particulier celle des genres»* (Adam, 1999:40). Inequivocamente, texto e discurso mantêm uma relação de interdependência, não podendo um ser pensado sem o outro, uma vez que o discurso actualizado apresenta-se, geralmente, sob forma de texto. Deste modo, a separação entre o que pertence ao nível do textual e do discursivo é, para este autor, essencialmente uma questão metodológica: *«Dans la pratique d'analyse qui nous retiendra prioritairement dans cet essai, cette distinction tend à s'estomper»* (Adam, 1999:40-41). Pode inferir-se da afirmação de Adam que a tendência é a de não distinguir os dois conceitos, a não ser que tal distinção seja benéfica para os objectivos da análise em questão.

Conclusivamente, falar em texto implica a noção de discurso, uma vez que o texto fixa um discurso que foi proferido por um indivíduo, num espaço e num tempo específicos. Na perspectiva de Ricoeur (1986:137), o texto é entendido como *«[...] tout discours fixé par l'écriture»*. Quanto a esta matéria, Rastier parece seguir a linha de pensamento de Ricoeur, uma vez que afirma: *«Il nous semble en effet que le concept de texte s'étend à toute pratique linguistique attesté, produite dans une pratique sociale déterminée, et fixée sur un support»* (1998:106). Deixa, assim, implícita uma aproximação teórica entre os conceitos de texto e de discurso.

A relação entre texto e discurso pressupõe um entrecruzar constante dos pressupostos teóricos que suportam os dois conceitos, sendo, porém, o nosso

objecto de estudo, o texto enquanto entidade singular e concreta resultante de uma produção verbal específica o verdadeiro objecto do nosso estudo.

Tendo em conta o exposto e a afirmação «*Le discours est conçu comme l'association d'un texte et de son contexte*» (Maingueneau, 1996:29), assim como a asseveração de Rastier :«[...] *le texte est le lieu de rencontre entre le contexte et l'intertexte*» (1998:107), julgamos pertinente abordar o texto através dos seus contextos sociodiscursivos e linguísticos.

## 2.2 Contextos

O texto pode, simultaneamente, ser entendido como produção e produto de uma comunidade de comunicação restrita (Maingueneau, 1992) ou nuclear (Beacco, 1995). Nele concentram-se todos os elementos linguísticos e extralinguísticos que dão conta da interacção da linguagem com a vida social, que fazem com que o texto possa ser analisado ao mesmo tempo como um processo e um resultado.

A importância dos parâmetros extralinguísticos no momento da produção textual, perceptíveis na análise do texto enquanto produto de uma actividade intelectual e social de um ou de um grupo de indivíduos é, presentemente, um facto irrefutável.

Diversas são as designações propostas para referir de forma não ambígua a realidade que não pertence à esfera da Linguística, mas sim às esferas social, psicológica, pragmática e/ou física, e que interfere na produção textual, reflectindo-se no produto final, o texto.

Designaremos, *a priori*, por contexto as realidades linguísticas ou não, que envolvem o texto e a sua produção. Rastier introduz o conceito de contexto como elemento indispensável à interpretação de um texto. Para ele, o contexto pode ser problematizado de distintas formas: «[...] *il est utilisé soit pour réintroduire une forme limitée de globalité dans l'interprétation, soit pour lier la langue à une exteriorité en conditionnant la construction du sens à la connaissance de cette exteriorité*» (Rastier, 1998:98).

Deste modo, o contexto pode assumir várias significações, podendo o seu uso ser gerador de ambiguidades. Bronckart é um dos autores que evita a utilização do termo contexto por o considerar ambíguo e por não o considerar suficientemente abrangente. Prefere recorrer ao conceito de «*extralangage*», que considera um conceito operatório, definindo-o como: «[...] *l'ensemble théoriquement infini de toutes les entités «mondaines» en dehors du langage*» (1995:26). Mas quando, por necessidade, se reporta ao contexto, fá-lo utilizando-o como sinónimo de «*extralangage*», distinguindo este último do conceito de co-texto que remete, inequivocamente, para questões do foro linguístico: «[...] *le concept de contexte s'adresse à l'extralangage, et qu'il faut éviter de le confondre avec l'environnement linguistique d'un énoncé. Pour désigner ce dernier, nous utiliserons le concept de «co-texte» [...]*» (1995:27).

Também Rastier distingue estes dois tipos de contextos considerando que o contexto linguístico se define como «*un voisinage local*» (Rastier, 1998:106), o contexto extraverbal como uma situação.

Para que o texto seja apreendido como um todo, é necessário que se tenham em conta todos os parâmetros linguísticos e extraverbais que intervêm no acto da escrita.

Na sua obra, publicada em 1996, Bronckart utiliza a expressão «*situation d'action langagière*» para referir os parâmetros que influem na análise do texto, entendido como objecto referencial. Diz-nos que «*Cette expression générale désigne les propriétés des mondes formels (physiques, sociales et subjectifs) qui sont susceptibles d'exercer une influence sur la production textuelle*» (1996:93). Tais parâmetros podem ser subdivididos em «*situation d'action langagière externe*» e «*situation d'action langagière interne*». A primeira diz respeito às características do mundo formal, isto é, do mundo observável, a segunda, está relacionada com a acção interiorizada dos indivíduos, tanto do autor como do destinatário.

Para se referir às mesmas realidades, Décles utiliza uma terminologia distinta. Distingue o contexto externo - «*Les contextes externes  $K_e$  font appel à des connaissances sur le domaine externe, sur les interlocuteurs, sur les conditions de dialogue, sur les buts poursuivis et les tâches à résoudre...*», (1997:216) - do contexto interno - «*Les contextes internes  $K_i$  reposent sur des connaissances linguistiques qui entrent dans le co-texte d'une unité linguistique*» (1997:216). Enquanto Bronckart utiliza o termo de co-texto para designar o ambiente linguístico de um enunciado, Décles utiliza-o para designar o ambiente linguístico que envolve a unidade linguística em análise. Para este autor, o conceito de co-texto abrange uma realidade linguística delimitada e curta, em comparação com o conceito de co-texto de Bronckart,

que remete para uma realidade linguística que pode ser tão vasta, quão vasta pode ser o enunciado.

Rastier sintetiza, de forma clara, as duas perspectivas de contexto linguístico empregues pelos dois autores supracitados: «[...] *on peut opposer deux conceptions du contexte linguistique: comme zone d'extension, relativement au signe; ou de restriction, relativement au texte*» (Rastier, 1998:99). O contexto linguístico remete para duas realidades diferentes. Por um lado, o contexto que envolve a unidade de significação em análise, correspondendo à noção de contexto utilizada por Décles; por outro, o contexto perspectivado como uma zona linguística mais alargada, o texto na sua amplitude, que está em consonância com o ponto de vista de Bronckart.

O contexto linguístico, entendido como uma zona de extensão, permite alargar a informação conceptual, semântica e sintáctica da unidade de significação através da sua análise. O contexto linguístico, entendido como uma zona de restrição, define-se como uma zona de localidade, sendo que «[...] *un contexte n'est plus le contexte d'un mot, mais un passage du texte. Un mot peu certes servir à choisir un passage dans une recherche [...]; mais ensuite, c'est le passage qui devient l'unité étudiée*» (Rastier, 1998:99).

Tanto Bronckart como Décles têm por objecto de estudo o texto, sendo que Bronckart desenvolve um método de análise baseado num modelo psicológico, ao passo que Décles desenvolve um método de exploração contextual que tem por intuito a construção de um sistema informático baseado em conhecimentos exclusivamente linguísticos, patentes no texto (cf. Décles, 1997:215).

Bronckart (1997) analisa o funcionamento do texto e a sua arquitectura interna partindo de pressupostos teóricos sociodiscursivos subjacentes à sua criação, enquanto que a abordagem de Décles é puramente linguística, pretendendo este autor criar um modelo que parta de uma descrição linguística fina, por considerar que as unidades linguísticas são índices pertinentes para o desenvolvimento de tarefas específicas. Assim sendo, isolar e observar o contexto que envolve uma unidade linguística é primordial, uma vez que este contém os elementos necessários para melhor entender a unidade linguística em análise, deixando marcas à sua volta, que são passíveis de serem identificadas.

Pensamos que Bronckart pretende chegar a um entendimento sobre o funcionamento do discurso e Décles à génese do texto, para dele retirar informação com vista ao tratamento automático do texto.

Novamente, pensamos que os contextos linguísticos e extralinguísticos se subjugam, na medida em que o indivíduo e o objecto produzem-se reciprocamente. O contexto desempenha uma função mediadora entre o indivíduo e o mundo objectivo, ideia que Rastier exprime do seguinte modo: *«[...] de même que l'environnement est un concept intermédiaire entre celui de monde propre du sujet et celui du monde objectif, il semble que le concept de «contexte» soit un pseudo-concept, qui tout à la fois souligne et voile l'unité entre le mot et le texte, comme le lien entre la situation et la pratique en cours»* (Rastier, 1998:110).

O contexto é um conceito operatório, mediador entre o intralinguístico, o linguístico e o extralinguístico, sendo os intervenientes os principais protagonistas da mediação.

### 2.3 Intervenientes

Para que haja comunicação, é necessário a intervenção de vários actores. No caso do acto da «*parole*», os intervenientes são, no mínimo, dois, que ora assumem o papel de falante, ora o de ouvinte. O outro, está presente, podendo ou não intervir no acto de comunicação.

Mas se entendermos a noção de «*parole*» no seu sentido mais lato, podemos considerar que o discurso é uma forma de «*parole*». Falar em «*parole*» na perspectiva saussuriana implica que se tenha subjacente a noção de língua, o segundo elemento da construção dicotómica. À semelhança desta, podemos construir uma outra dicotomia constituída pelo par língua - discurso: «*En effet, si on entend par parole, avec Ferdinand de Saussure, la réalisation de la langue dans un événement de discours, la production d'un discours singulier par un locuteur singulier, alors chaque texte est par rapport à la langue dans la même position d'effectuation que la parole*» (Ricoeur, 1986:138). Ricoeur considera que o acto da escrita pode ser equiparado ao acto da «*parole*», na medida em que estes actos representam uma forma de actualização da língua: «*La fixation par l'écriture survient à la place même de la parole, c'est-à-dire à la place où la parole aurait pû naître*» (Ricoeur, 1986:138).

Tal aproximação dicotômica deve-se ao facto de, em consonância com Saussure, se poder considerar a língua como um produto social e a «*parole*», tal como o discurso como um acto individual dependendo da vontade e da inteligência do indivíduo.

No caso da realização da comunicação através do texto, sendo este entendido como um produto material com contornos físicos finitos, que fixa um discurso proferido por um ou por vários indivíduos inseridos em contextos específicos, a relação entre o autor e o leitor caracteriza-se pelo facto de, regra geral, a comunicação se concretizar na ausência de um dos intervenientes, sendo o texto, objecto físico, o único elo de ligação entre eles.

A actualização da língua é uma actividade cooperativa que tem implicações ao nível dos seus intervenientes. Para Charaudeau «*Tout acte de communication est un objet d'échange entre deux instances, l'une d'énonciation l'autre de réception, dont le sens dépend de la relation d'intentionnalité qui s'instaure entre celles-ci*» (1997:15), o que pressupõe a existência de três lugares distintos associados à produção, ao produto e à recepção da transacção verbal mediatizada pelo texto. Assim, a produção integra-se no lugar das condições de produção, o produto, no lugar da construção do discurso e a recepção no lugar da interpretação.

O processo de transacção verbal realiza-se segundo quatro princípios que se complementam: o princípio da alteridade, o princípio da pertinência, o princípio da influência e o princípio da regulação (cf. Charaudeau, 1995:98-100).

Para que a comunicação se realize, é necessário que os intervenientes tenham referentes em comum e que estejam aptos a reconhecer o universo de referências do outro (princípio da pertinência), não obstante o autor e o leitor se reconhecerem como sujeitos estrutural e psicologicamente diferentes (princípio de alteridade): «*Ainsi, ce principe dit que chacun des partenaires est engagé dans un processus réciproque (mais non symétrique) de reconnaissance de l'autre, dans une interaction le légitimant du même coup dans son rôle, ce qui est une condition de validation de l'acte du langage*» (1995:99). O autor produz um discurso ou um texto com a intenção de exercer influência sobre o outro, com o intuito de fazê-lo agir, modificá-lo (princípio da influência). O leitor, por seu turno, consciente ou inconscientemente, regula o jogo de influências de modo a viabilizar o intercâmbio verbal (princípio da regulação).

Charaudeau considera que o processo acima explanado entra no campo daquilo que ele convencionou designar por «*machine médiatique*» (1997). Este autor trabalha com o discurso de informação mediática, podendo, no entanto, a sua perspectiva ser alargada ao discurso científico, na medida em que a institucionalização do discurso científico permite estabelecer simetrias com outros discursos, onde a noção de espaço é uma componente importante da legitimação do discurso.

Os intervenientes são uma das componentes que Bronckart insere nos parâmetros extraverbais, mais especificamente na «*situation d'action langagière interne*», que identificámos como sendo a acção interiorizada dos indivíduos, tanto do autor como do destinatário.

Este autor qualifica os intervenientes uns de activos, os autores, e os outros de passivos, os destinatários. O interveniente activo é o produtor de texto que incorpora toda a situação social e psicológica. O interveniente passivo é o destinatário do texto, o leitor, que tal como o interveniente activo, também incorpora uma situação social e psicológica, cuja função é ler, compreender e interpretar o texto.

O interveniente activo, neste caso, o especialista produtor de um texto de especialidade, estrutura o seu saber em função de um público modelo, do qual possui uma imagem prototípica. O público é assim constituído por um conjunto de indivíduos que, numa primeira instância, são considerados intervenientes passivos, na medida em que, no momento em que o texto está a ser construído, eles não exercem qualquer tipo de acção directa sobre o mesmo.

No entanto, são intervenientes, porque estão presentes na mente do autor que constrói uma imagem do potencial leitor, instituindo deste modo a competência do leitor-modelo: «[...] *prever o próprio Leitor-Modelo não significa apenas «esperar» que exista, significa também conduzir o texto de forma a construí-lo*» (Eco, 1979:59). É com base na imagem que construiu da competência do seu leitor-modelo que o especialista estrutura a complexidade conceptual e linguística que se reflecte num todo coeso: o texto de especialidade. A elevada competência atribuída ao leitor-modelo, permite ao especialista elaborar um texto com um alto nível de cientificidade que se repercute, geralmente, num texto denso e rigoroso, incompreensível para o

não-especialista não conhecedor dos conceitos, das redes conceptuais e referenciais que compõem o discurso patente no texto.

Qualquer texto tem por função ser lido: «[...] lire, c'est en toute hypothèse, enchaîner un discours nouveau au discours du texte» (Ricoeur, 1986:152). O leitor intervém, no momento em que o autor do texto considera o seu trabalho de escrita acabado e o põe à disposição do outro. A mudança dos meios histórico e social em que o texto foi produzido tem repercussões imanescentes ao nível dos conhecimentos veiculados pelo texto: «Si l'entour change, le contenu du texte change aussi puisqu'il est immanent à une situation de communication maintenant modifiée. En règle générale, dans le cas d'un changement d'époque ou de culture, il s'appauvrit, par déperdition des connaissances» (Rastier, 1989:51).

Nestas circunstâncias, deixamos de considerar o leitor um interveniente passivo para assumir a função de interveniente activo, com funções e graus de intervenção diferentes da do autor.

Segundo Rastier, o leitor adopta duas posturas: a da leitura produtiva ou a da leitura descritiva. No caso da leitura produtiva, o leitor «[...] réinterprète le texte au gré du récepteur, pour la faire correspondre à des situations et des référents nouveaux, quitte à les réécrire en partie» (Rastier, 1989:51); no caso da leitura descritiva, o seu objectivo é a de restituir o conteúdo do texto, «[...] en reconstituant l'entour de la communication initiale» (Rastier, 1989:52).

Estamos convictos de que o leitor especialista faz, regra geral, uma leitura produtiva. O leitor especialista apropria-se do texto, reestrutura a informação recebida à luz do seu próprio saber e conhecimento. Ao ler o texto,

o leitor - especialista acrescenta uma mais-valia ao seu saber. acrescenta saber ao saber que já detém e, teoricamente, leitura após leitura deste e de outros textos, adquire competências para produzir novos textos. Os textos por ele produzidos são adicionados aos textos já existentes e, conjuntamente, desencadeiam processos semelhantes noutros leitores. Como refere Rastier: «*En somme, tout texte est un centon*» (Rastier, 1989:39).

O conjunto de todos estes textos constituem arquivos documentais e arqueológicos (Foucault, 1969) comuns aos membros de uma comunidade científica: «*Dans cette chaîne les premiers interprétants servent de tradition pour les derniers interprétants [...]*» (Ricoeur, 1986:158).

### **3. Tipologia de tipos ou tipologia de géneros?**

Para atingirmos o objectivo que nos propomos levar a cabo, é imprescindível seleccionar criteriosamente um conjunto de textos da especialidade que será o objecto de análise, e que, de ora em diante, será considerado o principal objecto de estudo. Tal decisão obriga-nos a reflectir sobre os critérios subjacentes à selecção, à organização e à sistematização dos textos que passarão a constituir o *corpus* de referência.

O conceito de texto de especialidade é demasiadamente genérico e complexo por forma a permitir um tratamento uniforme de todas as ocorrências de textos produzidos no seio de uma mesma comunidade científica. O(s) público(s) a quem se dirige(m), a pluralidade das situações e o enquadramento

espacio-temporal em que os diversos textos são produzidos e consumidos são tão divergentes, que se impõe a constituição de uma tipologia.

Por tipologia entende-se: «[...] *un ensemble de procédures permettant de reconnaître et d'établir des corrélations entre deux ou plusieurs objets sémiotiques, ou leur résultat*» (Greimas, 1979:403). O estabelecimento de uma tipologia implica a classificação e a reunião de um conjunto de textos, que mantém entre si relações de semelhança ao nível das respectivas macro e/ou microestruturas, sob uma mesma etiqueta, através da identificação de regularidades de uns conjuntos de textos, por oposição às regularidades de outros conjuntos. Para Petitjean, classificar significa «[...] *opérer un rapprochement entre des objets langagiers, toujours univoques et différents, à partir de caractéristiques (propriétés) qui leur sont communes*» (Petitjean, 1989:93). Na identificação, designação e classificação das regularidades, que podem ser da ordem do sociológico, do psicológico, do histórico e/ou do linguístico, entre outros, é que se situa a essência da problemática da tipologia. Contudo, uma classificação traduz sempre, de forma mais ou menos clara, uma redução de informação das partes para se poder compreender o todo.

A construção de uma tipologia é, assim, um processo cognitivo que passa pelo estabelecimento de três parâmetros: «1) *Une base typologique*. 2) *Un domaine d'application caractérisé*. 3) *Une description des formes de mise en rapport entre la base et le domaine d'application*» (Petitjean, 1989:95). Identificar uma base tipológica é o ponto de partida para a estruturação de qualquer tipologia, uma vez que é absolutamente indispensável estabelecer os

critérios que devem ser usados para a sua construção. Os critérios de base devem ser adequados ao objecto de estudo, devendo ser reconhecidos e aplicados parcial ou totalmente, estabelecendo a relação entre a base, o objecto de estudo e a finalidade da análise.

Petitjean distingue três tipos de classificações: as classificações homogéneas, as classificações intermédias e as classificações heterogéneas.

Por classificações homogéneas entende-se « [...] *travaille[r] à partir d'une base typologique unique et homogène sous la forme d'un modèle abstrait, souvent déductif, qui se réalise complètement ou partiellement dans un domaine d'application donné*» (Petitjean, 1989:97). O resultado obtido através da classificação homogénea é designada por «tipo de texto».

Contrariamente à classificação homogénea, a classificação intermédia parte de uma base tipológica heterogénea, cujos critérios se baseiam no modo enunciativo, na intenção da comunicação e nas condições de produção. As tipologias que resultam dessas classificações reflectem a «*mise en situation des textes*», das quais procedem as tipologias enunciativas, as tipologias comunicacionais e as tipologias situacionais que o autor designa por «tipos de discursos».

Petitjean designa de heterogénea a terceira e última classificação: «[...] *la base typologique est totalement hétérogène et comprend des critères qui relèvent des foyers classificatoires aussi différents que l'intention communicative, le mode énonciatif, la stratégie illocutoire, [...]*» (Petitjean, 1989:117). Ao objecto de estudo, obtido através desta classificação, Petitjean atribui a designação de «*génères de textos*».

Se a tipologia das classificações, assim como as tipologias resultantes da classificação intermédia, vingaram e são referidas como tipologias pertinentes e funcionais, o mesmo não sucedeu relativamente às designações utilizadas para distinguir os objectos resultantes de cada uma das classificações, que Petitjean designou por tipo de texto, tipo de discurso e género de textos.

Existe, hoje, alguma confusão terminológica no que diz respeito à classificação de textos e discursos em géneros ou tipos.

Tal constatação leva-nos a repensar a problemática da distinção entre texto e discurso. Algumas questões podem ser colocadas a este respeito. A partir de que realidades é que podemos conceber tipologias? Tivemos oportunidade de referir autores como Adam que, em determinadas circunstâncias, optaram por não distinguir texto de discurso. Significa isso, que podemos falar indiferentemente de tipos de textos e de tipos de discursos, de géneros de textos e de géneros de discursos?

A tentativa de resposta a tais questões ocasiona a reflexão sobre o género, conceito que, embora geralmente associado ao de discurso, tem repercussões ao nível da classificação dos textos.

Para ilustrar o que atrás ficou dito, optámos por apresentar um quadro revelador da diversidade terminológica, reflexo de posicionamentos teóricos diferentes, face aos mesmos objectos de estudo:

	Tipos de texto	Tipos de discurso
Schaeffer (1986)	-	-
Petitjean (1989)	+	+
Rastier (1989)	+	-
Bronckart (1996)	-	+
Maingueneau (1998)	-	+
Adam (1999)	-	-

	Género de texto	Géneros de discurso
Schaeffer (1986)	+	-
Petitjean (1989)	+	-
Rastier (1989)	-	+
Bronckart (1996)	+	-
Maingueneau (1998)	-	+
Adam (1999)	-	+

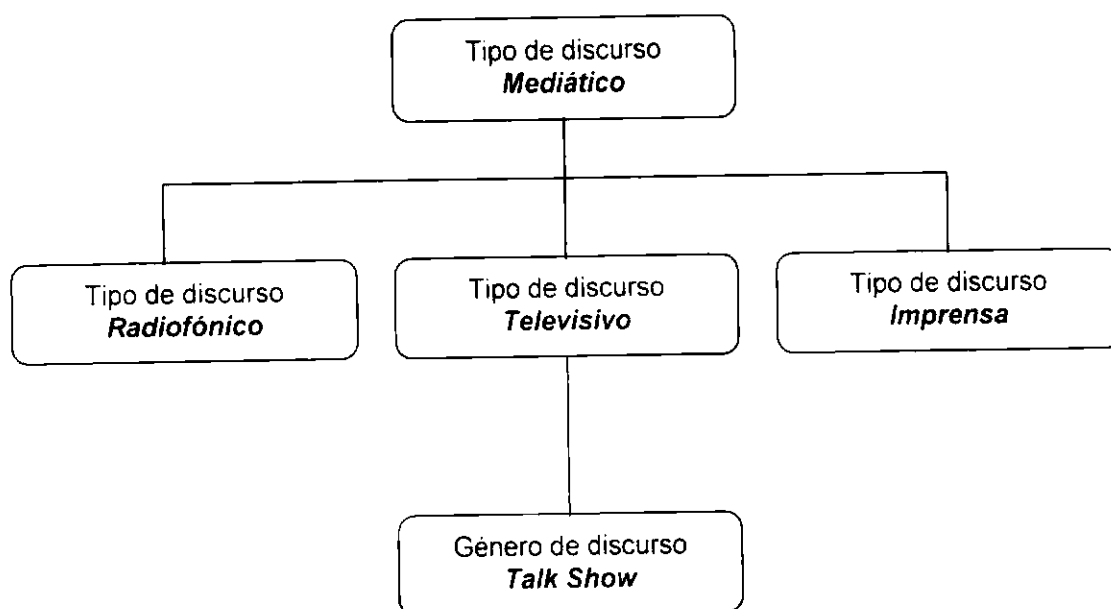
Classificar os textos em tipos é uma actividade de cariz sociológica e não linguística, sendo o género o conceito constitutivo da acção verbal: «*Les genres de discours relèvent de divers types de discours, associés à de vastes secteurs d'activité sociale*» (Maingueneau, 1998:47). Para este autor, a elaboração de tipologias de discursos e de textos só é pertinente caso se tenha em conta o género, conceito fundador da actividade verbal : «*Tout texte relève d'une catégorie de discours, d'un genre de discours*» (Maingueneau, 1998:45). Falar em tipos de discurso, significa estabelecer parâmetros em harmonia com os diversos sectores da sociedade; cada um desses sectores produz discursos e textos que podem ser classificados em tipos. A investigação científica é um

sector de actividade, cuja produção textual e discursiva constitui um tipo, porque é o produto de uma actividade social específica.

Quando o discurso ou texto é analisado como o resultado de uma actividade verbal, interdependente de parâmetros extralinguísticos, que são uma componente intrínseca da actividade da «*parole*» saussuriana, estamos perante os géneros, «[...] *c'est-à-dire des dispositifs de communication qui ne peuvent apparaître que si certaines conditions socio-historiques sont réunies*» (Maingueneau, 1998:47).

Existe, assim, uma relação de dependência hierárquica entre os conceitos de género e de tipo. É necessário que determinadas condições sociais sejam preenchidas para que exista um tipo de discurso ou de texto do qual decorrem géneros.

Na perspectiva defendida por Maingueneau, a relação entre tipo de discurso e género de discurso pode ser esquematizada do seguinte modo:



Os indivíduos reconhecem estruturas - do nível do textual, discursivo, histórico, pragmático, social, psicológico, etc. - que, obedecendo a protótipos interiorizados, lhes permitem ordenar os textos ou discursos em géneros. Segundo Kleiber, «[...] *le prototype est le meilleur exemplaire qui est reconnu comme étant le meilleur par les sujets*» (Kleiber, 1991:104). Kleiber relembra que o reconhecimento de um objecto como sendo prototípico implica uma percepção individual, estando, na globalidade, em conformidade com a percepção do colectivo.

O conceito de protótipo implica um modelo conceptual genérico que nos leva a referir Schaeffer, autor que considera a «generacidade» como sendo uma característica intrínseca à constituição do género. Segundo Schaeffer, o conceito de género pressupõe uma exterioridade genérica, que este define do seguinte modo: «[...] *c'est la procédure qui consiste à «produire» la notion d'un genre non à partir d'un réseau de ressemblance existant entre un ensemble de textes, mais en postulant un texte idéal dont les textes réels ne seraient que les dérivés plus ou moins conformes, de même que selon Platon les objets empiriques ne sont que de copies imparfaites des Idées éternelles*» (1986:190). O conceito de género surge, assim, como uma realidade abstracta ideal, do qual decorrem textos empíricos, que são ocorrências de um ou de diversos géneros.

Schaeffer acredita que um texto específico não tenta reproduzir de forma fiel um modelo genérico, tenta sim modificá-lo: «[...] *tout texte modifie «son» genre: la composante générique d'un texte n'est jamais (sauf exceptions rarissimes) la simple réduplication du modèle générique constitué par la classe*

de textos (supposés antérieurs) dans la lignée desquels il se situe. Au contraire pour tout texte en gestation le modèle générique est un «matériel» parmi d'autres sur lequel il «travaille» (Schaeffer, 1986:197).

Os conceitos de texto e de género implicam-se mutuamente, sendo que a não existência e o desconhecimento de géneros impossibilitaria a comunicação, na medida em que seria necessário estar constantemente a reinventar novas formatações ou modelos para poder proferir um discurso adequado a uma situação específica: «*Si les genres de discours n'existaient pas et si nous n'en avions pas la maîtrise, et qu'il faille les créer pour la première fois dans le processus de la parole, qu'il nous faille construire chacun de nos énoncés, l'échange verbal serait impossible*» (Bakhtine, 1984, p.285).

Independentemente de se privilegiar o texto em detrimento do discurso, ou vice-versa, ambos decorrem de géneros, cabendo-nos, neste âmbito, reflectir sobre que objectos teóricos é que se concebe tipologias: tipologias de tipos de discurso ou de texto ou tipologias de géneros de texto ou de discurso?

Optar, em consciência, recorrer a um género de discurso, implica o conhecimento das condições necessárias para que o discurso seja bem sucedido em termos sociais: «*Acte de langage d'un niveau de complexité supérieure, un genre de discours est soumis lui aussi à un ensemble de conditions de réussite*» (Maingueneau, 1998:51).

Tais condições obedecem a critérios diversos da ordem do social e do linguístico. No campo sociológico, são requeridas quatro condições: a finalidade do discurso, a legitimação dos intervenientes no discurso, a adequação do discurso ao espaço e ao momento em que é proferido e a

identificação ou a utilização do suporte adequado, em conformidade, particularmente, com a intenção do discurso (cf. Maingueneau, 1998:50-54). A última condição situa-se no campo do linguístico e diz respeito à própria organização textual.

A distinção entre tipo e género, preconizada por Maingueneau, não é adoptada por Adam, que rejeita, na globalidade, o conceito de tipo de texto: *«J'ai consacré plusieurs articles, entre 1987 et 1992, et en un livre (1992) à tenter d'expliquer pourquoi il est à mon sens, profondément erroné de parler de «types de textes»»* (Adam, 1999:82). Para este autor, o texto é uma unidade demasiadamente heterogênea e complexa para se poder identificar regularidades linguísticas, que possam ser consideradas o sustentáculo de uma tipologia. Sugere que se elaborem tipologias das sequencialidades presentes no texto, reduzindo para a sua análise a extensão do texto.

Partindo do pressuposto de que Adam tem o mesmo entendimento que Maingueneau da concepção de tipo, é compreensível que rejeite as tipologias de tipos de textos, uma vez que a sua abordagem é predominantemente linguística: *«C'est dire qu'au-delà des formes élémentaires de séquentialisation dont je vais parler, des codifications sociales - génériques - sont à l'oeuvre dans toute communication verbale (Rastier, 1989:37), codifications qui de toute évidence, ne relèvent pas d'une théorisation strictement linguistique et que je suis bien obligé, de ce fait, d'écarter provisoirement de ma réflexion»* (Adam, 1992:15).

Mais recentemente, Adam afirma aceitar o conceito de tipo, quando utilizado como sinónimo de género: *«Si l'on tient à parler de «types» au niveau*

*global et complexe des organisations de haut niveau, il ne peut s'agir que de pratiques sociodiscursives, c'est-à-dires de genres»* (Adam, 1999:83). Segundo este autor, o conceito de género permite congregiar análises linguísticas e sociais.

Em 1985, Bronckart defendia a existência de tipos de textos, conceito que abandona posteriormente: *«La notion de «type de texte» sera notamment abandonnée, au profit de celles de genres de texte et de types de discours»* (1996:13). Enquanto Maingueneau e Adam utilizam preferencialmente o conceito de género associado ao de discurso, Bronckart usa a noção de género associada ao conceito de texto: *«Et dans la mesure où tout texte s'inscrit nécessairement dans un ensemble ou dans un genre, nous avons adopté l'expression de genre de texte, de préférence à celle de discours»* (Bronckart, 1996:78). Pensamos, no entanto, que a definição de tipo utilizada por Bronckart não corresponde à de Maingueneau. Bronckart considera haver uma multiplicidade de géneros, provocada pela multiplicidade de actividades humanas, o que dificulta a abordagem ao género. Em contrapartida, os segmentos (segmentos de argumentação, de diálogo, ...) que entram na sua composição são finitos e podem ser identificados pelas suas regularidades linguísticas, o que justifica a sua escolha teórica: *«Ces segments différents entrant dans la composition d'un genre constituent le produit d'un travail particulier de sémiotisation ou de mise en forme discursive, et c'est pour cette raison que nous les qualifierons désormais de discours. Dans la mesure où ils présentent de fortes régularités de structuration linguistique, nous considérons qu'ils relèvent de types, et nous utiliserons donc pour les désigner l'expression*

*de type de discours, de préférence à celle de type textuel*» (Bronckart, 1996:78). Bronckart diverge dos outros autores, não só quanto à distinção de texto e de discurso, como também relativamente às definições de tipo e de género. Tal divergência deve-se ao posicionamento teórico assumido por cada um dos autores. Bronckart define, *a priori*, o seu enquadramento teórico: «*Nos propositions théoriques relèvent d'une psychologie du langage, elle-même orientée par les principes épistémologiques de l'interactionisme social*» (Bronckart, 1996:11). Tal nível de análise situa-se a um nível diferente da de Adam, cuja perspectiva se insere na óptica da linguística textual, posicionando-se Maingueneau na senda da análise do discurso.

Da definição de discurso dada por Rastier, depreende-se que a noção de género está ligada à de discurso e a noção de tipo à de actividade social. Para este autor discurso é «*[un] ensemble d'usages linguistiques codifiés attaché à un type de pratique sociale*» (Rastier, 1991:247). Esta ideia já tinha sido expressa anteriormente: «*Puisque le sens d'un texte est immanent à une situation de communication, et que les situations typiques déterminent des types de textes, les parcours interprétatifs qui permettent de re(construire) le sens textuel sont déterminés par le type de texte*» (Rastier, 1989:35). Contrariamente a Maingueneau e a Adam, Rastier admite distinguir claramente os conceitos de discurso e de texto: «*Nous écartons l'usage inspiré de la linguistique anglo-saxonne qui assimile discours et texte*» (1989:39). Consequentemente, segundo este autor, não pode proceder-se à constituição de tipologias indiferenciadas de textos e de discursos. Utiliza a expressão «tipo de texto» para referir-se às diversas funções perceptíveis num texto, que designa por

tipologia das funções ou por tipologia funcional dos textos, enquanto relaciona o conceito de género de discurso com as práticas sociais associadas a um tipo de uso linguístico, considerando que: «*Tout locuteur participe à plusieurs pratiques sociales et doit donc posséder plusieurs compétences discursives. Chacune suppose la maîtrise d'un ou plusieurs genres*» (1989:40). Se o texto e o discurso estão interligados através do género, então poder-se-á identificar as funções textuais mais usuais no seio de um género de discurso, permitindo assim chegar a um maior entendimento do mesmo.

Pensar o texto implica pensar o discurso, e pensar texto e discurso implica pensar tipo e género. Para Rastier «[...] *un genre est ce qui rattache un texte à un discours*» (Rastier, 1989:40), não deixando este autor, porém, de considerar as relações de dependência mútuas que unem os tipos aos géneros: «*Les régimes interprétatifs des types de discours (politiques, religieux, etc.) sont enfin spécifiés par les contrats interprétatifs propres aux genres*» (Rastier, 1998:109).

Na sequência deste debate, adoptamos a perspectiva de Rastier que defende a distinção entre texto e discurso, perspectiva contrária à defendida pela escola anglo-saxónica. O ponto de vista, segundo o qual se aborda o enunciado é que pré-determina a sua aproximação textual ou discursiva.

Deste modo, consideramos que a elaboração de tipologias de tipos, assim como as tipologias de géneros resulta da observação das condições sociodiscursivas em que foi produzido um texto, testemunho representativo de uma colecção de textos que, no seu conjunto, caracteriza um discurso.

Se, pelo contrário, os critérios de classificação forem consequência de uma análise linguística apurada, estamos, em nosso entender, claramente perante uma tipologia de textos, organizada segundo denominadores comuns aos textos que a compõem. Esta tipologia é designada de tipologia *a posteriori*: «*La typologie a posteriori part des données textuelles elles-mêmes pour proposer des catégories issues de l'examen et du regroupement de traits linguistiques*» (Habert, Fabre, Isaac, 1998:40).

A tipologia *a priori* corresponde ao que nós convencionámos denominar de tipologia de tipos de discursos e de tipologia de géneros de discursos, que Habert, Fabre, Isaac, influenciados por Sinclair e Ball (1996) definem como: «*La typologie a priori s'appuie sur des catégories pré-établies de genre [...], de domaine ou sur la détermination du contexte de production [...]*» (Habert, Fabre Isaac, 1998:40).

O texto decorre de um discurso, porém não podemos tomar a parte pelo todo. O estudo minucioso de um texto científico, não nos permite afirmar conhecer e dominar o discurso científico. Julgamos, contudo que um *corpus* reflecte as propriedades de um tipo de discurso ou de vários tipos de discursos, tendo em conta que um texto é uma ocorrência de um ou vários géneros, que estabelece(m) o elo de ligação entre o texto e o(s) discurso(s).

#### 4. Constituição do *corpus* em detecção remota

##### 4.1. Definição da área de especialidade em análise: Detecção Remota

O termo Detecção Remota, ou Teledetecção, designa todo o sistema de aquisição de informação sobre os objectos terrestres - reunidos nos três grandes domínios: água, solo e vegetação – sem contacto físico entre o instrumento de medida e o objecto. Nesta acepção geral, a Detecção Remota reúne um vasto conjunto de conhecimentos científicos, técnicos e tecnológicos que conduzem à observação da Terra a partir do espaço, essencialmente por satélite e por avião, utilizando as propriedades da radiação electromagnética.

Lillesand e Kiefer definem Detecção Remota como «(...) *the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area or phenomenon under investigation*» (Lillesand; Kiefer, 2000:1). Estes autores baseiam a definição, exclusivamente, na distância que separa o objecto observado e o sensor, entendido este como dispositivo de medida de energia electromagnética emitida e/ou reflectida pela superfície terrestre.

Campbell (1996:5), pelo contrário, não dá ênfase à distância entre objecto observado e o sensor, mas sim ao papel da radiação electromagnética (REM), ou seja, à propagação da energia através do vazio ou de um meio natural, sob forma de ondas vectoriais transversais, à velocidade da luz e em interacção com a matéria: «*Remote sensing is the practice of deriving information about earth's land and water surfaces using images acquired from an overhead perspective, using electromagnetic radiation in one or more*

*regions of the electromagnetic spectrum, reflected or emitted from earth's surface» (Campbell;1996:5).*

As definições anteriores, baseadas, como referimos, na noção de distância física e de REM, não contemplam, contudo, todos os processos que envolve a Detecção Remota, nem fazem referência à tecnologia que lhe está associada. Aquilo que é omissa nas definições de Lillesand e Kiefer (2000:1) e de Campbell (1996:5) está claramente expresso na definição de Bonn, segundo o qual: *«La télédétection est la discipline scientifique qui regroupe l'ensemble des connaissances et des techniques utilisées pour l'observation, l'analyse, l'interprétation et la gestion de l'environnement à partir de mesures et d'images obtenues à l'aide de plates-formes aéroportées, spatiales, terrestres ou maritimes. (...)» (Bonn,1992:3).* Este autor admite, na sua definição, embora de forma indirecta, a evidente noção de distância que incorpora o formante "tele-" - *« (...) à partir de mesures et d'images obtenues à l'aide de plates-formes aéroportées (...)»* -, mas não lhe atribui a importância dada por Lillesand e Kiefer e por Campbell.

A análise dos mais recentes manuais de referência em Detecção Remota dirigidos ao ensino e à investigação e escritos em francês, castelhano ou em inglês – além dos citados podemos referir, entre outros, Girard e Girard (1999), Chuvieco (1996), Jensen (1996) - permite concluir que a definição do conceito de Detecção Remota abarca:

- A noção de distância, a partir da qual se pode definir a Detecção Remota como o conjunto de métodos e técnicas utilizadas para determinar as características físicas e biológicas dos objectos por

intermédio de medidas efectuadas a distância, sem contacto material com os mesmos objectos. A noção de REM, meio pelo qual é possível conhecer um objecto ou um conjunto de objectos distribuídos espacialmente, em função: de um ou mais sinais electromagnéticos numa ou mais bandas do espectro; da evolução do sinal electromagnético do objecto no tempo (hora, ano, estação do ano ou década); da repartição espacial do objecto; das ligações espaciais e temporais do objecto com outro que lhe seja vizinho, mas de natureza química, geoquímica ou biológica diferente.

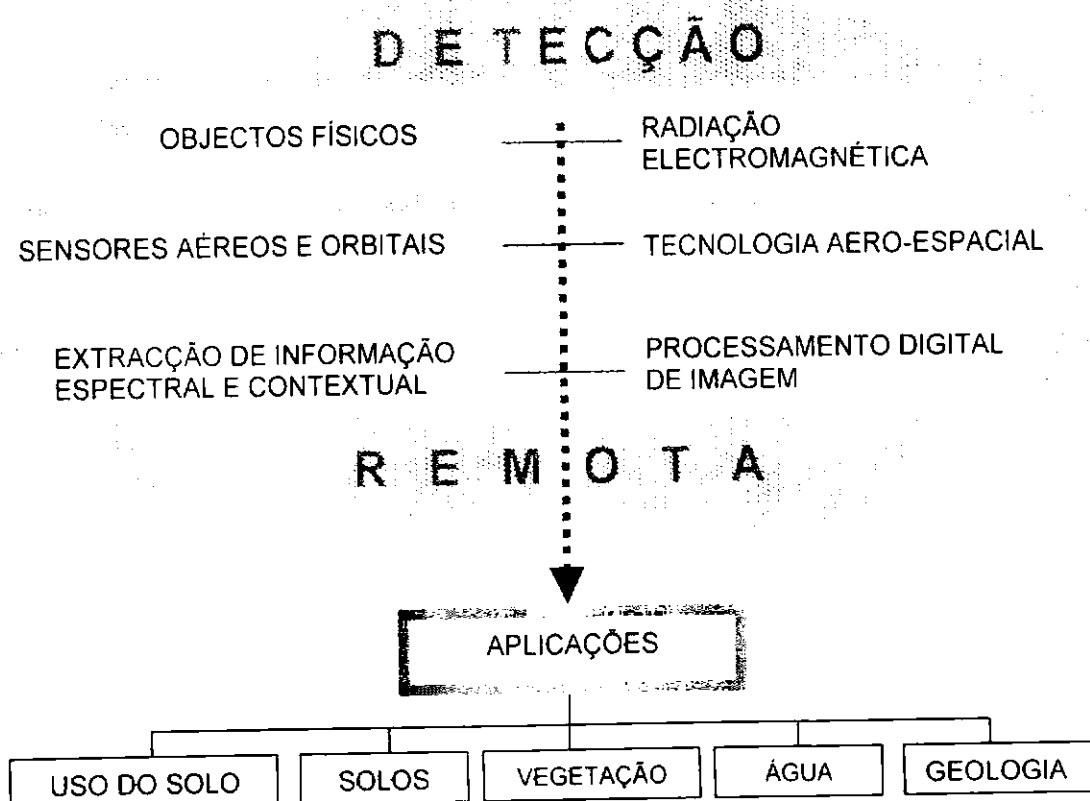
- A noção de aplicação, que faz da Detecção Remota um conjunto de métodos e de técnicas necessários para determinar a natureza de um objecto ou de um conjunto de objectos, considerando o seu comportamento espectral num contexto espacial e temporal definido. A noção de aplicação inclui a não menos importante noção de processamento digital de imagem – ou tratamento numérico de imagens – que incorpora os procedimentos para extrair os dados de Detecção Remota (fotografias aéreas, imagens de satélite, termografias, imagens radar, entre outras), a informação sobre a natureza e a forma e a função dos objectos à superfície da Terra.

No contexto das Ciências Sociais e Humanas, nomeadamente em Geografia – uma das Ciências que mais uso faz das imagens de satélite -, a Detecção Remota sustenta-se nos dois elementos seguintes:

- os pixels de uma imagem não são considerados apenas como uma medida de REM, mas sim como um meio de acesso a uma informação característica do espaço geográfico estudado;
- a análise, visual ou numérica, reporta-se, frequentemente, aos objectos geográficos, às texturas, às estruturas e às formas de organização espacial que traduzem o estado de interacções múltiplas e as respostas – actuais ou passadas – das sociedades às modificações das paisagens.

Pode concluir-se, então, que o conceito de Detecção Remota, em sentido amplo, não engloba apenas os meios (plataformas e sensores aéreos e orbitais) e os processos (REM, radar, sonar, etc.) que permitem obter dados sobre a Terra mas abrange também o seu posterior tratamento no âmbito de uma determinada aplicação ou duma “reflexão pura”, que conduz ao conhecimento científico.

O esquema seguinte ilustra a forma como pode ser entendido o conceito de Detecção Remota.



A ideia central esquematizada é a de a Detecção Remota não representa apenas o conhecimento tecnológico exigido pela observação da Terra; esta noção é entendida também como uma cadeia de processamentos numéricos, que dá forma às aplicações nas mais variadas áreas do conhecimento (Ciências Agrárias, Geografia, Oceanografia, Geologia, Climatologia, Ecologia, Pedologia, Urbanismo, etc.).

Os textos produzidos pela comunidade científica portuguesa utilizam sempre as denominações Detecção Remota e Teledetecção como sinónimos totais. Este fenómeno constata-se, por exemplo, em artigos publicados em revistas de especialidade, em comunicações apresentadas em colóquios, em seminários e em congressos, em teses de mestrado e de doutoramento e até na designação oficial das disciplinas dos planos de estudos das Licenciaturas e dos Mestrados.

Em Portugal, existe um número considerável de universidades onde se lecciona Detecção Remota. Esta disciplina ensina-se, mais precisamente, em 15 universidades e institutos politécnicos, incluindo os sectores público e privado, de acordo com o apuramento realizado a partir dos dados disponíveis na Internet em Julho de 2001 – relativos ao número de instituições de ensino superior e politécnico e conteúdos programáticos de Detecção Remota. O ensino é ministrado quer ao nível das Licenciaturas (74%) quer ao nível dos Mestrados (26%); cerca de 50% das disciplinas estão circunscritas às universidades e institutos politécnicos da região de Lisboa. Os conteúdos programáticos compreendem, na sua maioria, áreas de aplicação terrestre ou costeira; as aplicações ao domínio oceânico são ainda muito incipientes.

Excluindo apenas um caso, em que o título evidencia uma clara articulação da Detecção Remota com os Sistemas de Informação Geográfica – o que significa, no nosso entender, uma transversalidade clara da Detecção Remota em muitas áreas de aplicação –, as restantes disciplinas ou se intitulam «Detecção Remota» ou de «Teledetecção» (51%) ou especificam, na designação, a área de especialização (49%); por exemplo: «Detecção Remota em Ecologia» (Departamento de Ecologia da Universidade de Évora); «Cartografia e Teledetecção» (Departamento de Geologia da Faculdade de Ciências da Universidade de Lisboa).

Da análise efectuada podemos concluir, quer do ponto de vista técnico e tecnológico quer do ponto de vista científico, nada de substancial diferencia a definição dos conceitos de «Detecção Remota» e de «Teledetecção», sendo estes dois termos igualmente correctos do ponto de vista da sua formação. Por esta razão, são usados indiscriminadamente, sem gerar ambiguidades na comunicação entre especialistas. O que pode afirmar-se, ainda que tendo por base uma observação empírica, é que existe uma relação entre a utilização de uma ou outra designação e a "escola" que serviu de formação científica aos docentes e investigadores: «Detecção Remota» - do inglês *Remote Sensing* - é mais utilizada pelos docentes e investigadores de formação anglófona e «Teledetecção» - do francês *Téledétection* - pelos especialistas de formação francófona.

## 4.2. Constituição de tipologias

### 4.2.1. Tipologia de tipos de discurso

A tipologia de tipos de discurso em Detecção Remota que propomos baseia-se na observação e na análise dos textos escritos em português europeu que constituem o *corpus*. A primeira síntese dessa análise, esquematizada na figura seguinte, pressupõe que:

a) a Detecção Remota tem o seu próprio discurso e estando também presente nesta área do saber o discurso das outras ciências, das outras tecnologias e das outras técnicas que lhe são conexas. Na verdade, entre, por exemplo, a Detecção Remota e a Geografia existe um objecto de estudo comum: a superfície da Terra. Contudo, enquanto a Detecção Remota pertence, em primeiro lugar, ao largo espectro das Ciências da Informação e de Observação da Terra, a Geografia está orientada para a explicação das relações espaciais entre lugares geográficos caracterizados por diferentes atributos. Neste sentido, a Detecção Remota poderá ser entendida como uma técnica auxiliar da Geografia, fornecendo-lhe a informação espacial necessária à compreensão dos fenómenos localizados. Este exemplo pretende mostrar a dificuldade - e até a relativa artificialidade - em estabelecer uma tipologia de tipos de discurso numa área do conhecimento com elevado nível de transdisciplinaridade.

b) a "atomização" das aplicações da Detecção Remota produz "discursos temáticos" fundamentados na ciência, na tecnologia e na técnica que a Detecção Remota incorpora. Esses discursos, que designamos "temáticos" por terem origem nas mais diversas ciências, associam o

## TIPOS DE DISCURSO EM DETECÇÃO REMOTA

### DISCURSO CIENTÍFICO

#### DISCURSOS DAS CIÊNCIAS FUNDAMENTAIS:

- Física (física da atmosfera, electromagnetismo, óptica, optico-electrónica, metrologia, acústica)
- Matemática (lógica, derivação, primitivação, trigonometria)
- Estatística (estatística descritiva, estatística inferencial e probabilidades, estatística multivariada e classificações de dados)
- Informática (algoritmia, linguagens de programação)

#### DISCURSOS CIENTÍFICOS INTERDISCIPLINARES:

- Ciências Geográficas
- Ciências da Terra
- Ciências Sociais
- Telecomunicações

### DISCURSO TECNOLÓGICO

#### DISCURSOS DAS TECNOLOGIAS:

- Tecnologia aérea
- Tecnologia espacial
- Tecnologia acústica
- Instrumentos de registo de dados
- Sistemas de armazenamento e de transmissão de dados
- Instrumentos de tratamento de dados
- Instrumentos de edição de dados
- Instrumentos de gestão e de difusão de dados

#### DISCURSOS TECNOLÓGICOS INTERDISCIPLINARES:

- Geodesia
- Topografia
- Cartografia
- Sistemas de Informação Geográfica (SIG)

### DISCURSO TÉCNICO

#### DISCURSOS DAS TÉCNICAS:

- Processamento Digital de Imagem
- Análise visual de imagem
- Técnicas analógicas
- Técnicas digitais
- Técnicas aerofotogramétricas
- Técnicas cartográficas
- Técnicas topográficas
- Técnicas infográficas

#### DISCURSOS TÉCNICOS INTERDISCIPLINARES:

- "Técnicas temáticas"
- Técnicas de trabalho de campo
- Técnicas geomáticas

## DISCURSOS EM DETECÇÃO REMOTA APLICADA

- Ambiente
- Cartografia e Geodesia
- Geografia
- Geologia
- Hidrologia
- Meteorologia
- Oceanografia
- Ordenamento do Território
- Pedologia
- Outras aplicações

conhecimento fundamental - no sentido estrito de discurso das ciências fundamentais – ao conhecimento tecnológico e ao conhecimento mais próximo da acção, para darem origem aos discursos da Detecção Remota Aplicada.

Apesar de a Detecção Remota ser uma área em que é grande a transdisciplinariedade, podem aí ser reconhecidos quatro tipos pertinentes de discurso, isolados a partir da intervenção conjugada de dois tipos de fronteiras:

- a fronteira da natureza do conhecimento – responsável pela determinação do conhecimento científico, do conhecimento tecnológico e do conhecimento técnico;
- a fronteira dos conteúdos, determinada a partir da utilização das unidades terminológicas em contexto de discurso de especialidade - conhecimentos de Detecção Remota aplicados em disciplinas muito diversas.

A fronteira da natureza do conhecimento implica a utilização de um conjunto de conceitos característicos e próprios de cada um dos supracitados tipos de discurso. Esta especificidade, reconhecível nos três tipos de discurso em Detecção Remota – discurso científico, discurso tecnológico e discurso técnico – não encontra paralelo no quarto tipo, designado discurso em Detecção Remota Aplicada, dado que, por exemplo, termos diferentes podem apresentar significados iguais em contextos discursivos de especialidade. Por exemplo, os termos *resolução espectral* e *separação espectral* – muito frequentes no *corpus* constituído - denominam o mesmo conceito; o primeiro é vulgarmente utilizado por geólogos, geógrafos, entre outros especialistas, enquanto o segundo é

especialmente utilizado pela comunidade científica da Engenharia do Ambiente e da Biologia.

Assim, entendemos que as fronteiras mais rígidas que servem esta tipologia de tipos encontram-se na natureza dos discursos e não natureza dos conteúdos temáticos.

O primeiro tipo de discurso refere-se ao que designamos por discurso científico. Este tipo decorre dos domínios científicos fundamentais a que está associada a Detecção Remota: a Física, a Matemática, a Estatística e a Informática. Pela própria natureza do conhecimento que estas disciplinas envolvem – o conhecimento das «ciências duras» - os discursos, recorrendo à formulação matemática, revelam o uso de métodos quantitativos e experimentais (por exemplo, da física atmosférica e do electromagnetismo). Os fenómenos são apresentados como fenómenos isolados e controlados cientificamente para se proceder à modelação no espaço e no tempo (é o caso, por exemplo, dos discursos que apresentam os resultados da espectrometria da vegetação ou dos solos). A complexidade da formulação matemática, nomeadamente da primitivação e da derivação, é, claramente, o traço mais emblemático deste tipo de discurso.

O segundo tipo, o discurso tecnológico, envolve uma componente fortemente ancorada na descrição dos sistemas de observação da Terra. A este nível, os discursos apresentam a panóplia de tecnologia disponível (sondas espaciais, satélites de observação passiva da Terra, sistemas RADAR e LIDAR e SONAR, aviões para observação a muito grande escala) para medir, armazenar e/ou transmitir às estações de recepção em Terra, o sinal que

contém, de forma codificada, a informação sobre os objectos. A utilização intensiva de noções tecnológicas relativas aos instrumentos de medida, e à sua importância estratégica para a evolução do conhecimento sobre a Terra, constitui o denominador comum dos discursos que designamos por tecnológicos.

O discurso técnico trata o conhecimento tácito da Detecção Remota. Isso significa que se trata de um discurso que agrupa os fundamentos dos discursos enunciados anteriormente para, a partir deles, construir a base de conhecimentos técnicos necessários à percepção espectral, espacial e temporal dos objectos observados a distância.

O discurso técnico integra, por isso, o tratamento da informação contida em imagens (de satélite, de avião, de RADAR, etc.) para, a partir dele, reconhecer as características físicas, biológicas e funcionais dos objectos. Este discurso baseia-se na descrição e análise da investigação e desenvolvimento das técnicas de processamento digital de imagem; integra uma forte componente de informática aplicada, na medida em que traduz formalizações matemáticas e estatísticas em algoritmos para reconhecer objectos, quer pelos atributos relativos à sua radiometria quer pelos atributos relativos à forma.

Estes discursos caracterizam-se, na maior parte dos casos, pela existência de noções frequentemente utilizadas pelas aplicações temáticas. Por isso, anunciam a passagem entre a Detecção Remota e aquilo a que poderemos chamar Detecção Remota Aplicada. Na verdade, as diferentes disciplinas e as diferentes comunidades de investigadores fazem uso do discurso técnico da Detecção Remota, quer para equacionar problemas de

reconhecimento dos objectos (geográficos, ambientais, pedológicos, meteorológicos, oceanográficos, etc.), usando a ciência e a tecnologia aeroespacial, quer para renovar os métodos de investigação.

O quarto tipo de discurso em Detecção Remota decorre da utilização implícita dos três discursos anteriores. Os discursos em Detecção Remota Aplicada, conforme o designamos, são discursos claramente disciplinares; ou seja, utilizam intensivamente os conceitos do ramo de saber especializado - Ciências do Ambiente, Engenharia Geográfica, Geografia, Urbanismo, Biologia, Engenharia Florestal, Engenharia Agrícola, etc. – sem deixarem de reconhecer os fundamentos científicos, tecnológicos e técnicos dos outros discursos. O que distingue este tipo de discurso dos anteriores (trata-se de um problema de fronteira de conteúdos, já referida) é, por outro lado, a elaboração da problemática a partir da qual se elaboram as hipóteses, se tratam os dados espectrais, se criticam os resultados e se retiram as conclusões.

Estas etapas (equação do problema, elaboração das hipóteses, tratamento dos dados, crítica dos resultados e conclusões) estão fortemente condicionadas pela especificidade do discurso disciplinar no seio do qual ocorrem. É, por isso, frequente encontrar tendências para a sinonímia, para a polissemia e para a homonímia.

Em síntese, podemos considerar que a fronteira da natureza do discurso funciona bem na delimitação dos três primeiros tipos, porque a sua unidade discursiva é mais forte, por apresentar maior identidade. Por outro lado, a fronteira da natureza dos conteúdos - com manifesta repercussão nas unidades terminológicas - é fundamental para balizar, por vezes dificilmente, os

discursos da Detecção Remota Aplicada, dado os numerosos campos de aplicação disciplinar e a diversidade de comunidades científicas que produzem estes discursos de especialidade.

#### 4.2.2. Tipologia de géneros de discursos

A prática da análise dos tipos de discurso revela, também, o género de discurso. Esta prática denuncia, com relativa facilidade, uma separação entre as abordagens estritamente científica, estritamente técnica e estritamente didáctica.

Considerando o conjunto de textos recolhidos e informatizados, podemos identificar sete géneros, de acordo com o discurso dominante, conforme se indica no quadro seguinte:

	Científico	Técnico	Didáctico
Teses (Mestrado e Doutoramento)	•		
Artigos em revistas de Detecção Remota*		•	
Artigos em revistas temáticas**	•		
Comunicações em eventos de Detecção Remota*		•	
Comunicações em eventos temáticos**	•		
Manuais e sebatas			•
Relatórios de projectos de I&D	•	•	

\* As revistas e os eventos tendem a incluir a Detecção Remota no grupo das Tecnologias de Informação Geográfica (TIG), grupo que associa os Sistemas de Informação Geográfica e a Cartografia Digital às tecnologias e sistemas de observação da Terra.

\*\* "Temático" significa, neste contexto, a produção científica das diferentes áreas disciplinares (Engenharia Geográfica, Geografia, Geologia, Urbanismo, Cartografia, etc.).

As teses, de Mestrado e Doutoramento, denotam estruturas discursivas predominantemente científicas; aliam o quadro teórico, baseado no tipo de discurso científico – leia-se das ciências fundamentais -, à experimentação técnica desenvolvida com recurso à algoritmia. Neste género, o texto revela forte conceptualização matemática, estatística e informática, quase sempre aliada à ilustração através de imagens resultantes do processamento digital de dados de satélite. Embora as imagens tenham sido subtraídas ao texto, assim como as fórmulas, julgamos que elas podem servir como material para explorar uma vertente ainda pouco investigada: a ligação do texto à imagem e o “modelo semântico” que lhe está associado. Na realidade, a imagem constitui outro tipo de discurso que aqui deixamos como pista de investigação futura: Que discurso substitui uma imagem? Que terminologia estará associada a esse discurso?

Os artigos em revistas de Detecção Remota evocam, em primeiro lugar, a técnica como justificação da investigação; o 'paradigma dominante' é, por isso, a resolução de um problema concreto decorrente da vontade explícita da comunidade científica em determinar uma solução ou melhorar um processo já existente (por exemplo, a aplicação de redes neuronais no melhoramento das classificações automáticas de dados de satélite).

Os artigos em revistas temáticas articulam o discurso predominantemente científico (proposição de um problema novo, de índole geográfica, geológica, urbanística, etc.) com o discurso técnico (uma técnica nova, ou reformulada), ao serviço da resolução de um problema. É frequente encontrar neste género a

reflexão teórica sobejamente debatida à luz de novos dados e de novas técnicas de Detecção Remota.

As comunicações em colóquios, congressos e seminários de Detecção Remota privilegiam o discurso técnico, enquanto as apresentadas em eventos temáticos revelam o lugar primordial da reflexão científica conjugada com a utilização de novas tecnologias. Muitas vezes é difícil entender se os resultados são apresentados como evolução real da temática analisada (novas interpretações e novas descobertas sobre um mesmo ou diferente problema), se apresentados como 'propaganda científica' de novos materiais (os dados de Detecção Remota) à disposição da comunidade. Este género mostra, por vezes, marcas muito fortes da oralidade, facto que não se estranha considerando as circunstâncias sociodiscursivas em que ocorre.

Os manuais e sebentas são um género muito distinto dos restantes, com dominância do discurso didáctico, combinando o detalhe e o rigor dos discursos científico, tecnológico e técnico. Apesar de a produção deste género não proliferar em Portugal, podemos referir que os manuais existentes mostram grande abundância terminológica. Nos textos representativos deste género de discurso acedemos, por exemplo, às formas extensas de siglas e de acrónimos, sendo a maior parte destas formas abreviaturas de unidades terminológicas multilexémicas de línguas inglesa e francesa. É o caso dos acrónimos pixel (picture element), RADAR (radio detection and ranging) e de HRV (Haute Résolution dans le visible). Por outro lado, também é neste género de discurso que o especialista tem a preocupação de descrever e de definir os conceitos que utiliza no decorrer do seu discurso.

Os relatórios de projectos de I&D revelam uma combinação equilibrada dos discursos científico e técnico. O atributo mais peculiar é o de insistir na pertinência da demonstração dos resultados e no erro envolvido na formulação das hipóteses iniciais. Constituem, frequentemente, um género que denota ambiguidades na definição de conceitos, facto que se explica pelo contexto em que tais textos se produzem (o da investigação de novas soluções para novos problemas) e pela necessidade, aparente, de designar a mesma coisa com termos diferentes. Esta observação também se justifica na medida em que este género é elaborado em circunstâncias disciplinares muito diversas, quer quanto à substância quer quanto à forma como é produzido.

O *corpus* que, no âmbito desta dissertação, serve de base à análise e descrição é composto por textos representativos dos géneros acima referidos, que decorrem de discursos científicos, tecnológicos, técnicos, assim como da detecção remota aplicada.

## **Capítulo III**

---

### **Denominação: expressões nominais multilexémicas**

1. Denominação
2. Expressões nominais multilexémicas
  - 2.1. Lexias
  - 2.2. Sinapse
  - 2.3. Nome composto
  - 2.4. Colocação
  - 2.5. Frasemas
3. Fraseologia

## **Capítulo III - Denominação: expressões nominais multilexémicas**

### **1. Denominação**

A criação e a produção científicas implicam a sistematização, a estruturação e, eventualmente, a modelização do conhecimento, sendo estas actividades o resultado da conceptualização dos objectos. O objecto é entendido, aqui, como o que é dado à experiência, o que é independente do espírito. A organização e a hierarquização da estrutura conceptual são factores essenciais para a apropriação dos conhecimentos em qualquer área de especialidade, o que demonstra que a identificação e a delimitação dos conceitos pressupõem um julgamento por parte do especialista, tarefa que não é isenta de dificuldades. Diz-nos Moles a este respeito: *«Nous vivons au milieu des phénomènes vagues, de choses imprécises, de situations perpétuellement variables dans lesquelles il faut nous décider, réagir ou agir, prendre position. Si vagues soient-elles, pourtant, toutes les choses apparaissent à notre conscience comme des objets conceptuels, nous leur donnons des noms, et nous faisons sur elles des opérations, mentales d'abord, pratiques ensuite, à nos risques et périls»* (1995:13).

A terminologia, enquanto teoria, debruça-se sobre a relação entre o conceito e o termo. A identificação do conceito e sua definição estão na base da terminologia uma vez que, teoricamente, a terminologia utiliza uma metodologia onomasiológica, isto é, isola o conceito e estabelece,

seguidamente, a relação entre os vários conceitos que constituem o sistema conceptual dentro de uma área do saber. Para que o conceito seja perceptível e a comunicação possível é necessário denominar o conceito. Segundo Felber, «*Les notions ne peuvent pas être perçus par les sens. C'est pourquoi il faut aux fins de la communication, utiliser des symboles linguistiques, qui sont aussi des objets individuels*» (Felber,1987:141). Repare-se que, na versão francesa da referida obra, o termo inglês «*concept*» é traduzido pelo termo francês «*notion*», em detrimento do seu equivalente mais próximo, «*concept*». Também a norma ISO/FDIS 1087-1:2000 segue esta orientação teórica, considerando conceito e noção sinónimos um do outro.

Por sua vez, Rey distingue claramente noção de conceito, definindo a noção como «*l'objet de connaissance - qui pose un objet et le définit en tant que visée de son activité*» e o conceito como «*l'acte de pensée de l'objet de pensée (idée) en tant qu'abstrait et que général*»(1979:29).

Não pretendemos debater mais longamente esta questão, optando por utilizar, preferencialmente, o conceito de conceito, respeitando, neste texto, a designação utilizada por cada um dos autores.

Muitas têm sido as definições que se tem dado de conceito numa perspectiva terminológica, isto é, funcionalista, com o intuito de contribuir para uma metodologia das práticas terminográficas. A mais conhecida, porque normalizada, é a definição que nos é dada pela norma ISO/FDIS 1087-1:2000 (E/F), que define conceito como sendo uma «*[...] unité de connaissance créée par une combinaison unique de caractères*». Sager apresenta outras definições, tais como a seguinte: «*A concept is a unit of thought, produced by*

*grouping of individual objects related by common characteristics»* (Sager 1990:23).

No entanto, pensamos que a definição mais próxima do que acabámos de expor é a definição que provém de uma alteração que a ex-União Soviética propõe à norma ISO 1087:1990 (E/F), que obedece à seguinte formulação: «*A concept is a coherent group of judgements concerning an object whose nucleus is made up of those judgements which reflect the inherent characteristics of the object»* (Sager, 1990:23).

Todas as definições de conceito aqui apresentadas são fluidas; fluidos são também os conceitos nas áreas de especialidade, particularmente nas ciências «moles» (cf. Moles, 1995). A não delimitação clara do conceito reflecte-se na transmissão dos conhecimentos, isto é, na comunicação especializada, assim como nas definições desses mesmos conceitos e, portanto, na relação biunívoca entre o conceito e sua denominação, isto é, o termo.

Efectivamente, o especialista e/ou o terminólogo apreende o conceito delimitando-o em função da sua própria formação, em função da visão que tem do mundo. O especialista olha e reflecte sobre o conceito, sob um determinado prisma que tentará retratar o melhor possível, no momento de transmitir o conhecimento que tem acerca desse conceito. Como observa Béjoint, «*On pourrait dire que le concept n'existe que dans l'esprit du spécialiste»* (1997:20). A comunicação, neste caso, especializada, diz respeito à transmissão de conhecimentos, havendo um intercâmbio de formas e de representações entre o emissor e o receptor. Para Lamizet, «*[...] le réel [...] est exclu [de la*

*communication], car le réel ne se communique pas: il se vit, il se perçoit, il fait l'objet d'expériences» (Lamizet, 1992:19). Embora seja o emissor a transmitir o seu conhecimento, impregnado da sua visão do mundo, é o receptor que filtra a mensagem e a reconstrói em função do seu conhecimento: «L'acte de pensée s'accomplit dans l'individu, mais les résultats de cet acte sont communiqués et contrôlables par d'autres. Car la communication langagière est toujours conceptuelle: elle est le résultat de l'utilisation des concepts» (Czap, 1989 : 71).*

Ao abordarmos os conceitos operatórios de «conhecimento» e de «conceito», temos consciência de que nos encontramos ao nível do extralinguístico, na medida em que a função do termo é a de veicular informação científica e técnica. No plano linguístico, a inserção do termo no enunciado, especializado ou não, é inteiramente dependente do sistema da língua em causa.

Em terminologia, é fundamental não confundir os dois planos distintos de análise, o extralinguístico e o intralinguístico, que coabitam, mas não se sobrepõem. Não pode confundir-se conceito com significado, uma vez que este último, na perspectiva de Saussure, não se dissocia do significante, formando ambos uma unidade indivisível, o signo linguístico. Em contrapartida, o conceito tem uma existência independente; existe além da denominação. Desta forma, são os traços conceptuais que compõem um conceito e o distinguem de outro, não devendo ser confundidos com os traços distintivos, traços semânticos ou semas dos significados (cf. Manuelito, 1995:137-141).

As componentes que constituem os conceitos são por sua vez, também elas, conceitos. Tal como defende Felber, «Le caractère est un élément de la

*notion qui sert à décrire ou à identifier une qualité d'un objet individuel. Le caractère lui-même est une notion»*(1984:99). No entanto, este autor faz a distinção entre as características, componentes dos conceitos e as propriedades, componentes dos objectos: «[...] *the concept is a thought unit composed of characteristics assigned to properties of objects»* (Felber, 1994 / 1995:164).

Sager utiliza, indistintamente, características e propriedades, afirmando que as características do conceito são cada uma das propriedades que o descreve: «*Characteristics can be expressed as properties of the concept»* (Sager, 1990:25).

Se não nos parece de grande relevo teórico a diferenciação entre propriedade e característica, não podemos dizer o mesmo relativamente à utilização indiferenciada de sema e de propriedade ou característica, uma vez que os planos de análise em que estes termos devem ser aplicados são distintos. Podemos, no entanto, estabelecer aproximações teóricas e considerar que os semas estão para a significação como as propriedades ou características estão para o conceito.

A actualização linguística do conceito faz-se através da denominação, não havendo, necessariamente, uma relação de correspondência entre os traços conceptuais que possui um conceito e os elementos de nominação que constituem a respectiva denominação. Thoiron (1996 : 514) defende que ao designar-se um conceito, nem todos os traços conceptuais são activados, considerando que os que não são designados podem, eventualmente, ser inferidos a partir dos conhecimentos enciclopédicos dos locutores.

O texto escrito é uma das formas privilegiadas ao qual o especialista recorre para transmitir e aceder ao conhecimento. Partindo de um corpus textual, o terminólogo acede às denominações, podendo, numa primeira instância identificar alguns traços conceptuais por intermédio dos elementos de nomação. Os outros, os que não são perceptíveis através da leitura do texto, são os que nos são dados a conhecer pelos especialistas: *«Les traits conceptuels les plus immédiatement accessibles aux études terminologiques (par opposition à ceux qui n'apparaissent que par l'interrogation des spécialistes) sont ceux auxquelles correspondent des éléments de nomination. L'accès au concept se fait alors par la médiation du terme et en particulier de ses éléments de nomination»* (Béjoint; Thoiron, 1997:515). Mas nem sempre o especialista consegue dar todos os esclarecimentos de que necessitamos porque, como refere Béjoint (1997:20), não se pode conhecer um conceito na sua globalidade, mesmo quando se é especialista. Além do mais, nem tudo se pode exprimir através da língua.

Porém, o que nos motiva são as actualizações da língua mais em concreto, as unidades terminológicas estruturalmente complexas, cuja função no texto é a de denominar conceitos.

Assim, as denominações são, na sua essência, unidades terminológicas que designam conceitos específicos de uma ou de várias esferas do conhecimento e que, teoricamente, são comuns aos indivíduos que constituem uma comunidade de comunicação especializada.

De um ponto de vista morfossintáctico, os termos estruturalmente complexos são constituídos por dois ou por mais lexemas separados por um espaço em

branco, podendo entre eles existir gramemas autónomos (artigos, preposições, ...) ou uma marca gráfica, tal como o hifen: *imagem de satélite*, *imagem-textura*, *filmes infravermelhos coloridos*.

As denominações que nos interessa analisa, são unidades terminológicas, cujas estruturas morfossintácticas são monolexémicas ou multilexémicas (cf. Capítulo 3, p. 112).

Numa perspectiva meramente linguística, as unidades terminológicas são, na sua maior parte, expressões nominais complexas; são essas as estruturas que nos interessa identificar, descrever e seleccionar através de processos inerentes ao tratamento automático da língua natural.

## **2. Expressões nominais multilexémicas**

### **2.1. Lexias**

A tentativa de determinar as fronteiras das expressões nominais complexas recorrendo para o efeito ao estabelecimento de critérios que permitem apartar uma sequência fixa de uma sequência livre, tem sido largamente explorada.

A multiplicidade de terminologias criadas pelas diferentes escolas linguísticas para designar, descrever e analisar as unidades de significação constituídas por dois ou mais elementos lexicais, originariamente com autonomia morfológica, sintáctica e semântica, separados por um espaço em branco interligados entre si por valores que se fundamentam em critérios linguísticos e/ou conceptuais, é um indício da dificuldade em estabelecer

condições necessárias e suficientes para a delimitação e subsequente definição dos vários tipos de expressões nominais complexas.

A dificuldade em designar e em definir as unidades complexas, que não são abrangidas pela composição, segundo a definição dada pela gramática tradicional, leva alguns autores a propor novas designações e a delinear critérios que permitem a demarcação das diferentes estruturas que podem assumir as unidades linguisticamente complexas.

Em 1967, Pottier introduz a noção de *lexia* como elemento de base da estruturação sintáctica, opondo-a a morfema, unidade mínima de significação e a palavra, unidade mínima construída. A *lexia* caracteriza-se por ser uma unidade de comportamento, constituída por palavras, que se realiza como uma unidade de comportamento significativa em discurso.

Pottier considera que as *lexias* se actualizam em *lexias* simples, em *lexias* compostas e em *lexias* complexas.

A *lexia* simples corresponde inequivocamente à palavra simples e à palavra derivada da gramática tradicional. Em contrapartida, a distinção entre *lexia* composta e *lexia* complexa não está suficientemente demarcada. Da *lexia* composta, Pottier dá-nos a seguinte definição: «*La lexie composée contient plusieurs mots déjà en partie ou totalement intégrés (graphiquement, ou dans leur comportement tactique: un brise-glace*» (Pottier, 1967:17). No que se refere à *lexia* complexa, esta é definida como: «[...] *une séquence plus ou moins figée de mots: faire une niche, en avoir plein le dos, pomme de terre, au fur et à mesure, [...]*» (Pottier, 1967:17).

Estas duas definições não são claramente distintivas, uma vez que ambas remetem claramente para uma coesão morfossintáctica e semântica. De um ponto de vista teórico, as duas classes de entidades linguísticas, pressupostamente distintas, que são cobertas pelas duas designações, *lexia composta* e *lexia complexa*, não são, em consonância com as respectivas definições, demarcadas umas das outras. A primeira definição está próxima da definição de palavra composta dada pela gramática tradicional, em que o hífen é um elemento identificador. Por sua vez, a definição de *lexia complexa*, e de acordo com os exemplos apresentados, remete pressupostamente para todas as estruturas que indiciam um determinado grau de coesão e que não são abrangidas pela definição de *lexia composta*.

Não nos parece que as definições apresentadas por Pottier sejam suficientemente esclarecedoras para permitirem a elaboração de critérios para demarcar a *lexia composta* da *lexia complexa*.

Pottier tem, no entanto, o mérito de apontar para a existência de sequências fixas que se distinguem das sequências compostas. Tal facto é evidenciado pela introdução do modificador «complexo», que deixa supor um grau de estabilidade linguística mais ou menos fixa entre os constituintes de uma estrutura morfossintáctica, que podem co-ocorrer com uma frequência mais ou menos elevada. Os critérios que permitem a identificação dos parâmetros para a avaliação da sua estabilidade não se confundem com os critérios que permitem definir a composição da gramática tradicional.

O termo de *lexia complexa* é, em nosso entender, um genérico extremamente produtivo e operacional, na medida em que é frequentemente

empregue para designar toda e qualquer sequência discursiva fixa, independentemente da descrição linguística a que é sujeita.

Também Mel'čuk (1995), no âmbito da Teoria Linguística Sentido-Texto, privilegia a noção de lexia. Para este autor, a lexia é a unidade de base da lexicologia, que deve ser considerada nas suas vertentes semântica, sintáctica e léxico-combinatória, e que engloba todas as suas possíveis actualizações: simples, composta e complexa: «[...] *il nous suffit de dire que lexie ou unité lexicale, est soit un mot pris dans une acception bien spécifique (=lèxeme), soit encore une locution, elle aussi prise dans une acception bien spécifique (=phrasème)*» (Mel'čuk, 1995:16).

Mel'čuk não se detém nas premissas que permitem distinguir um lexema de um frasema. A lexia assume o valor de qualquer sequência morfolexical que possa ser objecto de descrição lexicológica, conferindo este autor uma importância maior à lexia, em detrimento das regras que permitem agrupar o conjunto das lexias em unidades coesivas de um nível estrutural superior: «*En exagérant quelque peu, on pourrait dire que l'ensemble des lexies est la langue. En effet, une langue est constituée de lexies et de règles servant à la manipulation de ces dernières. Les règles qui réunissent les lexies en syntagmes, les syntagmes en phrases, et les phrases en discours sont donc nettement secondaires par rapport aux lexies - en ce sens que leur nature et leur forme sont déterminées par les lexies*» (Mel'čuk, 1995:17).

A descrição semântica do léxico está no centro do quadro teórico de Mel'čuk, sendo a lexia monolexémica ou multilexémica (cf. Mel'čuk, 1995:45),

que assume o valor de lexema e de frasema, respectivamente, o pólo de atracção, relegando a sintaxe para um plano secundário.

## 2.2. Sinapse

Em 1974, Benveniste isola estruturas morfossintácticas complexas que se distinguem dos compostos, propondo um novo termo para as designar: a sinapse: *«Il consiste en un groupe entier de lexème, reliés par divers procédés, et formant une désignation constante et spécifique»* (Benveniste, 1974:172). Para este autor, a sinapse designa realidades linguísticas estruturalmente complexas, cujo estatuto não está claramente definido. No entanto, a sua elevada produtividade, particularmente, nas nomenclaturas técnicas, justifica uma observação mais minuciosa.

A sinapse é identificada como um fenómeno predominantemente sintáctico, que se define por um conjunto de propriedades, que se elevam a sete: 1. a natureza sintáctica que liga os constituintes; 2. a utilização dos determinantes *de* e *a* que unem os seus constituintes; 3. a ordem linear determinado mais determinante, 4. a conservação da forma plena dos seus constituintes e a livre escolha do substantivo ou do adjectivo; 5. a ausência de artigos antes do determinante; 6. a possibilidade de expansão de um ou de outro constituinte e, finalmente, 7. o carácter único do significado (cf. Benveniste, 1974: 172-173).

Os critérios elaborados por Benveniste demonstram a pertinência do estabelecimento de parâmetros para a delimitação das fronteiras entre

composição e outras estruturas coesas. Porém, os critérios apresentados nem sempre são distintivos e, conseqüentemente, funcionais.

Repare-se, a título de exemplo, no segundo critério, que diz que para se estar perante uma sinapse é necessário a presença das preposições *de* ou *a*. No entanto, podemos apresentar inúmeros exemplos em que tal facto não acontece: *imagem textura*, *camião reservatório*, entre muitos outros. Estas duas unidades não são consideradas compostas, uma vez que não obedecem aos critérios que nos permitem identificá-las como tal, mas encaixam na definição dada por Benveniste, apesar de não obedecerem aos parâmetros por ele elaborados. Nestes dois exemplos, além de não recorrerem às preposições *de* ou *a*, ficando o segundo parâmetro anulado, também o terceiro critério não pode aplicar-se, na medida em que a ordem linear determinado mais determinante não está claramente identificada.

Os critérios elaborados são da ordem do sintáctico, uma vez que Benveniste classifica os compostos e as sinapses como pertencentes às funções sintácticas. No entanto, o último critério situa-se no âmbito do semântico e do conceptual: «*C'est toujours et seulement la nature du désigné qui permet de décider si la désignation syntagmatique est ou n'est pas une synapsie [...]*» (Benveniste, 1974:173).

Benveniste acaba por nos dar indícios que revelam de modo claro que não nos podemos reger exclusivamente por critérios sintácticos para delimitar fronteiras entre as várias sequências estruturalmente complexas que abundam, essencialmente, nos textos de especialidade.

Acreditamos que os critérios que determinam as regras linguísticas para a delimitação de fronteiras dos diversos tipos de unidades complexas, derivam de diferentes níveis de análise linguística, que devem ser explorados caso se demonstrem profícuos e operacionais.

### **2.3. Nome composto**

O nome composto é a designação mais frequentemente utilizada para se designarem as expressões nominais complexas, sendo na composição que este fenómeno linguístico encontra a sua maior expressão. No entanto, os critérios para a delimitação da entidade linguística para a qual o nome composto remete, nem sempre são muito rigorosos e nem sempre reúnem o consenso dos linguistas.

Em 1992, Corbin coloca algumas questões fundamentais, que permitem reequacionar a noção de nome composto, bem como os critérios para o estabelecimento das suas fronteiras. A autora aponta duas abordagens linguísticas distintas às palavras compostas que, embora frequentemente efectuadas de forma uniaxial, se complementam.

A primeira abordagem à palavra composta é considerada externa, uma vez que tem por objectivo a sua identificação enquanto unidade lexical, sendo o resultado final o reconhecimento de uma sequência complexa lexicalizada. A segunda abordagem é interna, na medida em que tenta compreender a formação de palavra a fim de entender os processos linguísticos que lhe estão

subjacentes, sendo o objecto de estudo, neste caso, as sequências complexas cuja estrutura interna corresponde aos pressupostos teóricos pré-definidos.

Propondo-se descrever o papel que assumem os constituintes na estruturação do sentido das palavras construídas por derivação e por composição, que se opõem às palavras simples, definida pela incapacidade de se decompor em elementos mais pequenos, independentemente do seu valor, Corbin entende que: «*Un mot complexe est une unité lexicale complexe construite par des règles lexicales conjoignant des unités lexicales à pouvoir référentiel*» (Corbin, 1992).

Embora a diferenciação entre a derivação e a composição possa por vezes ser fluida e as denominações destes dois processos difíceis de fixar, tal diferenciação é no mínimo reveladora de uma percepção intuitiva que deve ser perscrutada.

Corbin parte do pressuposto teórico de que as palavras construídas podem ser definidas como: «[...] *l'association, gouvernée par une règle, d'un schéma d'interprétation à une structure*» (Corbin, 1992), concluindo que o modo de construção do sentido se faz por duas vias distintas, consoante se trate de palavras construídas por afixação ou por composição. Contrariamente aos afixos, os constituintes das palavras compostas têm um poder referencial, situado ao mesmo nível da base da palavra construída: «[...] *un mot composé est un mot complexe dont le sens est construit par la combinaison de constituants à pouvoir référentiel. La règle de composition régit la combinaison de ces constituants, leur nature, leur ordre dans lequel ils apparaissent et impose un schéma d'interprétation plus ou moins déterminé*» (Corbin, 1992).

Esta inferência é inerente à estrutura interna da palavra composta, sendo objecto de estudo da morfologia lexical.

Mas a descrição morfolexical da palavra composta não nos permite, só por si, distingui-la de outras estruturas livres ou de outras construídas por outras vias, que não sejam as regras de composição ou de afixação. Por este motivo, Corbin restringe a significação da etiqueta palavra composta a um «[...] *sous-ensemble d'unités lexicales construites par des règles lexicales de composition*» (Corbin, 1992), sendo todas as estruturas que não se definem por esta via outra coisa que não palavras compostas, não obstante serem formadas por palavras construídas.

Em 1997, Corbin define palavra composta como «[...] *des unités lexicales construites morphologiquement à partir des constituants à sens référentiel*» (Corbin, 1997:82), o que corresponde à lexicalização de unidades morfologicamente construídas. As unidades que correspondem à lexicalização de sequências sintacticamente construídas, são designadas por «*mots construits syntaxiquement*», recorrendo a autora à etiqueta unidade polilexemática para designar, genericamente, as unidades lexicais morfossintácticas complexas.

O que nos interessa reter da abordagem teórica de Corbin, é a distinção que estabelece entre sequências lexicalizadas, sequências em via de lexicalização e sequências coesas.

Tal como G. Gross, Corbin sustenta a sua teoria na identificação das propriedades que determinam a solidariedade entre os constituintes que firmam as expressões complexas. Enquanto que G. Gross elabora uma teoria dos

compostos, partindo do pressuposto de que a coesão é escalar, Corbin considera a coesão uma consequência da lexicalização, produto terminado, resultante de um processo lexical que conduz à coesão: «*En effet, la lexicalisation envisage soit un processus en devenir, soit un processus accompli, mais pas le processus en tant que potentialité, c'est-à-dire dans sa dimension prédictible*» (Corbin, 1997:58). O que a motiva, não é o produto acabado, mas os factores ligados ao modo de construção que favorece a lexicalização, que se caracteriza por: a) a sua extensão longitudinal entre duas extremidades e a complexidade da sua estrutura, b) o tipo de estrutura e c) a natureza do mecanismo gramatical que está na origem da estrutura (cf. Corbin, 1997:58-59).

A coesão de uma sequência é o produto de um processo de transição da sintaxe para o léxico, em que as propriedades sintácticas e semânticas de cada um dos constituintes que compõem a sequência ficam diluídas pelas relações de solidariedade que mantêm entre si. Deste modo, «[...] *c'est [...] la lexicalisation qui favorise le figement et donc la fixation de certaines propriétés*» (Corbin, 1997:87).

As unidades polilexemáticas podem, assim, ser classificadas em função de critérios sintácticos e morfológicos, que permitem a identificação da estrutura e da construção das palavras compostas, que ocasionam a produção de sequências complexas lexicalizáveis.

Deste modo, propõe a designação de palavra composta para denominar sequências morfológica e sintacticamente construídas e etiquetas como «GN lexicalizado» para designar o tipo de categoria lexicalizada:

1. palavras compostas	Ex: timbre-poste
2. predicado lexicalizado	Ex: pousse-au-crime
3. GN lexicalizado	Ex: coffre-fort
4. SN prép lexicalizado	Ex: hors-la-loi
5. predicado complexo lexicalizado	Ex: va-et-vient
6. frase formulário lexicalizada	Ex: on-dit

Enquanto Corbin restringe a significação de palavra composta nominal, G. Gross expande a significação desta, na medida em que reformula a definição de nome composto, propondo uma designação nova. Tal reformulação deriva da influência crescente do desenvolvimento da informática aplicada aos estudos da linguística. Assim, o seu ponto de partida é «[...] *lister les noms composés sur la base de leur structure et décrire avec précision l'ensemble des paramètres qui séparent les suites figées des suites libres, et montrer par là que le figement est un phénomène scalaire*» (Gross, G., 1996:26-27).

Também G. Gross preconiza a noção de palavras construídas, que podem ser derivadas ou compostas, sendo estas últimas mais difíceis de identificar uma vez que existem palavras intermédias que podem ser uma ou outra das realidades. Quando os elementos constitutivos das palavras são nomes ou palavras com autonomia dentro da língua, impõe-se uma análise mais depurada.

G. Gross aborda a composição numa perspectiva sintáctico-semântica e considera o composto o núcleo privilegiado do grupo nominal, sendo o nome a categoria gramatical mais produtiva do ponto de vista da estabilidade

morfossintática e semântica, estabilidade essa que designa de «*figement*» e que nós traduzimos por coesão.

Também Mathieu-Colas aceita a noção de composição de G. Gross afirmando: «[...] *nous convenons de désigner, sous l'appellation de noms composés, toutes les formes nominales non soudées présentant un certain "degré" de figement, quelles qu'en soit la présentation*» (1997:72).

Uma definição estritamente formal não serve para distinguirmos uma sequência livre de uma sequência composta. No caso da sequência N + Adj temos, por exemplo a sequência *imagem digital* que segundo a terminologia de G. Gross é um composto e *imagem original* que é uma sequência livre, embora as suas estruturas sejam as mesmas de um ponto de vista meramente gramatical.

A coesão é para G. Gross a propriedade que mais fortemente caracteriza os compostos, sendo esta o resultado de um processo que une de forma durativa os constituintes de uma sequência, à qual G. Gross atribui a designação composto sempre que este constitua o núcleo de um grupo nominal. Para este autor, uma sequência adquire o estatuto de nome composto quando existem constrangimentos sintáticos e semânticos que coagem a liberdade individual de cada um dos seus constituintes. A redução parcial ou total da liberdade exercida por cada um dos constituintes que formam um composto, leva G. Gross a determinar a existência de uma gradação de estabilidade linguística que agrega os seus elementos lexicais, sendo o grau de coesão variável e medível, preconizando a coesão como «[...] *un phénomène scalaire*» (Gross, G., 1996:27).

Para a medição desse grau de estabilidade, G. Gross recorre a critérios sintáticos e semânticos: *«Une séquence est figée du point de vue syntaxique quand elle refuse toutes les possibilités combinatoires ou transformationnelles qui caractérisent habituellement une suite de ce type. Elle est figée sémantiquement quand le sens est opaque ou non compositionnel, c'est-à-dire quand il ne peut pas être réduit au sens des éléments composants [...]»* (Gross, G., 1996: 154).

Deste modo, os constituintes têm uma actuação sintáctica e semântica que nos permite distinguir uma sequência livre de uma sequência composta; esses constituintes mantêm relações entre si, no seio do grupo nominal, único grupo que nos interessa abordar.

A coesão deve ser calculada tendo em conta a quantidade de palavras que compõem a sequência, podendo a coesão do grupo nominal ser parcial ou total.

O grupo nominal é constituído por um núcleo, que no caso dos compostos é designado por *«substantif-tête»*, e pelas suas extensões à esquerda e à direita: *«On pourrait qualifier ce figement de périphérique par rapport au noyau du groupe nominal»* (Gross, G., 1996:38). São as relações privilegiadas entre os elementos periféricos e o seu núcleo que originam a coesão.

A coesão total caracteriza-se por nenhum dos seus constituintes poder ser substituído ou sofrer alterações sintáticas. Assim, G. Gross tem a opinião de que uma sequência totalmente coesa tem de ser memorizada pelos falantes e o seu tratamento linguístico requer esforços reduzidos, uma vez que *«Ils ne relèvent d'aucune analyse syntaxique interne car il n'existe ni paradigmes ni*

*transformations. Les seules modifications sont de nature formelle: [...]*» (Gross, G., 1996:40) .

O grau de coesão, dentro do grupo nominal, é analisável a partir das combinações que unem os seus constituintes, sendo possível identificar relações sintáticas e semânticas motivadoras da coesão, estando essas relações dependentes da natureza semântica do núcleo nominal.

G. Gross considera compostos os grupos nominais: « [...] *dont aucun élément n'est actualisé de façon autonome et, en particulier, dont la détermination interne ne peut faire l'objet d'une variation; qui ne constituent pas de prédication interne; entre les éléments ne peut faire l'objet d'une substitution synonymique; et dont le sens global correspond à un concept existant dans la langue et qui pourrait, à l'occasion, être exprimé par un substantif unique*» (Gross, G., 1996:42).

Com base nesta definição, G. Gross propõe uma tipologia de nomes compostos, constituída por 17 tipos, que corresponde a uma versão elementar da tipologia de Mathieu-Colas (1996), composta por 700 tipos.

Da tipologia apresentada por G. Gross (1996:48-49), só daremos conta dos tipos cujo núcleo é constituído por um nome, obtendo-se para a língua francesa a seguinte tipologia:

- |                              |                    |
|------------------------------|--------------------|
| 1. compostos adjetivo + nome | Ex: beau-frère     |
| 2. compostos nome + adjetivo | Ex: table ronde    |
| 3. compostos nome + nome     | Ex: poche revolver |
| 4. compostos nome + de + X   | Ex: pomme de terre |

- |                              |                                       |
|------------------------------|---------------------------------------|
| 5. compostos nome + à + X    | Ex: cuillère à soupe, machine à laver |
| 6. compostos nome + Prep + X | Ex: service après-vente               |
| 7. compostos nome + divers   | Ex: traction avant                    |

Para completar a tipologia, G. Gross acrescenta um conjunto de oito tipos que permitem dar conta da estabilização de algumas expansões de compostos acima mencionados:

- |                              |   |
|------------------------------|---|
| 1. Expansões de A + N        | Ex: bon vieux temps                         |
| 2. Expansões de N + A        | Ex: produit national brut                   |
| 3. Expansões de N + N        | Ex: opération ville morte                   |
| 4. Expansões de N + de + N   | Ex: offre publique d'achat                  |
| 5. Expansões de N + à + N    | Ex: film à grand spectacle                  |
| 6. Expansões de N + en + N   | Ex: requête en suspicion légitime           |
| 7. Expansões de N + Prep + N | Ex: acte sous seing privé                   |
| 8. Expansões diversas:       | Ex: non-assistance en personne en<br>danger |

G. Gross defende a necessidade da criação de tipologias de compostos para o tratamento automático da língua, por considerar não ser possível obter resultados objectivos tendo por base generalidades; a identificação dos limites da coesão é assim uma prioridade: *«En effet, pour reconnaître automatiquement les noms composés, il faut être en mesure d'en montrer les limites et de prédire leur morphologie [...] et établir au préalable une typologie*

*de la composition, afin de percevoir les problèmes spécifiques à chaque type»*  
(Gross, G., 1996, 48).

Um conjunto de parâmetros sintático-semânticos (que não vamos aqui abordar) permite-lhe testar a união dos constituintes de uma sequência, por forma a considerá-la, no seu todo, um composto.

#### **2.4. Colocação**

As escolas anglo-saxónica (Halliday (1996), Sinclair (1996), Benson (1986,1997)) e alemã (Hausmann (1989), Heid (2001)) optam, preferencialmente, pelo termo colocação para designar os conjuntos de unidades que co-ocorrem em contiguidade com uma determinada frequência no eixo sintagmático.

No quadro teórico da análise do discurso de Halliday, são os conjuntos de unidades lexicais coesas que merecem particular atenção.

Para Halliday, é a coesão que fundamenta a estabilização dos constituintes que co-ocorrem, podendo inferir-se a partir dessa co-ocorrência uma relação sintático-semântica privilegiada: *«Cohesion occurs when the interpretation of some elements in the discourse is dependent on that other. The one presupposes the other, in the sense that it cannot be effectively decoded except by recourse to it»* (Halliday, 1996:4).

Neste quadro teórico, a colocação pode ser analisada numa perspectiva gramatical ou lexical, sendo a coesão, essência da colocação enquanto

entidade linguística, identificada pelas relações estabelecidas entre as diversas unidades lexicais.

Numa perspectiva lexical, a colocação dá conta de uma forma de coesão que se atinge através da «[...] *association of lexical item that regularly co-occur*» (Halliday, 1996:284). Porém, a unidade lexical não é intrinsecamente portadora de uma relação coesiva, sendo a coesão determinada por referência ao texto em que esta unidade ocorre: «[...] *It is the occurrence of the item in the context of related lexical items that provides cohesion and gives to the passage the quality of text*» (Halliday, 1996:289). As unidades lexicais que formam uma colocação lexical são actualizadas como um todo em discurso, porque existem proximidades lexicais privilegiadas que o sistema linguístico permite. As proximidades lexicais não são todas da mesma ordem, nem da mesma natureza: «*There are degrees of proximity in the lexical system, a function of the relative probability with which one word tends to co-occur with another. Secondly, in the text there is relatedness of another kind, relative proximity in the simple sense of the distance separating one item from another, the number of words or clauses or sentences in between*» (Halliday, 1996:290). Desta forma, a densidade da coesão está relacionada com o modo como os locutores activam o sistema linguístico e, conseqüentemente, com o modo como actualizam as unidades lexicais, em função da construção de um discurso ou de um texto.

Também Sinclair aborda a colocação na sua dupla faceta gramatical e lexical, conferindo ao método estatístico um papel essencial na delimitação da colocação, para melhor poder descrevê-la: «*Collocation is the occurrence of*

*two or more words within a short space of each other in the text*» (Sinclair, 1996:170). Sinclair preocupa-se, particularmente, com a vertente lexical da colocação, numa perspectiva lexicográfica, detendo-se nas suas descrições formais.

A colocação é, assim, abordada de acordo com metodologias lexicais e estatísticas, que permitem, recorrendo ao conceito de distância, compreendida como o comprimento do segmento de recta definida por dois pontos, calcular e medir a densidade entre as unidades que constituem a colocação. As análises lexical e estatística são, neste âmbito, ferramentas complementares. A frequência com que cada uma das unidades constituintes da colocação se actualiza numa determinada ordem sintagmática é um indicador para a identificação, descrição e classificação da colocação.

Cada uma das unidades que compõem a colocação pode assumir importância e valor distintos. A unidade em observação é constituída por um nó [*«node»*] e por um co-ocorrente [*«collocate»*], podendo qualquer unidade assumir o estatuto de nó ou de co-ocorrente, consoante o valor atribuído a cada uma das unidades.

Quando, na colocação AB, o A é considerado o nó, estamos perante uma *«downward collocation»*; quando a unidade B assume o valor de nó, fala-se em *«upward collocation»*. As diferentes designações adoptadas permitem-nos inferir que a abordagem teórica a uma e a outra é distinta: *«There appears to be systematic difference between upward and downward collocation. Upward collocation, of course, is the weaker pattern in statistical terms, and the words tend to be elements of grammatical frames, or superordinates. Downward*

collocation by contrast gives us a semantic analysis of a word» (Sinclair, 1996:116).

Para Sinclair, em termos metodológicos, a noção de nó está intimamente relacionada com a noção de concordância informática:

		Infravermelho	47
T1	1a.	ível dos Registos de Satélites em Infravermelho	Térmico na Climatologia
T1	1a	imo da radiação solar ( Visível e Infravermelho	próximo , entre 0 , 4 e
T1	1b	adiação da superfície terrestre ( Infravermelho	médio ou térmico , fig .
T1	2a	s imagens das bandas do Visível e Infravermelho	térmico permite localiza
T1	13i	ral das três emulsões de um filme infravermelho	colorido . O facto de to
T1	13a	ntes do processamento de um filme infravermelho	não correspondem às core
T1	13b	s cores apresentadas por um filme infravermelho	colorido sejam generalment
T1	13d	filme colorido vulgar e um filme infravermelho	colorido Recentemente ,
T2	12f	to absorver quase inteiramente no infravermelho	, dando origem a valores
T3	8d	da terra emersa , nesta banda do infravermelho	próximo . Inicialmente o
T3	11b	criminação nos canais verde e do infravermelho	próximo ( Figs . 2 , 4 ,
T3	7c	mo , quando secos , no do próximo infravermelho	. Constata - se , ainda
T3	9b	ns captadas nas bandas do próximo infravermelho	. Esse facto deve - se à
T3	12g	reflectância na banda do próximo infravermelho	, associados a baixos va
T3	13f	em especial dos canais vermelho e infravermelho	, e a relação que se pod
T5	14h	s espectrais , Verde , Vermelho e Infravermelho	próximo . Neste caso , a
T5	17b	s canais XS2 ( vermelho ) e XS3 ( infravermelho	próximo ) mostra que é p
T6	21i	rtura aéreo - fotográfica de filme infravermelho	( falsa cor ) existente
T7	23h	banda do visível e um na banda do infravermelho	próximo . A outra imagem
T7	23i	nda do visível , dois na banda do infravermelho	próximo e um na banda do
T7	23i	vermelho próximo e um na banda do infravermelho	médio . As diferentes
T7	24e	rico , a um sinal mais elevado no infravermelho	do que nas outras bandas
T7	24f	m a 0 . 7 ( m ) . - Região do infravermelho	próximo , com compriment
T7	24f	té aos 1 . 3 ( m . - Região do infravermelho	médio , que inclui valor
T7	24g	3 ( m até 8 ( m . - Região do infravermelho	térmico ( 8 ( m - 14 ( m
T7	24c	o de carbono absorve radiações no infravermelho	térmico e no infravermel
T7	24c	ões no infravermelho térmico e no infravermelho	médio , entre 2 . 5 e 4
T7	25e	amente nula a partir da região do infravermelho	. O comportamento espe
T7	25a	tividade decresce . Em relação ao infravermelho	próximo e médio , a resp
T7	26a	re a região visível e a região do infravermelho	próximo traduz - se pelo
T7	26b	onsoante as espécies vegetais , o infravermelho	próximo revela - se uma
T7	26c	sta espectral entre o visível e o infravermelho	próximo , permite conclu
T7	26c	ividade da vegetação na região do infravermelho	médio , principalmente p
T7	26a	ções principalmente no domínio do infravermelho	médio ( em tomo dos 2 .
T7	28j	que permitem abranger a região do infravermelho	próximo ( estes filmes s
T7	28l	brangidas pelo verde , vermelho e infravermelho	próximo do espectro .
T7	32g	o do visível , quer no domínio do infravermelho	. O sistema RGV nunca fu
T8	33a	spectro ) e imagens no domínio do infravermelho	próximo ( uma situada en
T8	33d	ndo - se os 120 m para a banda do infravermelho	térmico ( banda 6 ) ;
T8	33f	abrangida , tal como a região do infravermelho	. Desta forma , existem
T8	33f	s bandas que actuam no domínio do infravermelho	médio e uma banda na reg
T8	33f	ho médio e uma banda na região do infravermelho	térmico , mantendo - se
T8	33f	mico , mantendo - se uma banda no infravermelho	próximo , tal como o MSS
T8	37a	o os níveis digitais da imagem do infravermelho	próximo e Nd <sub>i</sub> , j , iv o
T8	37b	ectral da vegetação no domínio do infravermelho	próximo ser maior e as d
T8	37b	ser maior e as diferenças entre o infravermelho	e o vermelho são deste m
T9	32e	s aéreas em filmes pancromático e infravermelho	. A classificação atinge

A concordância permite organizar em coluna a unidade que assume a função de nó, com todas as suas expansões à esquerda e à direita, isto é, os seus co-ocorrentes. Desta forma, a colocação é identificada de forma eficaz, proporcionando um levantamento exaustivo num dado *corpus*.

Também, Benson distingue as colocações lexicais das colocações gramaticais; o seu interesse pelas colocações prende-se com questões do foro lexicográfico. Na introdução ao *BBI Combinatory Dictionary of English*, Benson afirma: «*In English, as in other languages, there are many fixed, indentifiable, non idiomatic phrases and constructions. Such groups of words are called recurrent combinations, fixed combinations, or collocations. Collocations fall into two major groups: grammatical collocations and lexical collocations*» (Benson e alii, 1997:XV).

A utilização do termo colocação é empregue com uma dupla acepção. Por um lado, considera-o sinónimo de combinatórias fixas e recorrentes; por outro lado, considera a colocação um subtipo de combinação lexical claramente identificado, distinguindo-a dos outros tipos.

Benson parte de uma tipologia, elaborada com base em possíveis combinações lexicais, organizado em quatro grupos, representando as colocações um subtipo:

1. combinações livres;
2. expressões idiomáticas;
3. colocações;
4. composto.

As combinações livres lexicais são aquelas estruturas que ocorrem com mais frequência em discurso, uma vez que apresentam um grau reduzido de coesão, por as suas hipóteses combinatórias serem livres: «*Free combinations,*

[...], consist of elements that are joined in accordance with the general rules of English syntax and freely allow substitution» (Benson et alii, 1997:XV).

O facto de determinadas combinações lexicais ocorrerem com frequência num *corpus* não nos autoriza a afirmar, desde logo, estarmos perante colocações. Com efeito, tais combinações lexicais discursivas ou livres, que se distinguem das colocações, unidades de língua apresentando restrições lexicais que permitem a sua identificação e conseqüente delimitação. Só estas últimas merecem um tratamento lexicográfico, tendo, nesta óptica, a frequência um interesse relativo.

As expressões idiomáticas que ocorrem com menor frequência em discurso são definidas por Benson como: «[...] *relatively frozen expressions whose meanings do not reflect the meaning of their component parts*» (Benson, 1985:4).

A meio caminho entre a combinação livre e a expressão idiomática está a colocação que é definida como: «[...] *semi-fixed combinations*» (Benson, 1997:IX). Segundo este autor, a particularidade da colocação encontra-se no facto de a totalidade do seu significado ser o resultado do significado individual dos seus constituintes, o que permite distingui-la da sequência livre.

Para além do mais, não é comum encontrarem-se sinónimos dos significantes e dos conceitos veiculados pelas colocações, o que as torna particulares.

Dentro desta categoria, Benson distingue as colocações gramaticais das colocações lexicais. A colocação gramatical «[...] *is a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure*

*such as an infinitive or clause*» (Benson, 1997:IX), enquanto a colocação lexical, em contraste com a colocação gramatical, se define por: «[...] *normally do not contain prepositions, infinitives, or clauses. Typical lexical collocations consist of nouns, adjectives, verbs and adverbs*» (Benson, 1997:XXX).

Os compostos, a quarta categoria identificada, é definida por Benson como «*completely frozen*» (Benson, 1985:4).

Para identificar as colocações de forma sistemática numa perspectiva lexicográfica, Benson subdivide as colocações lexicais constituídas por [V + N] em dois tipos, que designa de colocações CA e de colocações EN, respectivamente. Estes tipos de colocações são reconhecidas como tal, a partir de uma tipologia de funções lexicais elaboradas por Benson, tendo por base a tipologia de funções lexicais de Mel'čuk (1984).

As colocações CA exprimem a função «criação» («*creation*») (C) e a função «ativação» («*activation*») (A); as colocações EN são estruturas combinatórias que traduzem a função «irradiação» («*eradication*») (E) e a função «neutralização» («*nullification*») (N).

Além destas duas combinações lexicais que apresentam uma estrutura idêntica, mas um formalismo distinto, Benson identifica mais cinco tipos de estruturas combinatórias, que dão conta de relações paradigmáticas e sintagmáticas que se estabelecem entre os lexemas, cuja delimitação se fundamenta nas diferenças que permitem descrever as ligações entre os vários lexemas que constituem a colocação: [Adj+N]; [N+V]; [N+N]; [Adv+Adj] e [Adv+V].

Nestes conjuntos apresentados, Benson imputa a um dos lexemas o papel de determinante e a outro o de determinado, concedendo aos lexemas que intervêm na colocação o estatuto de um ou de outro, de acordo com a função que lhe é atribuída: «*Lexical collocations [...] do not have a dominant word, [...]*» (Benson et alii, 1997:IX)

Em 1989, Hausmann define a colocação como: «[...] *la combinaison caractéristique de deux mots dans une des structures suivantes: a) substantif + adjectif (épithète); b) substantif + verbe; c) verbe + substantif; d) verbe + adverbe; e) adjectif + adverbe; f) substantif + (prép.) + substantif*» (Hausmann, 1989:1010). Para Hausmann, a colocação define-se por oposição a combinação livre e a expressão idiomática, pela restrição combinatoria das unidades que compõem a colocação, pela sua transparência e pela sua não coesão, sendo apreendida e utilizada como uma unidade de língua e não como uma unidade de «*parole*» na acepção saussuriana. O facto de a colocação poder assumir o estatuto de unidade lexicográfica, implica, num determinado momento, a passagem do discurso para a língua.

Hausmann defende tal afirmação pelo facto de considerar que um estrangeiro pode perceber o significado da colocação como um todo, acto que resulta da sua transparência, mas no acto da «*parole*» não ser capaz de utilizar as combinações lexicais adequadas, o que comprova a imprevisibilidade da colocação. Ele precisa aprender e memorizar as combinações lexicais para poder usá-las, do mesmo modo que aprende e memoriza qualquer outro signo linguístico: « [...] *parce que les langues, dans la totalité des combinaisons logiquement possibles, font un choix idiosyncratique*» (Hausmann, 1989:1010).

Para este autor, a descodificação de uma colocação não traz qualquer tipo de contrariedade. A dificuldade situa-se ao nível da codificação: «[...] *l'étranger ne peut pas prévoir à l'aide de quel verbe la langue française «met fin» au silence: \*briser, \*casser, rompre le silence»* (Hausmann, 1997:282).

Hausmann considera que as colocações são identificáveis através dos elos que unem cada uma das suas unidades constituintes, o que designa por «*Wortverbindungen*». Assim, distingue as «*fixierte Wortverbindungen*» das «*nicht fixierten Wortverbindungen*». Nas unidades fixas, ele inclui expressões como «*se laver la tête*», assim como as tradicionais palavras compostas, «*chambre forte*», sendo as unidades fixas apreendidas como um único signo linguístico, enquanto as outras estruturas são apreendidas como dois signos linguísticos distintos.

Independentemente de se tratarem de unidades fixas ou não, Hausmann identifica três tipos de combinações lexicais: a co-criação («*Ko-Kreationen*»), a colocação («*Kollokationen*») e a contra-criação («*Konter-Kreation*»).

O primeiro tipo de combinação lexical é a co-criação («*Ko-Kreationen*»), que Bahns interpreta do seguinte modo: «*Als Ko-Kreationen bezeichnet Hausmann Verbindungen von Wörtern, deren Kombinationsfähigkeit praktisch unbegrenzt ist; bei ihrer Verbindungen sind lediglich gewisse semantische Mindestregeln zu beachten. [...] Ko-Kreationen werden entsprechend den Regeln des Sprachsystems kreativ zusammen gestellt»* (Bahns, 1996:23).

As colocações («*Kollokationen*») resultam de relações de palavras que possuem uma capacidade combinatória limitada. Por oposição às co-criações,

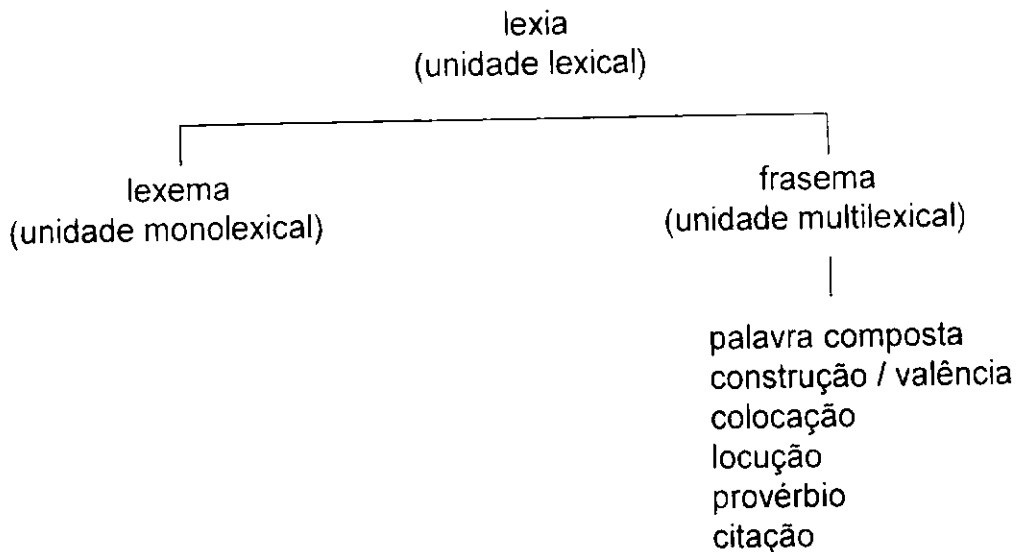
as colocações não são construídas de forma criativa. «[...] *sondern werden als Kombinationen aus dem Gedächtnis abgerufen*» (Bahns, 1996:24).

Por último, temos as contra-criações («*Konter-Kreation*»): «*Bei Konter-Kreationen handelt es sich um die Verbindungen zwischen einem begrenzt kombinierbaren Wort und einem Wort, mit dem es normalerweise nicht gemeinsam aufzutreten pflegt*» (Bahns, 1996:24). Normalmente, nestes casos as regras semânticas são violadas, requerendo por parte do falante uma capacidade criativa superior à que é activada no acto da co-criação.

Para Hausmann, uma colocação é uma combinação orientada, isto é, as unidades que a constituem não têm o mesmo estatuto, sendo uma delas o pólo responsável pelas relações lexicais privilegiadas que mantém com a sua vizinhança local. Por esse motivo, é essencial distinguir a base do colocador («*Kollokator*»), equivalente às noções de nó e de co-ocorrente, uma vez que a sua identificação é indispensável para a descrição linguística e o tratamento lexicográfico da colocação, assim como para a sua aprendizagem e consequente assimiliação. Estes dois elementos que constituem a colocação têm estatutos de parceria diferentes, sendo também diferente o tratamento a que são sujeitos consoante a ênfase seja dada à base ou ao colocador.

A colocação é, assim, constituída por uma base com autonomia sintáctica e semântica e por um colocador, que acrescenta uma característica à base, não modificando a sua identidade.

Em 1997, Hausmann (1997:288) apresenta uma organização hierárquica das denominações das estruturas lexicais que, em parte, reflecte a teoria de Mel'čuk, sendo, no entanto, a sua classificação mais pormenorizada:



Adaptado de Hausmann (1997), «Tout est idiomatique dans les langues», *La locution entre langue et usages*, Saint-Cloud. ENS Éditions, p. 288.

Heid segue, parcialmente, a linha de pensamento de Hausmann, destacando-se deste pelo facto de a sua abordagem à colocação ser terminológica. Heid defende, explicitamente, que a colocação pode corresponder a um termo com características e propriedades distintas das dos termos identificados, tradicionalmente, como nomes compostos.

Heid (2001) também se refere à polaridade da colocação. A colocação é constituída por dois lexemas e por possíveis determinantes, quantificadores e preposições, sendo um dos lexemas o determinado e o outro o determinante, correspondendo estas noções ao «*node*» e ao «*collocate*» de Sinclair e à «*Basis*» e ao «*Kollokator*» de Hausmann.

Sendo a sua abordagem à colocação terminológica, um dos lexemas tem de ser, obrigatoriamente, um termo, podendo ambos os lexemas assumir esse estatuto.

Do ponto de vista linguístico, as colocações são « [...] a *phenomenon of lexical combinatories: they involve the lexical, semantic, and syntactic*

*properties of lexical items and their syntagmatic co-occurrence*» (Heid, 2001:788), que tal como os signos linguísticos, resultam de uma convenção, «*rather than being explicitly rule-governed*» (Heid, 2001:788).

Os elementos que constituem uma colocação, podem, já por si, ser colocações que, recorrendo às mesmas estruturas, permitem a criação de novas colocações. Heid faz um levantamento das estruturas constituintes mais relevantes das colocações:

1. [N+V]
2. [N+Adj]
3. [N+N]

Estas três estruturas são as mais frequentes nas línguas de especialidade e são aquelas que, segundo Heid, constituem, regra geral, a estrutura das unidades terminológicas. Além destas, ocorrem outro tipo de estruturas, menos frequentes, contudo também representativas:

1. [V+Adv]
2. [Adj +Adv]

A identificação das propriedades linguísticas das colocações é essencial para o tratamento automático das línguas naturais. Por esse motivo, Heid considera que as colocações são delimitáveis e descritas através das suas propriedades sintáticas, semânticas, mas também conceptuais.

Quando se refere às propriedades sintáticas, Heid estabelece uma relação entre a colocação e a palavra composta, uma vez que considera que a escolha das componentes dos compostos está determinada do ponto de vista colocacional: «*The choice of the components in such noun groups, like the choice of the components of the compounds, is often collocationally determined: there are clear combinatory preferences, often merely conventional, that in many cases go as far as the complete terminological "fixing" of the compounds and noun groups*» (Heid, 2001:791).

Heid assume a dificuldade que sente em distinguir a colocação da composição com base em critérios meramente linguísticos, sabendo que a fronteira é, do ponto de vista teórico, muito ténue. Numa perspectiva terminológica, tal distinção não se revela muito operacional, uma vez que em terminologia é a denominação que está na base da identificação da realidade linguística, independentemente, da etiqueta que lhe é atribuída: «*From a terminological point of view, we may be more interested in whether the combination of term and collocate can be seen as the domination of a new concept in its own right*» (Heid, 2001:791). Numa perspectiva meramente terminológica, não podemos deixar de concordar com o seu posicionamento, independentemente, de em função da estrutura linguística, determinada unidade ser considerada uma colocação ou um composto, o que nos motiva verdadeiramente na abordagem terminológica é a denominação.

Contudo, na qualidade de linguistas, e numa perspectiva de tratamento automático da língua natural, preocupa-nos encontrar os critérios formais para delimitar as fronteiras da expressão linguística da denominação que ocorre em

texto. São as teorias linguísticas que permitem a criação de formalismos conducentes ao tratamento automático das línguas naturais.

Por isso, pensamos não ser produtivo confundir o plano linguístico com o plano extralinguístico, sendo todavia essencial acautelar as suas complementaridades.

No que diz respeito às propriedades semânticas, Heid inspira-se na teoria de Mel'čuk. Mel'čuk defende que os falantes, no pleno uso da «*parole*», recorrem às colocações para exprimir a significação genérica, descrevendo deste modo as colocações a partir de funções lexicais que lhe permitem dar conta dessa genericidade e que Heid exprime do seguinte modo: «*In lexicography, examples of collocations are usually treated in terms of a given collocate with a given base being arbitrary phenomenon that must be memorized*» (Heid, 2001:793).

Em língua de especialidade, tal fenómeno também ocorre, com a diferença que a escolha do co-ocorrente é, regra geral, o resultado de uma convenção e não de um livre arbítrio.

Desde modo, Heid (2001) e Martin (1992) distinguem as colocações lexicais das colocações conceptuais, que Martin define como se segue: «*By conceptual structure we can mean the semantic valency or argument structure of a concept (or the conceptual meaning of a lexeme). This way modifying concepts in "collocations" [...] are conditioned or expected from "definitional knowledge". In other words: the conceptual meaning of disease contains a slot for CAUSE, that of system contains one for FUNCTION, that of vowel one for articulation, that*

*the dictionary one for INFORMATION, etc. In this respect the above 'collocations' are concept bound» (Martin, 1992).*

Segundo estes autores, as diferenças entre as colocações lexicais e as colocações conceptuais prendem-se com o facto de as primeiras serem determinadas por questões meramente lexicais e idiomáticas, que ocorrem com mais frequência em língua corrente.

Por seu turno, as colocações conceptuais são mais relevantes nas línguas de especialidade e podem assumir o estatuto de termos: «*Terminologically relevant collocations are more often than not situated at the content-related end of the spectrum, with the meaning of the collocates fairly often derivable from definitional elements of the base, its properties, typical actions usages, etc.*» (Heid, 2001, 796). Em terminologia, a relação entre o conceito e as unidades que constituem a colocação justificam a coesão lexical, uma vez que a base subcategoriza determinadas unidades terminológicas constituintes das colocações, sendo a colocação um reflexo da conceptualização.

A colocação conceptual é um conceito que poderá permitir a distinção teórica entre «colocação livre» e «colocação restrita» numa perspectiva puramente terminológica. No entanto, rejeitamos tal denominação, uma vez que confunde os dois planos de análise que nos são muito caros: o linguístico que se reflecte na colocação e o extralinguístico que está presente no conceito de conceptual.

Do ponto de vista da sua estrutura, a colocação organiza-se segundo as regras da língua, enquanto que o conceptual se rege por critérios extralinguísticos que, em língua, é denominado por signos linguísticos.

O facto de uma colocação denominar um conceito, não nos impele a considerá-la conceptual. Preferimos considerá-la uma colocação terminológica, estando, deste modo, a relação entre o termo e conceito ressalvada.

Tal como a nossa exposição deixa antever, a colocação não é definida, nem entendida do mesmo modo por todos os autores, nem a realidade linguística para a qual remete é consensual entre os especialistas.

Todos os autores citados abordam a colocação numa perspectiva lexical, sendo, no entanto, as áreas de investigação distintas: textual para Halliday, lexicográfica para Sinclair, Benson e Hausmann, terminológica para Heid.

Para uns, a coesão é o que justifica a existência da colocação (Halliday), para outros, é a combinação restrita dos lexemas que a compõem, sendo a coesão uma característica da locução (Hausmann).

Quanto à designação e ao valor que assumem cada uma das unidades que constituem as colocações, também aí os autores usam terminologias distintas.

Por outro lado, as tipologias das entidades linguísticas não são consentâneas umas relativamente às outras. Benson apresenta a sua tipologia para a língua inglesa, Hausmann e Heid para a língua inglesa, francesa e alemã:

Benson	Hausmann	Heid
V + N	V + N	
ADJ + N		
N + V	N + V	N + V
N + N		N + N
ADV + ADJ		
ADV + V		
	N + ADJ	N + ADJ
	V + ADV	V + ADV
	N + PREP + N	
		ADJ + ADJ

Em síntese, concluímos que as colocações são constituídas por, exclusivamente, dois lexemas. Tendo em conta as tipologias apresentadas pelos autores, constatamos, com alguma perplexidade, que só a estrutura N+V é consensual. No que diz respeito às estruturas V+N; N+N; N+Adj e V+Adv, só dois autores em três estão em acordo. Hausmann apresenta ainda a estrutura N+Prep+N, Benson as estruturas Adj+N e Adv+Adj e Heid apresenta a estrutura Adj+Adj.

De todas as definições de colocações apresentadas e respectivas matrizes, só Heid faz referência à existência de gramemas autónomos no seio das colocações: «*Collocations involve two lexemes (plus possibly determiners, quantifiers, and prepositions)*» (Heid, 2001:788).

Curiosamente, nem todos os exemplos que Hausmann apresenta correspondem fielmente à sua tipologia, uma vez que neles inclui gramemas autónomos, preposições ou determinantes, que não estão representados nas suas matrizes.

A título de exemplo, para a estrutura V + Adv apresenta o exemplo em língua francesa *il pleut à verse*, onde ocorrem gramemas que não estão representados. No entanto, apresenta a matriz N+Prep+N, estrutura em que a preposição é claramente considerada uma classe gramatical de relevo, não justificando do ponto de vista teórico a sua inclusão na matriz.

Desta diversidade terminológica e conceptual, retemos os seguintes critérios para a definição de colocação:

1. ser constituída por dois lexemas;
2. poder conter determinantes, preposições ou quantificadores;

3. ser o resultado de combinações lexicais;
4. ser apreendida pelo falante como um único signo linguístico;
5. no caso da perspectiva terminológica, ser o resultado da conceptualização de um objecto.

## 2.5. Frasemas

Como já tivemos oportunidade de referir no ponto 1. deste capítulo, Mel'čuk considera a lexia o elemento de base de qualquer estudo lexicográfico, independentemente da sua forma linguística. O Dicionário Explicativo e Combinatório (DEC) é um dicionário de lexias, que pretende apresentar uma teoria lexicográfica, baseando-se numa descrição semântica, combinatória e explicativa das lexias de uma língua natural: *«Un DEC se doit de décrire non seulement les lexèmes de la langue L mais encore tous ses phrasèmes (dont le nombre est nettement supérieur à celui des lexèmes), en devenant de ce fait un dictionnaire de MOTS et de PHRASÈMES, donc un dictionnaires de lexies»* (Mel'čuk, 1995:45).

Mel'čuk refere-se à co-ocorrência lexical para dar conta da capacidade que os lexemas têm de se combinarem em sintagmas, por forma a exprimirem uma dada significação. Esta é uma propriedade inerente a todos os lexemas.

No entanto, esta co-ocorrência lexical é considerada restrita quando *«[...] un lexème A signifiant 'A' et un lexème B signifiant 'B' ne peuvent pas se combiner pour exprimer le composé 'A + B', cela n'étant pas interdit par la syntaxe»* (Mel'čuk, 1984:4). Podemos, deste modo, concluir que sempre que

A+B ≠ AB estamos perante uma co-ocorrência lexical restrita, que Mel'čuk denomina de frasema, caracterizada por, estruturalmente, ser uma unidade multilexémica, que detém propriedades que não são dedutíveis a partir da interpretação dos seus constituintes.

Para Mel'čuk, o frasema é um signo linguístico: «*A linguistic sign is an ordered triple:  $X = \langle 'X'; /X/; \Sigma_X \rangle$ , where 'X' is the signified of the sign X (= its meaning), /X/ is its signifier (=its phonetic form), and  $\Sigma_X$ , its syntactics (= the set of data on its co-occurrence with other signs)*» (Mel'čuk, 1998:26).

Especifica três tipos distintos de frasemas: os frasemas completos, os semi-frasemas e os quase-frasemas (cf. Mel'čuk, 1995: 45-48).

Os frasemas completos correspondem às estruturas sintagmáticas compostas pelos lexemas **AB**, cuja significação é igual a **C**, em que o **C** não contém nem o significado de A, nem de B: **C' ⊃ 'A' & 'C' ⊄ 'B'**

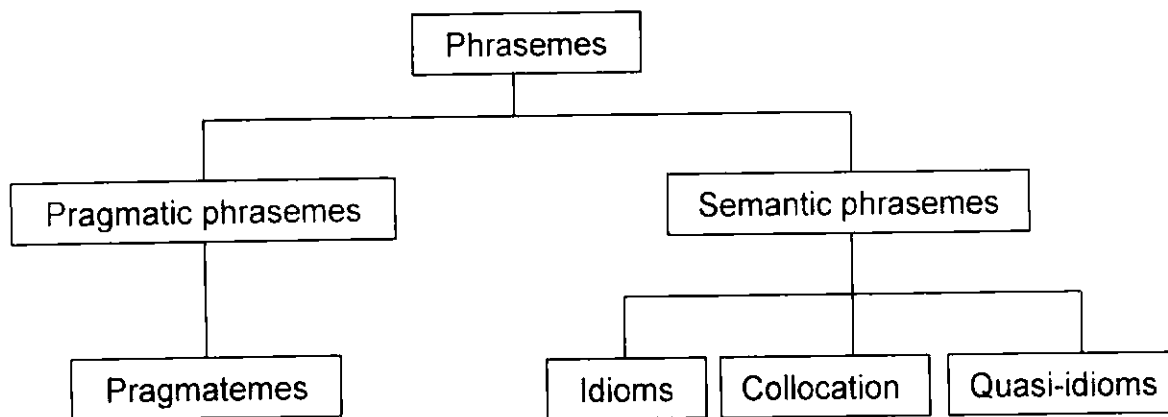
Os frasemas são um conceito equivalente às expressões idiomáticas definidas pela escola anglo-saxónica.

São considerados semi-frasemas todas as co-ocorrências restritas, constituídas pelos lexemas **AB**, cuja significação é igual a '**AC**' ou a '**BC**': «[...] *qui inclut le signifié de l'un des constituants, alors que l'autre ne garde pas son sens, soit - même s'il garde son sens - n'est pas sélectionné librement*» (Mel'čuk, 1995:46). Para este autor, semi-frasema é sinónimo de colocação.

Por fim, apresenta os quase-frasemas, que se compõem dos lexemas **AB**, cuja significação corresponde a **ABC**. Assim, o frasema comporta o significado dos respectivas lexemas **A** e **B**, sendo-lhe ainda acrescentado um outro significado não dedutível e imprevisível, o significado **C**.

Nas definições propostas dos diferentes tipos de frasemas, Mel'čuk não identifica as categorias gramaticais, nem as funções sintácticas dos lexemas que os constituem. Deste modo, não restringe as hipóteses de classificação dos lexemas, recorrendo em sua substituição, a expressões semânticas que lhe permitem obter uma tipologia, cuja finalidade é a descrição exhaustiva das lexias: «[...] *une description d'une langue L est un système de règles qui établit la correspondance entre le sens et les textes de L. Le DEC de L doit fournir au système toutes les données sur les mots individuels, données indispensables à son fonctionnement*» (Mel'čuk, 1984:3).

Mel'čuk representa, esquematicamente, a classificação geral dos frasemas do seguinte modo (cf. Mel'čuk, 1998:30).



Mel'čuk (1998), «Collocations and Lexical Functions», *Phraseology. Theory, analysis, and applications*, ed. A. P. Cowie, Oxford, Oxford University Press, p.30

No âmbito desta dissertação e na sequência das leituras efectuadas, optámos por considerar frasemas ou unidades multilexémicas todas as unidades terminológicas nominais que são constituídas por, no mínimo, dois lexemas ou unidades monolexémicas separados por um espaço em branco,

resultando das suas combinatórias morfolexémicas e sintácticas numa denominação.

Enquanto frasema, a denominação assume o estatuto de unidade lexicográfica e, em contexto terminológico textual, desempenha a função de uma unidade monolexémica ou lexema.

### 3. Fraseologia

Ao reflectirmos sobre o conceito de fraseologia, verificámos que grande parte dos autores consultados (Kjær (1990), Picht (1990), Rousseau (1993), Roberts (1993), Rey (1997)), distinguem a fraseologia que ocorre em língua corrente, e que é abordada de um ponto de vista lexicográfico, da fraseologia que ocorre em língua de especialidade, e que é objecto de estudo da Terminologia.

Kjær aborda a fraseologia na sua dupla vertente teórica e prática. Considera a fraseologia como uma teoria no seio da Linguística e, simultaneamente, como um inventário de palavras combinadas. Do ponto de vista da teoria, a fraseologia relaciona-se com: «[...] *syntagmatic combinations of words - or, to be more precise, of lexical units. In this sense phraseology can be paraphrased as the theory of phraseology*» (Kjær, 1990:4).

Do ponto de vista prático, a fraseologia, enquanto inventário, diz respeito a «*phraseological word combinations*» no seio de uma língua de especialidade, o que leva esta autora a estabelecer um paralelismo com a terminologia,

afirmando: «*In this sense, phraseology is comparable to terminology as denoting the inventory of terms of a sublanguage*» (Kjær, 1990:4).

Como referimos oportunamente (cf. Capítulo 1), não entendemos a língua de especialidade como uma sublíngua, mas sim como a língua utilizada em especialidade. O que nos interessa aqui debater é, no entanto, o conceito de fraseologia enquanto objecto de estudo da Terminologia.

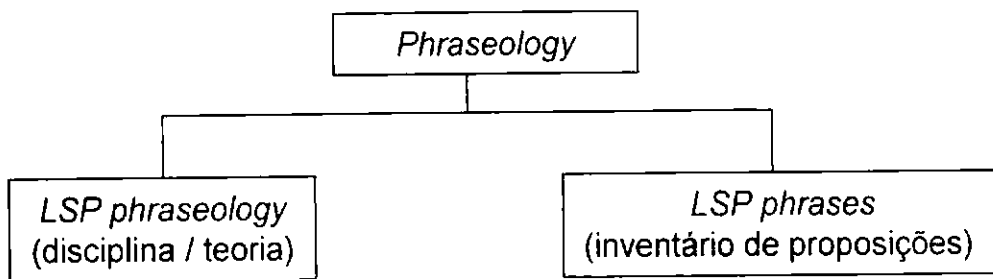
De forma pouco clara, Kjær tenta distinguir as combinações lexicais, objecto de estudo da Lexicologia, das combinações lexicais terminológicas, objecto de estudo da Terminologia.

Numa perspectiva terminológica, a fraseologia é entendida como «[...] *the object of linguistic study [that] denotes the environment of the terms*» (Kjær, 1990:4), o conjunto das unidades lexicais combinadas, designadas por «*terminological phrases*», em que o termo é o núcleo ou o pólo da fraseologia.

Num perspectiva lexicológica, a fraseologia é entendida numa dupla acepção: «*Phraseology denotes the branch of study concerned with the analysis of word combinations with make up lexical units of which have the character of fixed expressions for some reason other than their lexicalization*» (Kjær, 1990:4). Desta afirmação inferimos que, em Lexicologia, a fraseologia designa, por um lado, uma disciplina, por outro lado, as fraseologias que Kjær opõe às unidades lexicalizadas, designadas de «*multi-word units*» ou de «*fixed multi-word expressions*». As fraseologias desfrutam de características inerentes às expressões fixas, não ficando esclarecido quais os processos ou causas que determinam a sua coesão.

Kjær recorre a múltiplas denominações para dar conta de fraseologias em Lexicologia - frasemas, multi-palavras, expressões multi-palavras fixas -, que insere num grande grupo: «*phraseological word combinations*», da mesma forma que Hausmann (1997) e Mel'čuk (1995) consideram frasema um générico para colocações, palavras compostas ou locuções.

Picht (1990) aborda a fraseologia, numa perspectiva terminológica, de uma forma mais sistematizada. Também estabelece a distinção entre a teoria e o inventário de entidades linguísticas combinadas, sendo a «*LSP phraseology*» a disciplina que se debruça sobre as «*LSP phrases*»:



Para Picht, a «*LSP phraseology*» corresponde a uma disciplina, que se ocupa do estudo das línguas de especialidade, e que opera com os seguintes aspectos: «[...] *with deals on the on hand, with syntactic links which appear in LSP as expression level, their synonymy and equivalence, and on the other hand, with the conceptual relationships and their modifications of those LSP elements which may be combined so as to create a proposition which is both valid according to the percepts of the special field, and linguistically correct*» (Picht, 1990:43).

O objecto de estudo da «LSP phraseology» é, para Picht, a «LSP phrase», que define como «[...] *the product of a syntactic linkage between at least two elements in a proposition with an LSP content, whose inner coherence depends upon conceptual connectability*» (Picht, 1990:43).

Sendo assim, a «LSP phrase» deve ser analisada tendo em conta o nível do conceito e o nível da expressão que, conjuntamente, formam uma proposição.

No plano conceptual, Picht refere as propriedades combinatórias dos conceitos que se definem por possuírem características inerentes aos objectos: «[...] *the first type of relational characteristics are instrumental for the ordering of concepts, the second for the creation of a proposition with an LSP content*» (Picht, 1990:37).

Em segundo lugar, identifica as propriedades dos conceitos com características de verbos, que se diferenciam das características dos objectos por darem conta de uma acção: «[...] *all concepts which are expressed by means of a verb possess at least one characteristic which distinguishes them from the first category and which is essential for the creation of a proposition: the notion of DOING - the essential action*» (Picht, 1990:37). De facto, nas línguas de especialidade há determinados verbos que são claramente de especialidade, devendo ser tratados como termos. No caso da Detecção Remota, verbos como *digitalizar* e *colorir* são recorrentes, porque conceptualmente imperiosos, merecendo um tratamento terminológico ao mesmo nível dos substantivos.

No que concerne a abordagem conceptual da «LSP phrase», há que ter em consideração a existência de uma conexidade entre, pelo menos, dois conceitos no seio de uma proposição: «[...] *where both concepts exhibit relational characteristics which render capable of being joined together so as to form a proposition which fulfills the special requirements to the subject fields*» (Picht, 1990:37-38).

Pensamos, ser este um dos critérios mais relevantes para distinguir fraseologia de todos os outros conceitos apresentados nos pontos anteriores, e que iremos ter em especial consideração, no desenvolvimento desta dissertação.

Ao nível da expressão, a proposição é a manifestação de um relação linguística entre uma expressão nominal e uma expressão verbal: «*At this level a proposition manifests itself as the linguistic linkage of one term which is nominal in character, be it a simple or a compound term, an abbreviation, etc., with another, verbal in character and expressed by means of a verb or a verbal group, [...]*». (Picht, 1990:38).

Segundo Picht, para estarmos perante de uma fraseologia é necessário a congregação de dois conceitos numa única proposição, sendo um dos conceitos expresso por um verbo ou grupo verbal e o outro por um nome ou um grupo nominal. Os exemplos que apresenta são os seguintes: *to forge iron; elevar un bomba; der Strom flie }t; to retrieve data*.

Mais uma vez, nem todas as definições de fraseologia são concensuais e autores como Greciano (1993, 1997) referem-se ainda a fraseotermos e a fraseolexemas para designar termos, lexemas, frasemas, fraseologias, entre

outros. Mas não vamos desenvolver todas estas perspectivas, porque não se evidenciam produtivas, nem em termos teóricos, nem em termos metodológicos.

Parece-nos que a proliferação de terminologias, reflexo de uma profusão de teorias, vem dificultar o estabelecimento de critérios fundamentais para a identificação das realidades linguísticas estruturalmente complexas.

Só iremos deter-nos nas denominações, isto é, nas unidades terminológicas multilexémicas e nas suas combinatórias a fim de entendermos e descrevermos o seu comportamento num contexto textual terminológico, com o intuito de procedermos ao seu tratamento automático.

A extensão sintagmática das unidades terminológicas multilexémicas ou frasemas terminológicos são imprevisíveis, sendo as fronteiras entre elas e a fraseologia fluidas. Cowie (1998:5) designa as primeiras, genericamente, por «*Word-like (or semantic) unit*» e as segundas por «*sentence-like (or pragmatic) unit*».

Assim, entendemos a fraseologia como um conceito operatório que, no âmbito desta dissertação, serve para designar frasemas que possuem estruturas próximas da frase, entendida, neste contexto, como: «*Elemento teórico abstracto, no interior do qual funcionam outros elementos menores, [...]*» (Xavier, Mateus, 1990:185).

Tendo por base este pressuposto teórico, podemos definir a fraseologia como uma parte de frase, ou proposição, que se caracteriza por ser constituída por várias denominações ligadas entre si por elementos gramaticais que, regra geral, são extrínsecos às unidades terminológicas multilexémicas.

As denominações por sua vez são constituídas por unidades terminológicas monolexémicas e unidades terminológicas multilexémicas.

Pensamos, assim, que a denominação denomina uma realidade extralinguística, que pode ser tratada como unidade lexicográfica, enquanto a fraseologia é constituída por várias denominações, ligadas entre si por elementos característicos de frases.

Utilizaremos ainda o termo colocação para dar conta de sequências de lexemas terminológicos que co-ocorrem, frequentemente, mas que não designam conceitos, não podendo desta feita ser confundidos com unidades terminológicas multilexémicas.

Numa perspectiva de terminologia textual e de tratamento automático da língua natural, a fraseologia deve ser identificada como um todo, uma vez que as denominações surgem em discurso de forma encadeada, por vezes de forma ocasional, outra vezes de forma sistemática, podendo o seu conjunto ter interesses diversos para a tradução automática, para a constituição de bases de dados terminológicas, para tradutores humanos e para o tratamento da informação.

## Capítulo IV

---

### **Etiquetador gramatical automático: EtiqueLex**

1. Tratamento do corpus informatizado
2. Tabela de etiquetas
3. Constituição dos dicionários
  - 3.1. Dicionário de formas simples flexionadas
    - 3.1.1. Desdobramento de formas
    - 3.1.2. Variantes ortográficas
  - 3.2. Dicionário de locuções
  - 3.3. Dicionário de unidades terminológicas multilexémicas da Detecção Remota
4. Etiquetador gramatical: EtiqueLex
5. Tipologia de matrizes terminogénicas
  - 5.1. Tipologia elementar
  - 5.2. Análise da composição das unidades terminológicas multilexémicas
  - 5.3. Tipologia de base
  - 5.4. Dicionário de estruturas tipo

## Capítulo IV - Etiquetador gramatical automático: EtiqueLex

### 1. Tratamento do *corpus* informatizado

Para tirar o máximo proveito do *corpus* informatizado, é necessário procedermos à sua preparação, por forma a podermos explorar todas as suas potencialidades.

O objectivo desta dissertação é o tratamento automático dos signos linguísticos que compõem o texto. Mas o texto de especialidade caracteriza-se por ser constituído por outros símbolos, que não os signos linguísticos, tais como imagens, fórmulas, gráficos, mapas, esquemas, entre outros.

Assim, a primeira tarefa é a de subtrair ao texto todo e qualquer elemento que não seja considerado língua natural, para dar início à pré-fase da automatização, que se traduz na constituição de dicionários.

Os dicionários foram elaborados a partir do *corpus*. Para o efeito, utilizámos o *Hyperbase*, um programa de tratamento documental e estatístico que foi concebido para o tratamento de grandes *corpora*.

Com este programa obtivemos todas as ocorrências do *corpus*, isto é, 235.893 ocorrências para 12888 formas, que são apresentadas por ordem alfabética.

Apesar das sucessivas correcções pelas quais passou o *corpus*, ainda permaneceram elementos parasitários, que só se tornaram visíveis no dicionário. Da totalidade das formas constavam a pontuação, os ordinais, a

indicação de alíneas, as citações em língua estrangeira, os erros ortográficos, as gralhas, os nomes de autores citados, etc.

Muitas destas formas foram, posteriormente, suprimidas para que só subsistissem vocábulos, tendo sido, por esta via, eliminadas cerca de 1500 formas.

Cada uma das formas é acompanhada da respectiva frequência, isto é, do número de vezes que cada uma dessas formas ocorre no *corpus*:

Formas	FQ. Total	FQ. T1	FQ. T2	FQ. T3	FQ. T4	FQ. T5	FQ. T6
a	6909	3815	80	601	945	748	720
á	25	13	0	5	1	2	4
à	931	509	9	76	83	90	164
abaixo	5	2	0	0	2	0	1
abandonado	1	1	0	0	0	0	0
abandonadas	3	3	0	0	0	0	0
abandono	2	1	0	0	0	0	1
abarcando	1	1	0	0	0	0	0
abarcar	1	1	0	0	0	0	0
abastecer	1	1	0	0	0	0	0
abcissas	1	1	0	0	0	0	0
aberrantes	1	1	0	0	0	0	0
aberta	1	1	0	0	0	0	0
aberto	5	5	0	0	0	0	0
abertos	3	3	0	0	0	0	0
abertura	39	35	0	1	0	1	2
aberturas	5	5	0	0	0	0	0
aborda	1	0	0	0	1	0	0
abordada	1	0	0	0	0	0	1
abordados	5	2	0	0	0	2	1
abordagem	17	9	0	0	2	2	4

A primeira coluna contém as formas, a segunda a frequência total com que uma forma ocorre na totalidade do *corpus*, ao passo que as restantes

seis colunas indicam a frequência com que cada forma ocorre em cada sub-*corpus*.

O dicionário das formas é, assim, constituído, numa primeira instância, por todas as formas flexionadas que compõem o *corpus*.

Para que este dicionário cumpra a sua função, é indispensável atribuir etiquetas a cada uma das formas. É o dicionário de formas flexionadas etiquetadas que está na base do EtiqueLex.

## **2. Tabela de etiquetas**

Atribuir uma etiqueta aos diversos funções gramaticais que uma forma pode assumir, implica uma reflexão sobre o conjunto de etiquetas que deve ser criado, em consonância com os objectivos que nos propomos atingir.

Assim, a quantidade e o grau de explicitação da etiqueta devem adequar-se aos objectivos do tratamento dos dados linguísticos. Não é útil sobrecarregar o dicionário de informação metalinguística e, conseqüentemente, o *corpus*, se ela não tiver uma funcionalidade comprovada.

Por este motivo, ao longo de todo o processo de etiquetagem do dicionário e do *corpus*, fomos adaptando a bateria de etiquetas às nossas necessidades de investigação.

Passamos a apresentar os resultados deste processo e a reflexão que presidiu à criação das etiquetas.

Numa primeira fase, seleccionámos as classes gramaticais principais:

	<b>Classes gramaticais</b>	<b>Etiquetas</b>
1	Adjectivo	Adj
2	Advérbio	Adv
3	Artigo	Art
4	Conjunção	Conj
5	Nome	N
6	Preposição	Prep
7	Pronome	Pron
7	Verbo	V

As preposições sofrem um tratamento particular. Optámos por desdobrar a etiqueta `en` em duas sub-etiquetas `Prep1` e `Prep2`. Foram etiquetadas com `en` as preposições, tais como *de*, *em*, *com*, *trás*, etc., enquanto que todas as contracções das preposições foram etiquetadas com `epc`.

Além destas classes gramaticais, seleccionámos 'Outros elementos gramaticais', que privilegiámos ao mesmo nível das classes gramaticais acima referidas:

	<b>Outros elementos gramaticais</b>	<b>Etiquetas</b>
1	Gerúndio	Ger
2	Locução	Loc
3	Nome próprio	Np
4	Numeral	Num
5	Particípio passado	Pp
6	Sigla	Sigla

A etiqueta `gen` é atribuída a todas as locuções, independentemente do seu função gramatical, não fazendo qualquer distinção entre elas.

Numa primeira instância, a etiqueta `part` é atribuída ao participio passado, independentemente de ter uma valor adjectival ou verbal. Após o levantamento de ambiguidades entre um e outro, atribuímos a etiqueta `partv` aos participios passados com valor verbal e a etiqueta `parta` aos participios passados com valor adjectival.

A etiqueta `abv`, por sua vez, é atribuída a toda as formas abreviadas, sejam elas, trunicações, siglas, acrónimos, ou outras.

Consideramos, no entanto, que para a extracção automática de unidades terminológicas complexas, as classes gramaticais seleccionadas, assim como as classes que classificámos com 'Outros elementos gramaticais', são demasiadamente genéricas. Acrescentámos, a todas as classes que o permitem, a indicação de género, de número e de pessoa:

	<b>Género</b>	<b>Etiquetas</b>
1	masculino	m
2	feminino	f
3	dois géneros	2gen
4	invariável	inv

	<b>Número</b>	<b>Etiquetas</b>
1	singular	s
2	plural	pl
3	dois números	2num

	<b>Pessoa</b>	<b>Etiquetas</b>
1	1ª pessoa	1p
2	2ª pessoa	2p
3	3ª pessoa	3p

Aos verbos, adicionámos a seguinte informação:

	<b>Conjugação</b>	<b>Etiquetas</b>
1	conjugado	cj
2	infinitivo	inf

As etiquetas atribuídas aos verbos são muito pouco descritivas e precisas, uma vez não ser necessária mais informação do que aquela que está mencionada na tabela para a extracção que pretendemos efectuar.

Para os artigos, que designam 'modo indeterminado', criámos as etiquetas:

	<b>Definido / Indefinido</b>	<b>Etiquetas</b>
1	definido	def
2	indefinido	indef

Os pronomes podem ser pessoais, possessivos ou demonstrativos:

	<b>Pronomes</b>	<b>Etiquetas</b>
1	pessoais	pes
2	possessivos	poss
3	demonstrativo	dem

Combinando todas as propriedades que cada uma das classes gramaticais pode assumir, chegámos a 168 etiquetas, que apresentamos no quadro seguinte:

	<b>Classes gramaticais</b> <b>Outros elementos gramaticais</b>	<b>Etiquetas</b>
1	Adjectivo	Adj:m:s
2	Adjectivo	Adj:m:pl
3	Adjectivo	Adj:f:s
4	Adjectivo	Adj:f:pl
5	Adjectivo	Adj:2gen:s
6	Adjectivo	Adj:2gen:pl
7	Advérbio	Adv
8	Artigo	Art:def:m:s
9	Artigo	Art:def:m:pl
10	Artigo	Art:def:f:s
11	Artigo	Art:def:f:pl
12	Artigo	Art:indef:m:s
13	Artigo	Art:indef:m:pl
14	Artigo	Art:indef:f:s
15	Artigo	Art:indef:f:pl
16	Conjunção	Conj:e
17	Conjunção	Conj:ou
18	Gerúndio	Ger
19	Locução	Loc
20	Nome	N:m:s
21	Nome	N:m:pl
22	Nome	N:f:s
23	Nome	N:f:pl
24	Nome	N:m:2num
25	Nome	N:f:2num
26	Nome	N:2gen:s
27	Nome	N:2gen:pl
28	Nome próprio	Np
29	Numeral	Num:ord
30	Numeral	Num:card
31	Particípio passado	Pp:m:s
32	Particípio passado	Pp:m:pl
33	Particípio passado	Pp:f:s
34	Particípio passado	Pp:f:pl
35	Preposição	Prep1:a
36	Preposição	Prep1:após
37	Preposição	Prep1:até

38	Preposição	Prep1:com
39	Preposição	Prep1:conforme
40	Preposição	Prep1:consoante
41	Preposição	Prep1:contra
42	Preposição	Prep1:de
43	Preposição	Prep1:desde
44	Preposição	Prep1:durante
45	Preposição	Prep1:em
46	Preposição	Prep1:entre
47	Preposição	Prep1:excepto
48	Preposição	Prep1:fora
49	Preposição	Prep1:menos
50	Preposição	Prep1:para
51	Preposição	Prep1:perante
52	Preposição	Prep1:por
53	Preposição	Prep1:sem
54	Preposição	Prep1:sob
55	Preposição	Prep1:sobre
56	Preposição	Prep1:trás
57	Preposição	Prep2:à:f:s
58	Preposição	Prep2:ao:m:s
59	Preposição	Prep2:aos:m:pl
60	Preposição	Prep2:àquela:f:s
61	Preposição	Prep2:àquelas:f:pl
62	Preposição	Prep2:às:f:pl
63	Preposição	Prep2:da:f:s
64	Preposição	Prep2:dai
65	Preposição	Prep2:daquela:f:s
66	Preposição	Prep2:daquelas:f:pl
67	Preposição	Prep2:daquele:m:s
68	Preposição	Prep2:daqueles:m:pl
69	Preposição	Prep2:daqui
70	Preposição	Prep2:daquilo
71	Preposição	Prep2:das:f:pl
72	Preposição	Prep2:dela:f:s
73	Preposição	Prep2:delas:f:pl
74	Preposição	Prep2:dele:m:s
75	Preposição	Prep2:deles:m:pl
76	Preposição	Prep2:dessa:f:s
77	Preposição	Prep2:dessas:f:pl
78	Preposição	Prep2:desse:f:s
79	Preposição	Prep2:desses:f:pl
80	Preposição	Prep2:desta:f:s
81	Preposição	Prep2:destas:f:pl
82	Preposição	Prep2:deste:m:s
83	Preposição	Prep2:destes:m:pl

84	Preposição	Prep2:disso
85	Preposição	Prep2:do:m:s
86	Preposição	Prep2:dos:m:pl
87	Preposição	Prep2:doutra:f:s
88	Preposição	Prep2:doutras:f:pl
89	Preposição	Prep2:doutro:m:s
90	Preposição	Prep2:doutros:f:pl
91	Preposição	Prep2:dum:m:s
92	Preposição	Prep2:duma:f:s
93	Preposição	Prep2:dumas:f:pl
94	Preposição	Prep2:duns:f:pl
95	Preposição	Prep2:na:f:s
96	Preposição	Prep2:nalguma:f:s
97	Preposição	Prep2:nalgumas:f:pl
98	Preposição	Prep2:nalgum:m:s
99	Preposição	Prep2:nalguns:m:pl
100	Preposição	Prep2:naquela:f:s
101	Preposição	Prep2:naquelas:f:pl
102	Preposição	Prep2:naquele:m:s
103	Preposição	Prep2:naqueles:f:s
104	Preposição	Prep2:nas:f:pl
105	Preposição	Prep2:nela:f:s
106	Preposição	Prep2:nelas:f:pl
107	Preposição	Prep2:nele:m:s
108	Preposição	Prep2:neles:m:pl
109	Preposição	Prep2:nessa:f:s
110	Preposição	Prep2:nessas:f:pl
111	Preposição	Prep2:nesse:m:s
112	Preposição	Prep2:nesses:f:pl
113	Preposição	Prep2:nesta:f:s
114	Preposição	Prep2:nestas:f:pl
115	Preposição	Prep2:neste:m:s
116	Preposição	Prep2:nestes:f:pl
117	Preposição	Prep2:no:m:s
118	Preposição	Prep2:nos:m:pl
119	Preposição	Prep2:noutra:f:s
120	Preposição	Prep2:noutras:f:pl
121	Preposição	Prep2:noutro:m:s
122	Preposição	Prep2:noutros:m:pl
123	Preposição	Prep2:num:m:s
124	Preposição	Prep2:numa:f:s
125	Preposição	Prep2:numas:f:pl
126	Preposição	Prep2:nuns:m:s
127	Preposição	Prep2:pela:f:s
128	Preposição	Prep2:pelas:f:pl
129	Preposição	Prep2:pelo:m:s

130	Preposição	Prep2:pelos:m:pl
131	Pronome	Pron:pes:1p:s
132	Pronome	Pron:pes:1p:pl
133	Pronome	Pron:pes:2p:s
134	Pronome	Pron:pes:2p:pl
135	Pronome	Pron:pes:3p:m:s
136	Pronome	Pron:pes:3p:m:pl
137	Pronome	Pron:pes:3p:f:s
138	Pronome	Pron:pes:3p:f:pl
139	Pronome	Pron:poss:1p:m:s
140	Pronome	Pron:poss:1p:m:pl
141	Pronome	Pron:poss:2p:m:s
142	Pronome	Pron:poss:2p:m:pl
143	Pronome	Pron:poss:3p:m:s
144	Pronome	Pron:poss:3p:m:pl
145	Pronome	Pron:poss:1p:f:s
146	Pronome	Pron:poss:1p:f:pl
147	Pronome	Pron:poss:2p:f:s
148	Pronome	Pron:poss:2p:f:pl
149	Pronome	Pron:poss:3p:f:s
150	Pronome	Pron:poss:3p:f:pl
151	Pronome	Pron:dem:1p:m:s
152	Pronome	Pron:dem:1p:m:pl
153	Pronome	Pron:dem:2p:m:s
154	Pronome	Pron:dem:2p:m:pl
155	Pronome	Pron:dem:3p:m:s
156	Pronome	Pron:dem:3p:m:pl
157	Pronome	Pron:dem:1p:f:s
158	Pronome	Pron:dem:1p:f:pl
159	Pronome	Pron:dem:2p:f:s
160	Pronome	Pron:dem:2p:f:pl
161	Pronome	Pron:dem:3p:f:s
162	Pronome	Pron:dem:3p:f:pl
163	Sigla	Sigla
164	Verbo	V:cj
165	Verbo	V:inf

A relevância que assumem as preposições nas unidades terminológicas multilexêmicas e, conseqüentemente, na constituição das tipologias, levou-nos a discriminar minuciosamente as preposições, por forma

a obter melhores resultados na aplicação das regras linguísticas no momento da concepção do *ExtractTerm* (cf. Capítulo IV, ponto 5 e Capítulo V).

Também desdobrámos os pronomes, por serem classes gramaticais que se encontram na proximidade contextual de nomes, adjectivos e preposições, que são as classes mais frequentes nas matrizes que constituem a tipologia, sendo necessário levantar ambiguidades com base nestas etiquetas (cf. Capítulo V).

A informação metalinguística, sob forma de etiqueta, foi adicionada manualmente a todas as formas constituintes dos dicionários.

### **3. Constituição de dicionários**

Após madura reflexão, optámos por construir três dicionários distintos, mas complementares, que são os pilares do *EtiqueLex*:

1. um dicionário de formas simples flexionadas;
2. um dicionário de locuções;
3. um dicionário de termos complexos da detecção remota.

#### **3.1. Dicionário de formas simples flexionadas**

##### **3.1.1 Desdobramento de formas**

Tal como já demos conta no ponto 1 deste capítulo, o dicionário que construímos a partir do *corpus*, com a ajuda do *Hyperbase*, é constituído,

exclusivamente, por formas simples que ocorrem no *corpus*. Após sucessivas fases de eliminação de ruído, o dicionário de formas acumula, aproximadamente, 10.800 formas.

A cada uma das formas encontradas no dicionário é atribuída uma etiqueta. É necessário proceder-se ao desdobramento de cada uma das formas registadas a fim de obter todas as suas formas flexionadas.

Do dicionário elaborado a partir de um *corpus* pode, por exemplo, constar unicamente a forma *colorida*, sendo preciso acrescentar a sua forma no plural, *coloridas*.

No caso do adjectivo *colorido*, adicionámos a sua forma plural, *coloridos*, assim como a forma feminina no singular e no plural, *colorida* e *coloridas*.

Nas situações em que uma mesma forma pode assumir funções gramaticais distintas, é imperioso desdobrá-la e etiquetar as formas resultantes do desdobramento efectuado, com todas as etiquetas gramaticais correspondentes às funções gramaticais que podem assumir em diferentes contextos.

É o que acontece com as formas etiquetadas com as etiquetas *Adj:f:s* e

Formas	Etiquetas
colorida	Adj:f:s
colorida	Pp:f:s
coloridas	Adj:f:pl
coloridas	Pp:f:pl
colorido	Adj:m:s
colorido	Pp:m:s
coloridos	Adj:m:pl
coloridos	Pp:m:pl

A forma foi primeiro desdobrada em duas, para assumir o valor de adjectivo e o de participio passado. De seguida, acrescentámos a forma masculina plural, assim como as formas femininas no singular, e o plural para cada uma das classes gramaticais.

### 3.1.2. Variantes ortográficas

Os termos cujas grafias não estão fixadas levam o especialista a hesitar no momento da redacção. Desta hesitação nascem diversas variantes ortográficas para um mesmo termo, em textos redigidos pelo mesmo autor.

Assim, o acrónimo [« »], importado do inglês para o sistema linguístico português, surge no *corpus* com as seguintes grafias:

Formas	Frequência
pixel	198
píxel	2
pixels	88
pixeis	51
píxeis	16

A noção de acrónimo já se perdeu, sendo este termo sentido como uma unidade monolexémica. Não estando o termo normalizado, os especialistas optaram pelas distintas situações, acima indicadas:

1. mantém a grafia inglesa, sendo o seu singular, e o seu plural

. Em termos de frequência, esta é a opção mais recorrente;

2. efectuam um decalque e acentuam o *i*, assinalando uma vogal aberta não tónica, , passando o termo a pertencer à classe das palavras graves. formando-se o seu plural de acordo com a regra de formação do português:

Nenhuma das outras opções é desejável, na medida em que nem respeitam a grafia portuguesa, nem a grafia inglesa.

Mas como a nossa função não é normalizadora, decidimos etiquetar todas as variantes ortográficas do termo que ocorrerem no *corpus*.

Enquanto para a grafia da forma , a hesitação dos especialistas é justificada, outras situações há em que a hesitação advém de desconhecimento das regras da língua.

Que fazer perante termos da especialidade sistematicamente mal grafados? Temos o caso de e toda a sua família morfológica que está, frequentemente, grafada com um só *r*, numa só palavra. Nestes casos, decidimos corrigir o *corpus*, por consideramos estar perante um erro ortográfico. Mantivemos, no entanto, as variantes ortográficas:

Formas	Etiquetas
geo-referenciação	N:f:s
geo-referenciada	Pp:f:s
geo-referenciada	Adj:f:s
geo-referenciadas	Pp:f:pl
geo-referenciadas	Adj:f:pl
geo-referenciados	Pp:m:pl
geo-referenciados	Adj:m:pl

geo-referenciado	Pp:m:s
Georreferenciado	Adj:m:s
Georreferenciação	N:f:s
Georreferenciada	Pp:f:s
Georreferenciada	Adj:f:s
Georreferenciadas	Pp:f:pl
Georreferenciadas	Adj:f:pl
Georreferenciados	Pp:m:pl
Georreferenciados	Adj:m:pl
Georreferenciado	Pp:m:s
Georreferenciado	Adj:m:s

Tal como podemos verificar, a partir do exemplo anterior, o especialista junta o formante à sua base, recorrendo ao hífen ou ao processo tradicional da derivação por prefixação.

Neste *corpus*, estas situações são muito frequentes, uma vez que a terminologia da Detecção Remota recorre com alguma regularidade a formantes, tais como: *multibanda*, *multi-banda*, *multibandas*, *multi-bandas*, *multiespectral*, *multi-espectral*, *multiespectrais*, *multi-espectrais*, ....

Em todas estas situações, mantivemos a dupla grafia:

<b>Formas</b>	<b>Etiquetas</b>
Multibanda	N:f:s
multi-banda	N:f:s
Multibandas	N:f:pl
multi-bandas	N:f:pl
Multiespectral	Adj:2gen:s
multi-espectral	Adj:2gen:s
Multiespectrais	Adj:2gen:pl
multi-espectrais	Adj:2gen:pl

Por este processo, o dicionário de formas flexionadas foi drasticamente aumentado, perfazendo a totalidade do dicionário, aproximadamente, 17200 formas.

### 3.2. Dicionário de locuções

No capítulo dedicado às expressões nominais complexas debatemos demoradamente as terminologias e os pressupostos teóricos que lhes estão subjacentes, utilizados pelos diversos autores para designar expressões coesas. Através da bibliografia consultada, verificámos que, para muitos autores, palavra composta, colocação, lexia complexa, frasema ou locução remetem para estruturas morfológicas, lexicais e sintácticas coesas (cf. Capítulo 3, p.113), sendo, frequentemente, empregues como sinónimos.

Porém, reservamo-nos o direito de conservar o termo locução para dar conta de uma sucessão de elementos lexicais e gramaticais que co-ocorrem de forma constante, não podendo os seus elementos constituintes ser actualizados individualmente, desempenhando o seu conjunto a função de um gramema. Estamos a referir-nos às locuções prepositivas, às locuções conjuncionais, assim como às locuções adverbiais. Esclarece-nos Rey [1997] a este respeito: *«Locution (du latin locutio , de loqui, «parler») est exactement «manière de dire», manière de former le discours, d'organiser les éléments disponibles de la langue pour produire une forme fonctionnelle. Ceci pourquoi on peut parler de «locutions adverbiales» ou «prépositives», alors que ces*



Uma locução como `de@Loc` seria etiquetada pelo EtiqueLex da seguinte forma:

Se a forma que preceder a preposição `de` for um nome e a forma que suceder a preposição `de` for um nome, o `ExtractTerm` gerará um falso termo, porque a regra `de@Loc` é um dos tipos que está incluído no dicionário de tipologias, sobre o qual irá actuar o `ExtractTerm`.

Assim, a etiquetagem prévia das locuções permite a redução do ruído no resultado da extracção dos termos, o que justifica a criação de um dicionário de locuções.

Em entrada, o dicionário de locuções apresenta a seguinte constituição:

a fim de@Loc  
à frente da@Loc  
à frente das@Loc  
à frente de@Loc  
à frente do@Loc  
à frente dos@Loc  
a jusante@Loc  
à medida de@Loc  
à medida que@Loc  
a par com@Loc  
a par de@Loc  
à parte@Loc  
a ponto de@Loc  
a princípio@Loc  
a propósito de@Loc  
a propósito@Loc

a	respeito	da@Loc
a	respeito	das@Loc
a	respeito	de@Loc
a	respeito	do@Loc
a	respeito	dos@Loc

Este dicionário será o primeiro a ser comparado com o *corpus*. O EtiqueLex procede à identificação da sucessão de elementos que se encontram no eixo sintagmático. Ao encontrar as locuções registadas no dicionário, o EtiqueLex prossegue, unindo os elementos constituintes com o sinal `&`, dando, deste modo, conta da sua coesão, acrescentando a etiqueta `Loc`, sendo o formato de saída o seguinte:

Neblinas ligeiras, poeiras, fumos mesmo não detectáveis na imagem da banda visível são, susceptíveis de distorcer bastante a temperatura aparente do solo verdadeiro. Não se pode, comparar directamente o mapa de registo termográfico de um espaço continental com o construído por interpolação a partir dos clássicos registos meteorológicos pontuais, estes não pretendem traduzir a temperatura da superfície terrestre mas escapar às suas irregularidades e fornecer a temperatura do ar, mantido à sombra do abrigo e um metro solo.

Esta etiquetagem evita que na comparação do *corpus* com o dicionário de formas flexionadas, o EtiqueLex atribua, autonomamente, etiquetas aos elementos constituintes das locuções.

Deste modo, previne-se uma etiquetagem errada, reduzindo a margem de erro do *ExtractTerm*. Ao ser-lhe vedado a hipótese de etiquetar os

constituintes das locuções, o ExtracTerm encontra-se na impossibilidade de aplicar as regras linguísticas do qual depende, em parte, a sua existência.

### 3.3. Dicionário de unidades terminológicas multilexémicas da detecção remota

A criação de um dicionário de unidades terminológicas complexas de uma área de especialidade só é pertinente caso haja recursos linguísticos já disponíveis. Pensamos, efectivamente, ser proveitoso em tempo e em esforço recorrer a trabalho previamente desenvolvido.

Os recursos linguísticos da Detecção Remota que dispomos são o resultado de um projecto, desenvolvido no seio do Praxis<sup>1</sup>. Para o efeito criámos uma base de dados terminológicos, utilizando para o efeito uma metodologia de processamento semi-automático de *corpora*. As unidades terminológicas recolhidas por este processo já foram, parcialmente, validadas por especialistas, o que nos permite utilizá-las com segurança.

Deste modo, chegamos ao terceiro dicionário que, também ele, faz parte do EtiqueLex. Tal como para o dicionário das locuções, este dicionário tem o propósito de fazer ganhar tempo ao ExtracTerm, evitando que ele aplique regras linguísticas a estruturas que, *a priori*, já conhece.

O recurso a este dicionário só faz sentido se o *corpus* tratado for da área de especialidade correspondente aos recursos linguísticos já existentes.

---

<sup>1</sup> «*Arquivos Electrónicos de Terminologias e "corpora"*», PRAXIS XXI (Projecto PRAXIS XXI - JNICT nº CSH - 717 - 95, U.N.L.), Projectos de Investigação Científica e Tecnológica de pequena e média dimensão, direcção científica a cargo da Prof.<sup>a</sup> Doutora Maria Teresa Rijo da Fonseca Lino.

O dicionário contém a seguinte informação:

absorção atmosférica@N:f:s  
agregação de objectos@N:f:s  
agregação espacial@N:f:s  
água líquida@N:f:s  
água profunda@N:f:s  
amplitude térmica@N:f:s  
ângulo azimutal@N:m:s  
ângulo de iluminação@N:m:s  
ângulo de varrimento@N:m:s  
ângulo zenital de iluminação@N:m:s  
ângulo zenital@N:m:s  
assinatura espectral@N:f:s  
detecção remota@N:f:s  
filmes pancromáticos@N:m:pl  
filmes pancromáticos@N:m:pl  
filtragem de imagens@N:m:s  
filtragens de imagens@N:m:s  
fluxo de calor@N:m:s  
fotointerpretação assistida por computador@N:f:s  
foto-interpretação assistida por computador@N:f:s  
imagem de textura@N:f:s

Tal como tivemos oportunidade de referir, o frasema, na sua totalidade, pode em situação textual comportar-se como uma unidade monolexémica. Assim, ao conjunto do frasema é atribuída uma única etiqueta, correspondente à função gramatical que desempenha no texto.

No momento em que o EtiqueLex compara o *corpus* ao respectivo dicionário e estabelece a correspondência, o formato final que se alcança é o seguinte:

Incluíram-se, nesta unidade, os pequenos bancos de ostra da Palma, Zambujal, Carrasqueira e Comporta, que, por estarem cobertos de algas verdes e castanhas, apresentam vegetação superior halófitas. Na metodologia seguida foi utilizada a mesma metodologia usada na classificação das classes de acordo com os histogramas bidimensionais, as curvas de perfil e os perfis radiométricos.

No caso do lexema que segue ou antecede a unidade etiquetada ser um elemento que conste numa das estruturas que constitui a matriz terminogénica, o sistema reconhece a totalidade do frasema como um elemento e opera sobre esta da mesma forma que opera sobre um lexema.

Apesar do termo ser identificado como um frasema, o EtiqueLex também identifica os termos e os morfemas em outros contextos, uma vez que eles se encontram no dicionário de formas flexionadas. Assim, registamos todos os frasemas em que ocorrem os lexemas terminológicos e os morfemas, embora não estejam armazenados no dicionário de Detecção Remota.

Este sistema permite-nos uma constante actualização do dicionário de frasemas, uma das possíveis aplicações resultante deste sistema.

#### 4. Etiquetador gramatical: EtiqueLex

O EtiqueLex foi concebido e construído tendo por base os três dicionários acima descritos. Na atribuição de etiquetas, o etiquetador EtiqueLex percorre o *corpus*, comparando este com os conteúdos do dicionário de locuções e do dicionário de Detecção Remota, uma vez que a estrutura de dados é a mesma.

Se, no momento da comparação, a primeira forma do *corpus* encontrada corresponder à primeira forma da sequência que constitui as locuções e os frasemas que constam dos devidos dicionários, o EtiqueLex continua a sua comparação no eixo sintagmático, até atingir o fim da sequência comparada, assinalando-a com a etiqueta correspondente, unindo os seus elementos com um

Caso a forma encontrada no *corpus* não corresponda à primeira forma de uma locução ou de um frasema previamente identificado, o EtiqueLex passa de imediato à comparação com o dicionário de formas simples flexionadas, atribuindo a etiqueta correspondente. As formas não etiquetadas equivalem a formas que não se encontram em nenhum dos dicionários.

Desta feita, o EtiqueLex percorre o *corpus* uma só vez, executando, em complementaridade, os dicionários de locuções e de frasemas e o dicionário de formas simples flexionadas.

O algoritmo que permite desempenhar esta tarefa é o seguinte:

```
{
    CDicionarioEx*
    dicLocucoesFrasemas;
    CDicionario*
    dicEtiquetas;
```

```

std::vector<std::string>          entrada;
std::string                      token,
etiqueta, tmpTok;                i = 0;
int
long
cracteres_tratados = 0;
std::vector<std::string>::iterator it;
std::vector<CEntrada>::iterator
iterator_m_corpus_dividido;
char                              tempoIni[1024];
char                              tempoFim[1024];
std::vector<std::string>::iterator icEtiquetas;
std::vector<std::string>          vazio_etiquetas;
CEntrada                          tmpEntrada;

vazio_etiquetas.clear();

dicLocucoesFrasemas = (CDicionarioEx*)
m_Dicionarios[FRASEMAS_LOCUCOES];
dicEtiquetas = m_Dicionarios[FLEXIONADAS];

std::vector<std::string> tt = dicEtiquetas-
>DevolveEtiqueta("os");

boost::tokenizer<> tok(m_Corpus);

_strtime(tempoIni);

std::cout << tempoIni << " : Inicio etiquetagem corpus (cada
',' 100k tratados)" << std::endl;

for(boost::tokenizer<>::iterator beg = tok.begin(); beg !=
tok.end(); ++beg, i++)
{
    token = *beg;

    if ( token[0] == '\n' )
    {
        tmpEntrada.entrada = "<br>";
        tmpEntrada.etiqueta = vazio_etiquetas;
        tmpEntrada.etiquetado = false;

        m_corpus_dividido.push_back(tmpEntrada);

        token = token.substr(1, token.length() ).c_str();
    }

    tmpTok = token.c_str();

    while ( dicLocucoesFrasemas-
>ProcuraEntrada(_strlwr((char*)tmpTok.c_str())) )
    {
        entrada.push_back( token );

        ++beg;
        token = *beg;
        tmpTok = token.c_str();
    }
}

```

```

        etiqueta = dicLocucoesFrasemas-
>DevolveEtiqueta(entrada);
        if ( etiqueta != "" )
        {
                CEntrada
                tmpEntrada;

                for ( it = entrada.begin(); it !=
entrada.end(); it++ )
                {
                        tmpEntrada.entrada += (*it);
                        if ( it != entrada.end() -1)
                                tmpEntrada.entrada += "_";
                }

                tmpEntrada.etiqueta.push_back(etiqueta);
                tmpEntrada.etiquetado = true;

                m_corpus_dividido.push_back(tmpEntrada);

                entrada.clear();
        }
}

for ( it = entrada.begin(); it != entrada.end(); it++ )
{
        CEntrada    tmpEntrada;

        tmpTok = (*it).c_str();
        tmpEntrada.entrada = tmpTok.c_str();
        tmpEntrada.etiqueta = dicEtiquetas-
>DevolveEtiqueta(_strlwr((char*)tmpTok.c_str()));
        if ( tmpEntrada.etiqueta.size() > 0 )
                tmpEntrada.etiquetado = true;
        else
                tmpEntrada.etiquetado = false;

        m_corpus_dividido.push_back(tmpEntrada);
}

entrada.clear();

if ( token.length() < 1024 )
{
        tmpTok = token.c_str();

        tmpEntrada.entrada = token;
        tmpEntrada.etiqueta = dicEtiquetas-
>DevolveEtiqueta(_strlwr((char*)tmpTok.c_str()));
        if ( tmpEntrada.etiqueta.size() > 0 )
                tmpEntrada.etiquetado = true;
        else
                tmpEntrada.etiquetado = false;

        m_corpus_dividido.push_back(tmpEntrada);
}

caracteres_tratados += token.size() + 1; // conto mais 1
(espaço)

```

```

        if ( caracteres_tratados > 1024 * 100 )
        {
            caracteres_tratados = 0;
            std::cout << ".";
        }
    }

    _strtime(tempoFim);

    std::cout << std::endl << tempoFim << " : Fim etiquetagem
corpus." << std::endl;

```

Para cada forma encontrada, o EtiqueLex atribui todas as etiquetas correspondentes a essa mesma forma, sendo o resultado como se segue:

A utilização possível dos Registos de Satélites em Infravermelho Térmico na Climatologia Regional. Os satélites destinados a facilitar a previsão do tempo registam, em geral, conjuntamente no comprimento de ondas que corresponde à visão humana e ao máximo da radiação solar (Visível e Infravermelho próximo), entre 0,4 e 1,1 (m) e no correspondente à janela atmosférica (entre 10 e 12,5 (m) que deixa escapar em direcção ao espaço uma parte da irradiação da superfície terrestre (Infravermelho médio ou térmico).

Partindo da observação dos resultados obtidos com o EtiqueLex, falta-nos percorrer duas grandes etapas.

A primeira etapa consiste no estabelecimento de matrizes terminogénicas que contenham as tipologias correspondentes às estruturas morfolexicais mais produtivas desta terminologia.

A segunda etapa consiste no estabelecimento de regras linguísticas que nos permitam levantar as ambiguidades das classes gramaticais por forma a obtermos o maior número possível de frasemas correctos, de acordo com a tipologia estabelecida.

## **5. Tipologia de matrizes terminogénicas**

### **5.1. Tipologia elementar**

No âmbito desta dissertação, pretendemos identificar e delimitar os frasemas que, num contexto terminológico textual, assumem o estatuto de denominações. Partimos do pressuposto teórico de que o frasma remete para um único conceito, correspondendo a sua denominação a uma unidade terminológica. As unidades terminológicas são denominações que, sendo monolexémicas, assumem, maioritariamente, o estatuto de nome. No caso de serem unidades multilexémicas comportam-se em contexto textual como uma unidade monolexémica, sendo, por esse motivo, a sua forma base um nome. Considera Sager que: *«Du point de vue formel, le terme est [...] un signe linguistique (il se présente alors sous la forme d'un nom), [...]. Dans le discours écrits ces signes sont considérés comme des noms [...]. S'il [terme]*

*est pris comme un signe linguistique et comme une unité syntagmatique, le terme est une variété fonctionnelle du nom comum» (Sager, 2000:53).*

Na observação do *corpus* da Detecção Remota, tal como já tinha acontecido com o *corpus* da Economia Monetária (Costa, 1993), verificámos que, na constituição das unidades terminológicas, as classes gramaticais mais produtivas são os nomes, os adjectivos e as preposições, que fazem parte das 100 formas mais frequentes do *corpus*:

Ordem	Frequência	Forma
1	13072	de
2	11938	,
3	9166	.
4	7762	a
5	5411	e
6	3932	da
7	3825	o
8	3626	-
9	3266	do
10	3157	que
11	2814	se
12	2708	(
13	2633	em
14	2540	)
15	2028	os
16	1942	uma
17	1875	para
18	1831	dos
19	1812	As
20	1769	um
21	1664	com
22	1607	é
23	1592	por
24	1554	das
25	1413	no
26	1337	na
27	1228	imagem
28	1031	ou

29	1010	imagens
30	987	:
31	931	à
32	806	1
33	791	"
34	782	não
35	767	ao
36	766	como
37	758	;
38	716	2
39	700	ser
40	685	são
41	628	mais
42	607	informação
43	577	3
44	552	entre
45	538	dados
46	512	satélite
47	488	sobre
48	445	cada
49	414	energia
50	401	valores
51	378	pode
52	377	pela
53	363	este
54	360	foi
55	356	pelo
56	353	sua
57	346	solo
58	345	sistema
59	343	forma
60	342	detecção
61	336	áreas
62	333	[
63	328	x
64	323	análise
65	322	4
66	319	superfície
67	316	]
68	314	região
69	306	terreno
70	305	classificação
71	297	nas
72	296	diferentes
73	294	tem
74	281	nos

75	278	0
76	274	Utilização
77	270	área
78	264	foram
79	262	resolução
80	261	conjunto
81	260	tipo
82	260	água
83	257	5
84	252	remota
85	252	partir
86	252	bandas
87	251	podem
88	249	reflectância
89	247	m
90	246	também
91	243	às
92	242	esta
93	236	espectro
94	233	assim
95	231	sistemas
96	231	pontos
97	230	Terra
98	228	/
99	226	banda
100	225	espectral

Além destas classes, privilegiámos os nomes próprios e as siglas que também desempenham um papel relevante nesta terminologia.

Partindo deste conhecimento, elaborámos uma primeira matriz que dá conta das estruturas elementares das unidades terminológicas.

Obtivemos a seguinte tipologia:

A tipologia é, deste modo, constituída por sete tipos principais, correspondentes às estruturas elementares das unidades terminológicas que desempenham a função de uma unidade monolexémica em texto.

## **5.2. Análise da composição das unidades terminológicas multilexémicas**

A extensão sintagmática das unidades terminológicas multilexémicas é imprevisível, sendo as fronteiras entre a colocação, o frasema e a fraseologia, a fraseologia e a proposição nem sempre muito transparentes. Identificar e proceder ao processamento automático das unidades terminológicas no seio destas diferentes realidades linguísticas, constitui um grande desafio, inerente ao tratamento automático dos textos de especialidade.

Em língua de especialidade, as denominações são na sua essência constituídas por lexemas e frasemas, podendo da sua conexidade resultar fraseologias. Apesar de, do ponto de vista teórico, a distinção entre umas e outras estar estabelecida, na prática as diferenças nem sempre são facilmente detectáveis.

Em 1991, Jacquemin (*in* Daille, 1994) utiliza o conceito de sobrecomposição para dar conta da forma como as estruturas elementares se conjugam entre si, dando origem a novas unidades multilexémicas.

De acordo com a tese defendida por Jacquemin, a sobrecomposição recorre a duas operações: a sobrecomposição por justaposição e a

sobrecomposição por «cobertura» («*recouvrement*»), podendo esta ser parcial.

Estamos perante um processo de sobrecomposição por justaposição quando, por exemplo, se acrescenta um adjectivo a uma estrutura elementar já atestada. No exemplo *le mot de la semaine*, o adjectivo *de la semaine* é adicionado à unidade terminológica *le mot*, cuja estrutura corresponde a uma das estruturas elementares que constam da tipologia acima apresentada:

A sobrecomposição por cobertura («*recouvrement*») equivale à junção de duas unidades terminológicas já atestadas, deixando cair o segundo elemento da primeira unidade ou o primeiro elemento da segunda unidade:

*le mot de la semaine* e *le mot de la semaine* resulta no termo *le mot de la semaine*, que corresponde a uma das seguintes estruturas:  $N_1$  ou então  $N_2$

Daille reformula o conceito de sobrecomposição. Aceita a sobrecomposição por justaposição, rejeita a sobrecomposição por cobertura («*recouvrement*»), preferindo neste caso utilizar o conceito de sobrecomposição por substituição, que define do seguinte modo: « [...] *étant donné un nom composé de longueur 2, l'une des unités lexicales pleines est substituée par un nom composé dont la tête est cette unité lexicale*» (Daille, 1994:58).

Podemos ilustrar esta definição com o exemplo seguinte:

*le mot de la semaine* + *le mot de la semaine* → *le mot de la semaine*, cuja tipologia corresponde à seguinte estrutura:  $N_1$  +  $N_2$  →  $N_1$

A unidade terminológica retoma, assim, o primeiro lexema da primeira unidade multilexémica e o último lexema da segunda unidade multilexémica, sendo criado, desta forma, um novo frasema. Este tipo de composição é possível quando o termo base desempenha a função de hiperónimo, sendo este o caso de , termo extremamente produtivo na área da Detecção Remota.

Mas, para Daille, o conceito de sobrecomposição é insuficiente para dar conta dos processos de modificação operados sobre os tipos elementares por esta autora definidos: *«Les modificateurs apparaissent avant, à l'intérieur ou après le nom composé. Les modificateurs qui apparaissent avant ou après ne cassent pas la structure, au contraire de ceux qui apparaissent à l'intérieur. Ces deux cas de modifications peuvent donner lieu à de nouveaux noms composés [...]»* (Daille, 1994:59).

Mas a distinção entre a sobrecomposição e a modificação também não é satisfatória, uma vez que um processo muitas das vezes pressupõe o outro, o que não os torna totalmente operativos. A própria autora assim o considera: *«Ces deux phénomènes ont des définitions linguistiques différentes mais les données étudiées révèlent que la frontière réelle est floue.»* (Daille, 1994:66).

Pensamos que o conceito de sobrecomposição pode ser utilizado como um genérico, para dar conta de processos morfológicos que permitem combinar entre si as várias estruturas que compõem as matrizes terminogénicas. O resultado desses processos morfológicos redundam em sequências que designamos por colocações, frasemas ou fraseologias, consoante os casos em análise.

Passamos a analisar algumas das estruturas que as unidades terminológicas complexas ou 'frasemas terminológicos' podem assumir e as diferenças que estes apresentam relativamente às fraseologias.

Para ilustrar os nossos propósitos, partimos de um enunciado retirado do *corpus* em Detecção Remota. Assinalámos a azul as sequências que correspondem aos tipos que constituem a tipologia elementar e a vermelho as formas que se repetem no fim e no princípio de dois tipos distintos:

*De entre os vários sistemas de mais geral utilização destacam-se os que permitem medir as densidades das tonalidades [N] (os microdensitómetros), efectuar a análise estereoscópica de imagens de várias composições coloridas e bandas distintas do espectro.*

De acordo com a tipologia elementar (cf. Capítulo IV, ponto 5, 159), obtém-se as seguintes sequências:

- (1) densidades das tonalidades
- (2) tonalidades das imagens
- (3) análise estereoscópica
- (4) pares de fotografias
- (5) obtenção de composições
- (6) composições coloridas
- (7) bandas distintas

Apesar das sequências (1) e (2) obedecerem à estrutura elementar e dos lexemas que as compõem serem terminológicos, não podemos

considerá-las unidades terminológicas complexas, porque não denominam nenhum conceito. Como o último lexema da primeira sequência corresponde ao primeiro lexema da segunda sequência, podemos fundir os dois tipos elementares, para passar a ter um estrutura combinada que integrará a tipologia de base (cf. Capítulo IV, ponto 5.3, p. 170):

$$N_2 + N_2 + \dots \rightarrow N_2$$

Esta fusão dá origem à sequência  $N_2 + N_2 + \dots$  tonalidades

, que não corresponde a uma unidade terminológica, nem a uma fraseologia. Podemos considerar esta sucessão de lexemas uma 'sequência terminológica livre', pois embora a concentração de lexemas monolexémicos terminológicos no eixo sintagmático seja elevada, da proximidade destes não resulta qualquer coesão, não podendo a sequência ser confundida com um colocação terminológica, um fraseo terminológico ou uma fraseologia.

Procedendo da mesma maneira para as sequências (5) e (6), alcançamos, porém, resultados diferentes dos anteriores:

$$N_2 + N_2 \rightarrow N_2$$

A sequência obtida é  $N_2 + N_2$  composições

Neste caso, consideramos estar perante um exemplo de fraseologia, da qual se pode extrair uma unidade terminológica complexa

, na medida em que esta última denomina um conceito. Identificamos

na sequência *desenvolvimento* composições *desenvolvidas*, uma fraseologia, por o lexema *desenvolvido* ser um nome deverbal, que deriva do particípio passado do verbo *desenvolver*, denotando este tipo de nome acção.

Assim, estamos perante dois conceitos, um com característica de verbo (cf. capítulo 3, ponto 3) se bem que expresso por um nome, *desenvolvido*, outro com característica de objecto, sendo expresso por uma unidade terminológica complexa *desenvolvimento desenvolvido*. O nome *desenvolvido*, facilmente, pode ser substituído pelo verbo *desenvolver*, passando a sequência a ser *desenvolvimento desenvolvido*.

A obtenção e a aquisição de informação estão na base da Detecção Remota, o que justifica a frequência dos paradigmas dos verbos *desenvolver* e *desenvolvido*. Num *corpus*, que conta com 12.888 formas, o nome *desenvolvido* tem uma frequência de 77 ocorrências, o particípio passado *desenvolvido* tem uma frequência 364 e a totalidade das formas conjugadas ocorre 129 vezes, perfazendo um total de 647 ocorrências.

As fraseologias, as colocações e algumas sequências terminológicas livres, embora não correspondam a unidades terminológicas, têm uma importância crescente em terminologia textual, uma vez que nos permitem identificar unidades discursivas recorrentes numa área de especialidade, fundamentais para o processamento automático da informação.

Em *desenvolvimento desenvolvido* (sequência 3), identificamos dois lexemas terminológicos, sendo o primeiro um termo genérico de orientação científica geral, considerando que é utilizado, em áreas tão diversas como a literatura, a química, a psicanálise, ...

Em contrapartida, o adjectivo *remota* é um lexema terminológico da Detecção Remota. No entanto, a sequência não corresponde a uma unidade terminológica, podendo ser retida como uma colocação, porque embora *remota* não seja um termo especificamente da especialidade, tem uma frequência 323 (cf. Capítulo 4, ponto 5.1), sendo a sexagésima quarta frequência mais elevada do *corpus*. Em conformidade com esta estrutura, encontramos sequências como: *remota*, *remota* e *remota*, ..., comprovando a produtividade desta combinatória.

No entanto, mais uma vez as realidades linguísticas em análise são distintas. *remota* e *remota* são frasemas terminológicos, enquanto que *remota* é uma colocação que está ao mesmo nível que *remota*. Os dois primeiros denominam conceitos, os dois últimos não.

Estas distinções compreendem-se pelas características dos adjectivos que modificam os nomes. Os adjectivos *remota* e *remota* são relacionais, porque derivam de uma base nominal, posicionando-se sempre à direita do nome, não aceitando gradação:

Por outro lado, *remota* são sempre subcategorizados pelos substantivos que modificam, mas não qualificam, podendo esta subcategorização dar conta de

uma relação estável ou efêmera entre o substantivo e o adjetivo. Se a relação for estável «[...] *la lexicalisation entraine le blocage de la formule, qui ne peut subir aucune modification sans perdre sa référence habituelle* [...]» (Noailly, 1999:22), se, em contrapartida, a relação for efêmera «[...] *une paraphrase réintroduisant en jeu le substantif qui est à la base de la dérivation permet de restituer le sens souhaité sans problème*» (Noailly, 1999:22).

No caso de *analyse de texture*, pensamos que a relação é efêmera por dois motivos: o primeiro, prende-se com uma questão extralinguística: esta combinação de lexemas não é coesa, porque não dá conta de nenhum conceito.

O segundo motivo, é de ordem linguística. Os adjectivos relacionais indicam uma relação com o referente do nome do qual derivam. O nome *analyse* não subcategoriza o adjectivo *de textura*, já que neste caso é o resultado da observação de *analyse* e não o resultado da análise de uma técnica em si. Assim, não podemos substituir *analyse de texture* por \**analyse de texture*, ou ainda *analyse de texture* por \**analyse de texture*.

Em compensação, os termos *analyse de spectre* e *analyse de texture* referem-se a análises de resultados obtidos sobre o espectro e sobre a textura, podendo por estes motivo serem substituídos por *analyse de spectre* e *analyse de texture*, constituindo, *analyse de spectre* e *analyse de texture* os equivalentes sintácticos e semânticos de um complemento de nome.

Só o nome *analyse*, lexema base da sequência seguinte (4), *analyse de spectre*, é que permite a subcategorização do adjectivo *de spectre*, já que

que este *sim*, é uma unidade terminológica, cuja estrutura elementar corresponde à expressão

O frasema *sim, sim* sofreu um processo de elipse, sendo o lexema terminológico, *sim*, omitido. O termo é, assim, *sim*, sendo a sua estrutura que não está contemplada na tipologia elementar.

Caso juntemos as sequências (3) e (4) obtemos a sequência *sim, sim, sim*, cuja estrutura também não pertence aos registos da tipologia elementar:

O adjectivo relacional *sim* está sintacticamente mal colocado, não podendo este ficar separado do lexema *meses*, uma vez que é por ele subcategorizado, concordando o adjectivo em número e em género com o respectivo nome. Apesar desta má colocação sintáctica somos induzidos a fazer o levantamento desta sequência, uma vez que ela ocorre tal e qual no *corpus* e é com ela que temos que operar:

(1) *sim, sim, sim*

→ *sim, sim, sim*

(2) *sim, sim, sim*

→ *sim, sim, sim*

Nestes dois exemplos, estamos perante duas fraseologias. No primeiro exemplo, a fraseologia é constituída por uma colocação, que através da análise se demonstrou ser incorrecta no seu

processo de formação, e uma unidade terminológica complexa,

. No segundo exemplo, a fraseologia é constituída por uma unidade terminológica simples, , e por uma unidade terminológica complexa,

.

Em ambos os casos estamos perante a junção de dois conceitos, com características de objecto.

Se o processo de elipse acima operado sobre *fraseologia* for aplicado neste contexto fraseológico, obtemos as seguintes variantes:

(3) *fraseologia* *fraseologia* *fraseologia*

→ *fraseologia* *fraseologia*

(4) *fraseologia*

→ *fraseologia*

A sequência 3 continua inaceitável, uma vez que o adjectivo continua mal colocado. Por sua vez, a sequência 4, sem a omissão *fraseologia*, equivale a um enunciado correcto, deixando de o ser com a aplicação do processo elíptico. Neste contexto, a elipse não é desejável, uma vez que altera a correcção sintáctica do enunciado, trazendo essa modificação alteração ao nível do conceito, introduzindo ambiguidade no discurso.

Resta-nos ainda analisar o último exemplo, *fraseologia* que corresponde a uma sequência livre de lexemas: *fraseologia* é um lexema terminológico monolexémico, acompanhado de um outro lexema, um adjectivo qualificativo de uso corrente.

No entanto, se juntarmos tal sequência à seguinte *do espectro*, Prep + N obtém-se uma combinatória interessante a reter *bandas distintas do espectro*, que corresponde à seguinte estrutura: N + Adj + Prep + N. Neste caso concreto, há o encaixe do adjectivo *distintas*, que vem modificar, temporariamente, o termo base *bandas* da unidade terminológica multilexémica *bandas do espectro*, acarretando, neste caso, uma alteração na sua estrutura tipo. Esta unidade faz parte das estruturas que nos interessa reter, podendo ser representada da seguinte forma:  $[N_1 + (Adj) + Prep_1 + N_2]_N$ .

Destas análises retemos alguns fundamentos para a construção do ExtracTerm:

- (i) As estruturas que compõem a tipologia elementar podem ser combinadas por forma a obter sequências máximas, colocações e fraseologias, que se revelam ser contextos mínimos de frasemas terminológicos.
- (ii) O alargamento da tipologia elementar impõe-se. As combinatórias gramaticais e estruturais comprovam-se proficuas, originando uma tipologia de base, derivada da tipologia elementar, que estará na cerne do dicionário de matrizes sobre o qual é concebido o ExtracTerm.
- (iii) A comprovação da importância das colocações e fraseologias para a identificação de frasemas terminológicos e para a apreensão de estruturas discursivas mais alargadas que podem servir, entre outros, a tradução automática.





**N<sub>1</sub> + Adj<sub>1</sub> + Sigla<sub>1</sub>**

⇒ [[imagens digitalizadas ] Landsat ]

**N<sub>1</sub> + N<sub>2</sub>**

⇒ [imagem textura ]

**N<sub>1</sub> + hífen + N<sub>2</sub>**

⇒ [imagem - textura ]

⇒ [imagens - carta ]

**N<sub>1</sub> + N<sub>2</sub> + Adj<sub>1</sub>**

⇒ [imagem satélite ] digitalizada

**N<sub>1</sub> + N<sub>2</sub> + Sigla<sub>1</sub>**

⇒ [[imagens textura ] Spot ]

**N<sub>1</sub> + Prep + N<sub>2</sub>**

⇒ [imagem de satélite ]

**N<sub>1</sub> + Prep + Np<sub>1</sub>**

⇒ [lei de Planck ]

**N<sub>1</sub> + Prep + Np<sub>1</sub> + Conj + Np<sub>2</sub>**

⇒ [efeito de Doppler e Lorentz ]

**N<sub>1</sub> + Np<sub>1</sub>**

⇒ [efeito Doppler ]

**N<sub>1</sub> + Np<sub>1</sub> + Conj + Np<sub>2</sub>**

⇒ [efeito Doppler e Lorentz ]

**N<sub>1</sub> + Sigla<sub>1</sub>**

⇒ [imagem Landsat ]

**N<sub>1</sub> + Sigla<sub>1</sub> + Adj<sub>1</sub>**

⇒ [[imagem SPOT ] colorida]

**N<sub>1</sub> + Sigla<sub>1</sub> + Sigla<sub>2</sub>**

⇒ [[imagem Landsat ] - MSS ]

**Sigla<sub>1</sub> + Sigla<sub>2</sub>**

⇒ [Landsat TM ]

Acrescentando as siglas e os nomes próprios às três classes gramaticais, nomes, adjectivos e preposições, construimos a tipologia de base, que conta com 19 matrizes, que resultam de processos de

sobrecomposição e de combinatórias de 7 estruturas elementares (cf. Capítulo IV, ponto 5.1.).

#### 5.4. Dicionário de estruturas tipo.

Os 19 tipos de base que compõem a matriz terminogénica foram desdobrados em todas as suas variantes, tendo em conta as regras da língua portuguesa e de acordo com a informação contida na tabela de etiquetas (cf. Capítulo 4, ponto 2).

Além das variantes, demos conta de todas as suas combinatórias, o que resultou num dicionário de estruturas tipo de unidades terminológicas complexas.

O Dicionário de tipologias é, por conseguinte, constituído por 666 matrizes, do qual apresentamos uma amostra exemplificativa:

#### **N + Adj**

N:m:s	Adj:m:s
N:m:pl	Adj:m:pl
N:f:s	Adj:f:s
N:f:pl	Adj:f:pl
N:2gen:s	Adj:2gen:s
N:2gen:pl	Adj:2gen:pl

#### **N+Prep+N ⇒ N + Prep1:de + N**

N:m:s	Prep1:de	N:m:s
N:m:s	Prep1:de	N:m:pl
N:m:s	Prep1:de	N:f:s
N:m:s	Prep1:de	N:f:pl
N:m:s	Prep1:de	N:2gen:s
N:m:s	Prep1:de	N:2gen:pl
N:m:pl	Prep1:de	N:m:s

N:m:pl	Prep1:de	N:m:pl
N:m:pl	Prep1:de	N:f:s
N:m:pl	Prep1:de	N:f:pl
N:m:pl	Prep1:de	N:2gen:s
N:m:pl	Prep1:de	N:2gen:pl
N:f:s	Prep1:de	N:m:s
N:f:s	Prep1:de	N:m:pl
N:f:s	Prep1:de	N:f:s
N:f:s	Prep1:de	N:f:pl
N:f:s	Prep1:de	N:2gen:s
N:f:s	Prep1:de	N:2gen:pl
N:f:pl	Prep1:de	N:m:s
N:f:pl	Prep1:de	N:m:pl
N:f:pl	Prep1:de	N:f:s
N:f:pl	Prep1:de	N:f:pl
N:f:pl	Prep1:de	N:2gen:s
N:f:pl	Prep1:de	N:2gen:pl
N:2gen:s	Prep1:de	N:m:s
N:2gen:s	Prep1:de	N:m:pl
N:2gen:s	Prep1:de	N:f:s
N:2gen:s	Prep1:de	N:f:pl
N:2gen:s	Prep1:de	N:2gen:s
N:2gen:s	Prep1:de	N:2gen:pl
N:2gen:pl	Prep1:de	N:m:s
N:2gen:pl	Prep1:de	N:m:pl
N:2gen:pl	Prep1:de	N:f:s
N:2gen:pl	Prep1:de	N:f:pl
N:2gen:pl	Prep1:de	N:2gen:s
N:2gen:pl	Prep1:de	N:2gen:pl

**N + Adj + Prep + N ⇒ N + Adj + Prep1:pele + N**

N:m:s	Adj:m:s	Prep2:pele:m:s	N:m:s
N:m:pl	Adj:m:pl	Prep2:pele:m:s	N:m:s
N:f:s	Adj:m:s	Prep2:pele:m:s	N:m:s
N:f:pl	Adj:f:pl	Prep2:pele:m:s	N:m:s
N:2gen:s	Adj:2gen:s	Prep2:pele:m:s	N:m:s
N:2gen:pl	Adj:2gen:pl	Prep2:pele:m:s	N:m:s

É com base neste dicionário de estruturas tipo que o `ExtractTerm` efectuará a primeira identificação de frasemas. (cf. Capítulo V).

## Capítulo V

---

### **Extractor automático de unidades terminológicas multilexémicas: ExtracTerm**

1. Conceção do ExtracTerm
  - 1.1. Aplicação do dicionário de tipologias
  - 1.2. Análise dos resultados
2. Regras de desambiguação
  - 2.1. Preposição, artigo definido feminino singular e pronomes
    - 2.1.1. Regras linguísticas de aprendizagem
  - 2.2. Adjectivo e participio passado pós-nominal.
    - 2.2.1. Regras linguísticas de aprendizagem
  - 2.3. Nome e adjectivo
    - 2.3.1. Regras linguísticas de aprendizagem
3. ExtracTerm: utilização das regras de aprendizagem
  - 3.1. Análise dos resultados
  - 3.2. Teste do ExtracTerm sobre um *corpus* alargado
    - 3.2.1. Análise dos resultados

## **Capítulo V - Extractor automático de unidades terminológicas multilexémicas: ExtractTerm**

### **1. Concepção do ExtractTerm**

#### **1.1. Aplicação do dicionário de tipologias**

O ExtractTerm foi concebido tendo por base conhecimentos previamente adquiridos na observação de textos de especialidade. O seu funcionamento assenta num dicionário de tipologias terminogénicas flexível, que pode ser alterado em qualquer momento do processo. O ExtractTerm funda-se em princípios teóricos deterministas, segundo os quais as condições de existência de um fenómeno são determinadas e fixadas, de tal forma que o fenómeno ocorre de modo obrigatório.

Extraír sequências reconhecidas como colocações, frasemas terminológicos ou fraseologias é o objectivo que nos propomos alcançar com o ExtractTerm, que opera, exclusivamente, sobre textos etiquetados.

Com base nos textos etiquetados, o ExtractTerm aplica as matrizes contidas no dicionário de tipologias (cf. Capítulo IV, ponto 5.4), apresentando o algoritmo correspondente a essa função o seguinte código:

```
for ( iterator_corpus_dividido = corpus_dividido.begin();
iterator_corpus_dividido != corpus_dividido.end();
iterator_corpus_dividido++ )
{
    if ( iterator_corpus_dividido->entrada == "<br>" )
    {
        iterator_corpus_dividido2 =
iterator_corpus_dividido;
```

```

        if ( iterator_corpus_dividido2 !=
corpus_dividido.begin() )
            iterator_corpus_dividido2--;

        while ( iterator_corpus_dividido2->entrada != "<br>"
&& iterator_corpus_dividido2 != corpus_dividido.begin() + 1 )
            iterator_corpus_dividido2--;

        out << "<blockquote>";

        for ( iterator_corpus_dividido3 =
iterator_corpus_dividido2; iterator_corpus_dividido3 !=
iterator_corpus_dividido; iterator_corpus_dividido3++)
        {
            if ( iterator_corpus_dividido3->entrada != ""
)
                out << " " << iterator_corpus_dividido3-
>entrada << " ";
        }

        out << "</blockquote>";
        out << "<hr><p>";
    }

    for ( iterador_tipologias = m_tipologia.begin();
iterador_tipologias != m_tipologia.end(); iterador_tipologias++)
    {
        // Quantas partes tem esta regra
        int partes = iterador_tipologias->size();

        if ( iterator_corpus_dividido->Compara_Etiqueta(
(*iterador_tipologias)[0] ) )
        {
            bool encontrei = true;
            iterator_corpus_dividido1 =
iterator_corpus_dividido;

            dbg = iterator_corpus_dividido1->entrada;
            for ( itEtiquetas =
(*iterator_corpus_dividido1).etiqueta.begin(); itEtiquetas !=
(*iterator_corpus_dividido1).etiqueta.end(); itEtiquetas++ )
            {
                dbg += " <font color=\\"#008080\ ">[";
                dbg += (*itEtiquetas);
                dbg += "]</font> ";
            }

            for ( int i = 1; i < partes ; i++ )
            {
                if ( iterator_corpus_dividido1 !=
corpus_dividido.end() -1 )
                    iterator_corpus_dividido1++;

                if ( !iterator_corpus_dividido1-
>Compara_Etiqueta( (*iterador_tipologias)[i] ) )
                    encontrei = false;
                else
                {

```

```

                                dbg += iterator_corpus_dividido1-
>entrada;
                                for ( itEtiquetas =
(*iterator_corpus_dividido1).etiqueta.begin(); itEtiquetas !=
(*iterator_corpus_dividido1).etiqueta.end(); itEtiquetas++ )
                                {
                                dbg += " <font
color=\\"#008080\\">{";
                                dbg += (*itEtiquetas);
                                dbg += "]</font> ";
                                }
                                }
                                if ( encontrei )
                                {
                                out << dbg << " <font
color=\\"#C0C0C0\\">{";
                                for ( int i = 0; i < partes ; i++ )
                                {
                                out <<
(*iterador_tipologias)[i].etiqueta[0];
                                if ( i < partes -1 )
                                out << " + ";
                                }
                                out << "}</font> <br>";
                                dbg = "";
                                }
                                }
                                }
                                palavras_tratadas++;

                                if ( palavras_tratadas > 400 )
                                {
                                palavras_tratadas = 0;
                                std::cout << ".";
                                }
                                }

```

É o comportamento das formas base, etiquetadas com e que, no âmbito desta dissertação, nos interessa observar.

O *ExtracTerm* procede ao reconhecimento consecutivo de cada forma contida no *corpus*, acompanhada da etiqueta metalinguística <sup>^</sup> ou <, sendo estas duas etiquetas os pontos em que o programa se detém para iniciar o processo subjacente à extracção.

Assim, o extractor inicia a seu percurso detendo-se no primeiro ou que encontrar. Verifica se a sequência de etiquetas que se posiciona à direita do ou corresponde à primeira das 666 sequências matriciais contempladas no dicionário.

Se a pesquisa for profícua, o ExtracTerm extrai a sequência, acompanhada da matriz correspondente à sua estrutura, para um ficheiro HTML. Caso não encontre a estrutura correspondente, passa à regra seguinte, até percorrer as 666 matrizes. Esta operação repete-se, sequencialmente, para todas as etiquetas e de uma pontuação forte a outra pontuação forte.

Assim, ao enunciado,

Teledeteccção em áreas periurbanas: combinação de índices temáticos para localizar a mudança de uso do solo com recurso às imagens digitais SPOT HRV.

é aplicado o EtiqueLex, sendo o formato de saída, como se apresenta de seguida:

Teledeteccção em áreas periurbanas :  
combinação de índices temáticos para  
localizar a mudança de uso [N:m:s] [V:c]] do solo  
com recurso às imagens digitais  
SPOT HRV

Com base no texto etiquetado, o ExtracTerm efectua a seguinte extracção:

1. Teledetecção em áreas {N:f.s + Prep1:em + N:f.pl}
2. áreas periurbanas {N:f.pl + Adj:f.pl}
3. combinação de índices {N:f.s + Prep1:de + N:m.pl}
4. combinação de índices temáticos {N:f.s + Prep1:de + N:m.pl + Adj:m.pl}
5. índices temáticos {N:m.pl + Adj:m.pl}
6. mudança de uso {N:f.s + Prep1:de + N:m.s}
7. uso do solo {N:m.s + Prep2:do:m.s + N:m.s}
8. solo com recurso {N:m.s + Prep1:com + N:m.s}
9. imagens digitais {N:f.pl + Adj:2gen.pl}
10. imagens digitais SPOT {N:f.pl + Adj:2gen.pl + Sigla}
11. imagens digitais SPOT HRV {N:f.pl + Adj:2gen.pl + Sigla + Sigla}
12. SPOT HRV {Sigla + Sigla}

O texto a preto corresponde às sequências extraídas, a informação a azul, entre parêntesis rectos, contém a informação metalinguística e a informação a vermelho, entre chavetas, equivale às regras que o *ExtractTerm* aplicou.

Os resultados obtidos, exemplificados com este pequeno enunciado, permitem-nos avançar com a sua análise, explanando o funcionamento dinâmico do *ExtractTerm*.

## 1.2. Análise dos resultados

Retomamos o exemplo apresentado no ponto 1.1. de forma a comentar os resultados obtidos a partir do funcionamento do *ExtractTerm*.

Para ilustrar o nosso propósito, apagámos parcialmente a informação metalinguística do texto etiquetado, deixando exclusivamente as etiquetas e

Teledeteccção em áreas periurbanas: combinação de índices temáticos para localizar a mudança de uso do solo com recurso às imagens digitais SPOT HRV

O *ExtractTerm* comparou o *corpus* ao dicionário de tipologias. Sempre que à direita do se encontra uma etiqueta que corresponde às sequências registadas no dicionário de tipologia, ele efectua uma extracção. Foi o que sucedeu com com mas não aconteceu com o nome por na tipologia não estar prevista uma sequência

. Para os restantes , o *ExtractTerm* extraiu todas as sequências previstas no dicionário de tipologias.

Vejamos as sequências (2), (3) e (4), que correspondem ao enunciado

(2) combinação de índices {N:f:s + Prep1:de + N:m:p}

(3) combinação de índices temáticos {N:f:s + Prep1:de + N:m:pl + Adj:m:pl}

(4) índices temáticos {N:m:pl + Adj:m:pl}

Para extrair a sequência máxima aplicámos uma regra complexa {N:f:s + Prep1:de + N:m:pl + Adj:m:pl}, que

resulta de duas regras elementares {N:f:s + Prep1:de + N:m:p} e {N:m:pl + Adj:m:pl}.

Nesta situação, verificámos que a regra {N:f:s + Prep1:de + N:m:pl + Adj:m:pl} é dispensável, uma vez que a fraseologia é recuperada com as primeira e terceira regras. O resultado desta opção é a seguinte:

(2) combinação de índices {N:f:s + Prep1:de + N:m:p}

(4) índices temáticos {N:m:pl + Adj:m:pl}

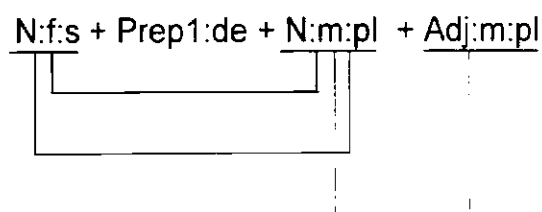
O reconhecimento da fraseologia resulta do facto de o último da primeira regra coincidir com o primeiro da segunda regra, obtendo-se a sequência \_\_\_\_\_ índices \_\_\_\_\_, em que um dos

é eliminado no momento da interpretação do contexto que envolve o \_\_\_\_\_ e que designamos de contexto lexical: «*Although not as common, the lexical context of a unit can be defined as the set of words, terms or candidate terms which appear in the same given dependency relation within a syntactic structure. For example, the lexical context of a noun could be defined as all the distinct nouns that adjoin onto it to make a compound noun*» (Bourigault, D. et alii, 2001:XIII).

Não obstante, preferimos manter as três regras, apesar da redundância manifesta da regra {N:f:s + Prep1:de + N:m:pl + Adj:m:pl} e desta estrutura corresponder, não a uma unidade terminológica multilexémica, mas a uma fraseologia. Em nosso entender, é preferível que o resultado da extracção contenha mais ruído e menos silêncio.

Enquanto a primeira extracção, {N:f:s + Prep1:de + N:m:p}, deve ser ignorada, por corresponder a uma 'sequência livre terminológica', a aplicação da segunda regra provocou a extracção de uma fraseologia composta pelo lexema *orientação científica*, termo genérico de orientação científica, e pelo frasema terminológico *Consideramos*. Assim, a fraseologia um contexto mínimo do frasema terminológico, ou seja, um contexto lexical, porque, em acordo com Bourigault: «[...] *the lexical context of a unit is the set of words which occur in the same dependency relation with it*» (Bourigault, D. et alii, 2001:XVI).

A composição da fraseologia pode ser representada desta forma:



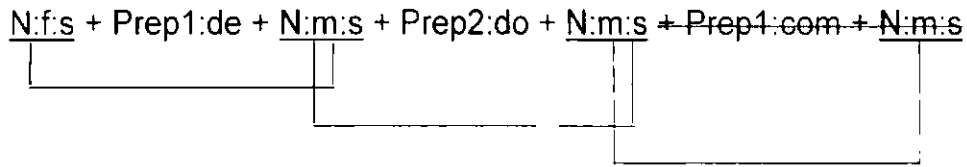
Das três sequências que se seguem:

- (5) mudança de uso {N:f:s + Prep1:de + N:m:s}
- (6) uso do solo {N:m:s + Prep2:do:m:s + N:m:s}
- (7) solo com recurso {N:m:s + Prep1:com + N:m:s}

a sequência (6) e (7) são excluídas de *per se*, na medida em que as classificamos como 'sequências livres terminológicas'.

No entanto, a sobreposição da matriz aplicada em (5) com a matriz aplicada em (6) permite-nos chegar a uma interpretação próxima da anterior:

é uma fraseologia da qual se pode extrair o termo



Neste caso, a primeira regra retém  $N:f:s + Prep1:de + N:m:s$ , a segunda regra retém a sequência  $Prep2:do + N:m:s$ . É a combinação de duas regras que nos permite isolar a unidade terminológica, sendo o termo  $N:m:s$ , o  $N:m:s$  de ligação entre as duas regras:

$N:m:s +$

As sequências  $N:f:s + Prep1:de + N:m:s$  e  $Prep2:do + N:m:s$  são 'sequências livres terminológicas', por a sequencialidade dos seus lexemas corresponder a um corte sintáctico pertinente no texto, não havendo, no entanto, combinações lexicais que os solidarizem.  $N:m:s + Prep1:com + N:m:s$  e  $Prep1:com + N:m:s$  têm ambos frequência 1 e ocorrem em contextos em que os seus especificadores, que se apresentam à sua direita, são omissos, por terem sido amplamente referenciados antes.

O mesmo não acontece com  $N:m:s + Prep1:com + N:m:s$ . Neste caso, a estrutura  $N:m:s + Prep1:com + N:m:s$  ocasiona uma cisão errónea de duas sequências sintácticas coesas: a unidade terminológica  $N:m:s$  e a locução preposicional  $Prep1:com + N:m:s$ .

Tal facto sucedeu, por o `ExtractTerm` se deter em  $N:m:s$  e constatar que a expansão à direita corresponde a uma matriz pré-registada. A falha deve-se, essencialmente, a dois factores.

O primeiro, relaciona-se com o facto de termos incluído a preposição na lista dos gramemas pertinentes para constituição de unidades terminológicas multilexémicas. Da observação dos resultados, verificámos que ela não é uma preposição constituinte, mas é uma preposição cuja função é a de associar dois frasemas, como podemos comprovar com os exemplos seguintes:

com  
com  
com  
com

A preposição tem a função de associar no eixo sintagmático dois termos dotados ambos de autonomia semântica.

O segundo factor, diz respeito à não inclusão da locução no dicionário das locuções, o que levou o EtiqueLex a etiquetar

agindo, desta feita, o ExtracTerm sobre a locução preposicional como se de uma sequência livre se tratasse.

Em resultado desta análise, acrescentámos a locução no dicionário de locuções e eliminámos as matrizes que continham a etiqueta na medida em que o ruído provocado por esta matriz é superior ao silêncio provocado pela sua ausência.

As quatro últimas sequências são exemplos de uma extracção fina, bem sucedida:

- (9) imagens digitais {N:f:pl + Adj:2gen:pl}
- (10) imagens digitais SPOT {N:f:pl + Adj:2gen:pl + Sigla}
- (11) imagens digitais SPOT HRV {N:f:pl + Adj:2gen:pl + Sigla + Sigla}
- (12) SPOT HRV {Sigla + Sigla}

Neste caso, os quatro exemplos correspondem a quatro unidades terminológicas multilexémicas, uma vez que correspondem a quatro denominações distintas, sendo que as expansões à direita de (10) e (11) vêm especificar o tipo das , tendo, para estes casos, o ExtractTerm funcionado de acordo com as suas máximas potencialidades.

[[[imagens digitais ] [[SPOT ] HRV ] ] ]

Se entendermos que desempenha a função de um nome, podemos, por extensão, considerar que e desempenham a função de 'siglas epítetos' ao mesmo nível dos substantivos epítetos: *«Les substantifs épithètes, s'ils passent à un statut d'adjectif, le font donc toujours par le truchement de la fonction épithète. C'est en s'attachant directement à un substantif recteur qu'ils se déchargent de leur poids référentiel et deviennent progressivement des indicateurs de propriétés»* (Noailly, 1999:24).

Para o especialista, o adjetivo pode ser omissa, sem que a ambiguidade conceptual se estabeleça. De facto, todas as imagens de satélite são, actualmente, digitais (forma de registo dos dados), podendo ocorrer o

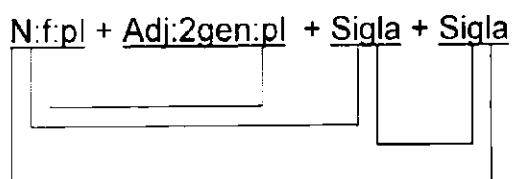
termo \_\_\_\_\_, sendo que a sigla \_\_\_\_\_, nome do satélite francês, qualifica o nome \_\_\_\_\_ ou o nome \_\_\_\_\_.

Por sua vez, a sigla \_\_\_\_\_, que significa *Haute Résolution dans le Visible*, é um sistema de aquisição de imagem inerente ao \_\_\_\_\_ o que nos permite inferir que a sigla \_\_\_\_\_ desempenha a função de um adjectivo qualificativo, tendo conseqüentemente, de ser subcategorizado por um nome ou por outro elemento que desempenhe a função de um nome, sendo este, na realidade, o caso da sigla \_\_\_\_\_.

A sequência \_\_\_\_\_ é uma unidade terminológica e pode, conseqüentemente, ser classificada como um nome, o que nos permite atribuir-lhe a função de nome epíteto, ao mesmo nível da sigla \_\_\_\_\_, uma vez que, inequivocamente, qualifica o termo \_\_\_\_\_.

A extracção correcta destes quatro exemplos leva-nos a concluir que as regras têm em determinadas condições uma boa aplicabilidade, o que nos induz a melhorá-las.

Nestes últimos exemplos apresentados, as regras permitiram extrair todos as unidades multilexémicas da sequência máxima, partindo do menos complexo \_\_\_\_\_, para atingir o mais complexo, \_\_\_\_\_ e \_\_\_\_\_, voltando, novamente, ao menos complexo, \_\_\_\_\_, ficando assim exploradas todas as hipóteses deste contexto lexical:



Uma das características do ExtracTerm é podermos visualizar todos os momentos de um processo que resultam do encaixe sucessivo de matrizes, permitindo-nos decidir, de entre o leque apresentado, quais as sequências a reter. Neste caso, o encaixe decorreu de um processo morfológico. Com excepção do último termo, este processo pode ser considerado um processo de sobrecomposição, na medida em que cada expansão à direita, permite a especificação de um conceito, o que em termos linguísticos resulta numa única denominação.

Com o ExtracTerm extraímos:

- i. colocações
- ii. frases terminológicas
- iii. fraseologias

No entanto, também extraímos sequências menos interessantes tais como 'sequências livres terminológicas' e correntes, unidades multilexémicas correntes, bem como sequências resultantes de cortes mal efectuados.

Para que os resultados tenham uma melhor resolução é necessário melhorar a qualidade da matéria prima do extractor: o texto etiquetado.

Se o ExtracTerm funcionar sobre informação metalinguística desambiguada, as hipóteses de erro são mais reduzidas. Pretendemos, assim criar regras de desambiguação, exclusivamente aplicadas a determinadas classes gramaticais.

## 2. Regras de desambiguação

O EtiqueLex atribui a cada forma todos os funções gramaticais que esta pode assumir nos diferentes contextos, podendo uma forma estar marcada por diversas etiquetas metalinguísticas. Esta etapa equivale à primeira fase de procedimento do ExtracTerm, que foi concebido para extrair informação tendo por base textos etiquetados. O EtiqueLex tem, assim, a particularidade de funcionar como ferramenta autónoma ou, ainda, como ferramenta integrada do ExtracTerm.

Após a etiquetagem efectuada, o ExtracTerm parte do dicionário de tipologias, que contém as estruturas que nos interessa identificar no *corpus*. Quando uma forma está duplamente etiquetada, o ExtracTerm extrai duas vezes a mesma sequência, caso esta corresponda a uma das matrizes contida no dicionário de tipologias.

Assim, para a sequência *o* *o*, em que *o* é acompanhada da etiqueta *N:f:s* e da etiqueta *N:f:s*, o ExtracTerm aplica duas regras de reconhecimento: {N:f:s + N:f:s} e {N:f:s + Adj:f:s} extraíndo, desnecessariamente, duas vezes a mesma sequência, estando neste contexto, a classificação *N:f:s* errada para o lexema *o*.

Todavia, esta classificação desacertada não afecta o resultado da extracção, na medida em que ao especialista só interessa saber se a realidade linguística extraída é ou não uma denominação, independentemente da classificação gramatical atribuída a cada um dos seus elementos constituintes.

[Pron:dem:3p.f.s] De entre estas classes gramaticais, só nos interessa distinguir a preposição, uma vez que esta é uma das classes que consta no dicionário de regras de reconhecimento.

Neste *corpus*, a forma *a* tem frequência 7762, não nos sendo possível saber quantas vezes ocorre cada uma das classes gramaticais. A primeira consequência contraproducente desta indeterminação é desconhecermos a posição diferencial que cada uma das classes ocupa no eixo sintagmático, sendo difícil a avaliação da sua importância na constituição das unidades terminológicas multilexémicas. Deste desconhecimento decorre a dúvida sobre a pertinência da elaboração de regras de reconhecimento que contemplem a etiqueta [Prep1:a].

Só depois da aplicação das regras de desambiguação é que iremos decidir se a estrutura  $N^* + \text{Prep:1a} + N^*$ , assim como as suas combinatórias, são regras enriquecedoras para a extracção automática ou se, pelo, contrário, são geradoras de ruído.

Tomemos como exemplo o enunciado seguinte:

[...] combinação [N:f.s] automática [Adj:f.s] de [Prep1:de] índices [N:m:pl] temáticos [Adj:m:pl] calculados [Adj:m:pl] [Pp:m:pl] a [Art:f.s] [Pron:pess:f.s] [Pron:dem:f.s] [Prep1:a] partir [V:inf] de [Prep1:de] [...].

Neste enunciado ocorrem duas formas, às quais são atribuídas mais do que uma etiqueta calculados [Adj:m:pl] [Pp] e a [Art:f.3p.f.s] [Pron:pess:3p.f.s] [Pron:dem:3p.f.s] [Prep1:a]. Embora a distinção entre participio passado verbal e participio passado adjectival seja tratada no ponto 2.2, importa desde já

No entanto, a dupla extracção não tem qualquer utilidade: sobrecarrega os resultados, aumentando, naturalmente, o volume de informação a descodificar.

Se neste contexto, o resultado não é afectado, outras situações há em que tal sucede. No texto:

[...] permite [V.cj] a [Art.f:s] [Pron.poss.3p:f:s] [Pron.dem:3p:f:s] [Prep1:a] extracção [N:f:s] dos [Prep2:do:m:s] níveis [N:m:pl] de [Prep1:de] reflectância [N:f:s] do [Prep2:do:m:s] coberto [Adj:m:s] [Pp:m:s] vegetal [Adj.2gen:s] [N:m:s]. Contudo [Conj], os [Art.def:m:pl] [Pron.poss] [Pron.dem] resultados [N:m:pl] obtidos [Pp:m:pl] [Adj:m:pl] se\_bem\_que [Loc] elucidativos [Adj:m:pl] quanto\_aos [Loc] níveis [N:m:pl] de [Prep1:de] actividade [N:f:s] clorofilina [Adj:f:s]. [...]

identificámos diversas situações de etiquetas erradas, que influenciaram negativamente a extracção:

- (1) níveis [N:m:pl] de [Prep1:de] reflectância [N:f:s] do [Prep2:do:m:s] coberto [Adj:m:s] [Pp:m:s] vegetal [Adj.2gen:s] [N:m:s].
- (2) resultados [N:m:pl] obtidos [Pp:m:pl] [Adj:m:pl]

No primeiro exemplo, o lexema *coberto* está duplamente mal etiquetado. Na construção do dicionário de formas flexionadas, a forma *coberto* não foi marcada com a etiqueta [N], porque descontextualizada não a reconhecemos como um lexema da especialidade.

O efeito imediato desta etiquetagem errada reflectiu-se na incapacidade do ExtracTerm extrair a unidade terminológica multilexémica *coberto vegetal*, por a etiquetagem atribuída, automaticamente, a esta sequência, não corresponder a nenhuma matriz pré-determinada.

Também o lexema *vegetal* está duplamente etiquetado, mas de forma correcta, embora, neste contexto, a etiqueta [N] seja inadequada, uma vez que se trata, inequivocamente, de um [Adj].

No caso do último exemplo, a sequência *informação obtida* é o resultado da aplicação de uma regra elementar {N:m:pl + Adj:m:pl}. A etiqueta [Pp] é ignorada, uma vez que não faz parte das classes constituintes mestras das unidades terminológicas multilexémicas retidas e, deste modo, obtém-se uma sequência livre de lexemas. Nesta fase, o ExtracTerm ainda não detém a informação necessária para distinguir um particípio passado com valor verbal de um particípio passado com valor adjectival, o que resulta numa extracção não produtiva.

O nosso objectivo é o de aumentar ao máximo a correcção linguística, para diminuir o erro na extracção. Para esse efeito, é imprescindível reduzir a ambiguidade provocada pela atribuição de múltiplas etiquetas a uma mesma forma, porque quanto mais monoetiquetadas forem as formas, mais probabilidades o ExtracTerm tem de melhorar a qualidade da extracção.

Desta feita, é preciso criar regras que permitam levantar as ambiguidades entre as diferentes classes gramaticais.

Para levar a cabo tal procedimento, tivemos que deliberar sobre que classes gramaticais devem incidir as regras de desambiguação e em que momento aplicá-las.

Assim, optámos por associar as regras de desambiguação ao processo da etiquetagem. O EtiqueLex identifica os contextos lexicais das classes gramaticais a desambiguar e aplica as regras, de maneira a conceder uma única etiqueta a cada forma para um dado contexto.

A partir do momento em que o *corpus* esteja, parcialmente, desambiguado, o ExtracTerm tem menos etiquetas a percorrer, uma vez que as formas monoetiquetadas provocam uma quebra do número de combinatórias possíveis, reduzindo ainda o seu tempo de pesquisa.

As regras de desambiguação foram elaboradas a partir de uma primeira extracção (cf. Anexo 2) que fizemos sobre *corpus* não desambiguado (cf. Anexo1). Dessa observação, constatámos que os conjuntos de classes gramaticais sobre as quais é pertinente criar regras são:

- participios passados pós-nominais e adjectivos
- artigos, pronomes e preposições
- nomes e adjectivos.

## 2.1. Preposição, artigo definido feminino singular e pronomes

No dicionário de formas flexionadas, à forma *a* estão associadas quatro etiquetas metalinguísticas: [Frec.1.a] [Art.def.f.s], [Pron.pess.3p:f.s],

[Pron:dem:3p:f:s]. De entre estas classes gramaticais, só nos interessa distinguir a preposição, uma vez que esta é uma das classes que consta no dicionário de regras de reconhecimento.

Neste *corpus*, a forma *a* tem frequência 7762, não nos sendo possível saber quantas vezes ocorre cada uma das classes gramaticais. A primeira consequência contraproducente desta indeterminação é desconhecermos a posição diferencial que cada uma das classes ocupa no eixo sintagmático, sendo difícil a avaliação da sua importância na constituição das unidades terminológicas multilexémicas. Deste desconhecimento decorre a dúvida sobre a pertinência da elaboração de regras de reconhecimento que contemplem a etiqueta [Prep1:a].

Só depois da aplicação das regras de desambiguação é que iremos decidir se a estrutura  $N^* + \text{Prep:1a} + N^*$ , assim como as suas combinatórias, são regras enriquecedoras para a extracção automática ou se, pelo, contrário, são geradoras de ruído.

Tomemos como exemplo o enunciado seguinte:

[...] combinação [N:f:s] automática [Adj:f:s] de [Prep1:de] índices [N:m:pl] temáticos [Adj:m:pl] calculados [Adj:m:pl] [Pp:m:pl] a [Art:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] partir [V:inf] de [Prep1:de] [...].

Neste enunciado ocorrem duas formas, às quais são atribuídas mais do que uma etiqueta *calculados* [Adj:m:pl] [Pp] e a [Art:f:3p:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a]. Embora a distinção entre participio passado verbal e participio passado adjectival seja tratada no ponto 2.2, importa desde já

referir que a identificação inequívoca da preposição *a* em determinados contextos é primordial para a elaboração das regras de desambiguação entre outros conjuntos de classes gramaticais.

Com base na regra de reconhecimento {N:m:pl + Adj:m:pl + Adj:m:pl}, o *ExtracTerm* extrai a sequência *índices temáticos calculados*, uma vez que a regra não contempla a etiqueta [Pp]. O facto de a classe gramatical que está à direita da forma *calculados* ser uma forma multietiquetada dificulta a criação de uma regra de desambiguação para distinguir o [Pp] do [Adj]. Deste modo, julgamos pertinente, a criação de uma regra de desambiguação que permita distinguir a preposição *a* das outras classes gramaticais, de maneira a criar regras para o conjunto de classes gramaticais seguintes.

### 2.1.1. Regras linguísticas de aprendizagem

As regras linguísticas de aprendizagem são construídas a partir da análise do contexto lexical da forma a desambiguar. Sempre que o *Extracterm* identifica a classe gramatical a desambiguar, ele pesquisa à esquerda e à direita da forma e aplica as regras, optando pela etiqueta adequada ao contexto em questão.

As regras podem, assim, ser construídas pela positiva. Quando se realiza a condição A e a condição B, obtém-se a situação C : [+A] + [+B] = [+C]. Em determinados contextos é mais produtivo construir regras pela negativa. Deste modo, quando se realiza a condição A e a condição B, não se obtém a situação C : [+A] + [+B] = [-C]:

## Regra 1

- $[Art:def:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] + [N:f:s] \rightarrow [Art:def:f:s] + [N:f:s]$

*Leitura da Regra 1:*

Sempre que a forma *a* é etiquetada com  $[Art:def:f:s]$   $[Pron:pess:f:s]$   $[Pron:dem:f:s]$   $[Prep1:a]$  e a classe gramatical que se encontra à sua *direita* é acompanhada da etiqueta  $[N:*)$ , então o `ExtractTerm` selecciona a etiqueta  $[Art:def:f:s]$  em detrimento das outras.

Exemplo: a  $[Art:def:f:s]$  imagem  $[N:f:s]$

## Regra 2

- $[Ger] + [Art:def:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] \rightarrow [Ger] + [Art:def:f:s]$

*Leitura da Regra 2:*

Sempre que a forma *a* é etiquetada com  $[Art:def:f:s]$   $[Pron:pess:f:s]$   $[Pron:dem:f:s]$   $[Prep1:a]$  e a classe gramatical que se encontra à sua *direita* é acompanhada da etiqueta  $[Ger]$ , então o `ExtractTerm` selecciona a etiqueta  $[Art:def:f:s]$  em detrimento das outras.

Exemplo: [...] facilitando  $[Ger]$  a  $[Art:def:f:s]$  sua leitura [...]

## Regra 3

- $[Pp] + [Art:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] + [Pron:poss:*] \rightarrow [Pp] + [Prep1:a] + [Pron:poss*]$

*Leitura da Regra 3:*

Sempre que a forma *a* é etiquetada com [Art:def:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] e a classe gramatical que se encontra à sua esquerda é acompanhada da etiqueta [Pp], e a classe gramatical que se encontra à sua direita é [Pron:poss:\*], então o *ExtracTerm* selecciona a etiqueta [Art:def:f] em detrimento das outras.

Exemplo: a [Art:def:f:s] sua [Pron:poss:3p:f:s] ocupação [N:f:s] temática [Adj:f:s]

#### Regra 4

- [N\*] + [Art:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] + [V:inf] → [Prep1:a] + [V:inf]

*Leitura da Regra 4:*

Sempre que a forma *a* é etiquetada com [Art:def:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] e a classe gramatical que se encontra à sua direita é acompanhada da etiqueta [V:inf], então o *ExtracTerm* selecciona a etiqueta [Prep1:a] em detrimento das outras.

Exemplo: [...] satélites [N:m:pl] destinados [Adj:m:s] [Pp] a [Prep1:a] facilitar [V:inf][...]

#### Regra 5

- [Art:f:s] [Pron:pess:f:s] [Pron:dem:f:s] [Prep1:a] + [Art:indef:f:s] → [Prep1:a] + [Art:indef:f:s]

*Leitura da Regra 5:*

Sempre que a forma *a* é etiquetada com [Art.def:f.s] [Pron.pess:f.s] [Pron.aem:f.s] [Pleb1:a] e a classe gramatical que se encontra à sua direita é acompanhada da etiqueta [Art.indef:f.s], então o ExtractTerm selecciona a etiqueta [Prep1:a] em detrimento das outras.

Exemplo: nebulosidade [N:f.s] ligada [Adj] [Pp] a [Prep1:a] uma [Art.indef:f.s] depressão [N:f.s] de [Prep1:de] altitude [N:f.s]

Com os resultados obtidos com a aplicação destas cinco regras de aprendizagem, que permitem desambiguar a forma *a* em cinco contextos diferentes, pensamos poder testar a matriz  $N^* \rightarrow Prep1:a \leftarrow N^*$  (cf. ponto 2.5.)

## 2.2. Adjectivo e participio passado pós-nominal

O adjectivo é, a par do nome, a classe gramatical que nos interessa identificar, na medida em que inclui as matrizes que constituem o dicionário de tipologias.

A distinção entre adjectivo verbal e participio passado com valor verbal impõe-se.

O participio passado é uma forma verbal que participa da natureza do verbo e do adjectivo, sendo denominado participio passado verbal no primeiro caso e participio passado adjectival, no segundo; ambas as formas do participio passado são variáveis em número e em género.

Os participios passados verbais caracterizam-se por não poder, regra geral, desempenhar a função de adjectivos e contribuírem para exprimir valores

temporais e aspectuais. Podem, por um lado, participar dos tempos compostos, por outro, podem ocorrer isoladamente, exprimindo deste modo uma acção temporal.

Formados a partir de verbos, os participios passados adjectivais cumprem a função de adjectivos qualificativos, em posição atributiva ou predicativa, exprimindo propriedade.

De acordo com os objectivos preconizados no nosso estudo, interessamos exclusivamente, observar o comportamento dos participios passados que ocorrem em posição pós-nominal, por terem um comportamento idêntico aos dos adjectivos epítetos.

Relembramos que o EtiqueLex atribui a dupla etiqueta [Pp] e [Adj] a formas como *digitalização colorido assistida*. No momento da desambiguação automática, a nossa opção consiste em atribuir a etiqueta [Pp] aos participios passados verbais e a etiqueta [Adj] aos participios passados adjectivais, uma vez que estes últimos têm uma função qualificadora, ao mesmo nível dos adjectivos qualificativos.

Para o tratamento automático das unidades terminológicas multilexémicas, a distinção entre participio passado com valor verbal e participio passado com valor adjectival passa pela análise do contexto lexical que antecede e sucede o participio passado.

Se o participio passado é construído com um complemento, introduzido por uma preposição, considera-se que o participio passado tem um valor verbal e, por isso, é etiquetado com a etiqueta [Pp]. Caso o participio passado pós-

nominal seja construído sem complemento, ele assume valor adjectival e, consequentemente, deve ser etiquetado com a etiqueta [Adj].

Perante a aplicação desta regra de aprendizagem, o EtiqueLex opta, exclusivamente, por atribuir uma das etiquetas à respectiva forma. Desta maneira, a informação metalinguística acrescentada ao *corpus* é simplificada, aumentando a precisão e a correcção da etiquetagem.

As regras de aprendizagem têm como consequência imediata um melhoramento da qualidade dos dados obtidos com o *ExtractTerm*, uma vez que este se depara com menos etiquetas, aplicando as regras de reconhecimento sobre formas monoetiquetadas, reduzindo as hipóteses de extracção erradas ou de duplas extracções. Por não se deter na etiqueta [Pp], o *ExtractTerm* reduz o tempo de pesquisa.

### 2.2.1. Regras linguísticas de aprendizagem

Para o texto não etiquetado,

A informação obtida sobre o uso do solo, por segmentação e interpretação de fotografia aérea, é extraordinariamente precisa e exaustiva e a unidade de análise, a zona urbana homogénea, responde satisfatoriamente à abordagem do crescimento urbano das periferias.

obtém-se a seguinte etiquetagem:

A [Art.f.s] [Pron.pess:3p.f.s] [Pron.dem:3p.f.s] [Prep1:a] informação [N:f.s]  
obtida [Pp.f.s] [Aq.f.s] sobre [Prep1:sobre] o [Art.def:m.s] [Pron.pess:3p.m.s]  
[Pron.dem:3p.m.s] uso [N:m.s] [V:cj] do [Prep2:do.m.s] solo [N:m.s], por  
[Prep1:por] segmentação [N:f.s] e [Conj:e] interpretação [N:f.s] de [Prep1:de]

fotografia [N:f:s] aérea [Adj:f:s], é [V:cj] extraordinariamente [Adv] precisa [Adj:f:s] [V:cj] e [Conj:e] exaustiva [Adj:f:s] e [Conj:e] a [Art:f:s] [Pron.pess.3p:f:s] [Pronidem.3p:f:s] [Prep1:a] unidade [N:f:s] de [Prep1:de] análise [N:f:s], a [Art:f:s] [Pron.pess.3p:f:s] [Pronidem.3p:f:s] [Prep1:a] zona [N:f:s] urbana [Adj:f:s] homogênea [Adj:f:s], responde [V:cj] satisfatoriamente [Adv] à [Prep2:à:f:s] abordagem [N:f:s] do [Prep2:do:m:s] crescimento [N:m:s] urbano [Adj:m:s] das [Prep2:das:f:pl] periferias [N:f:pl].

Com base na regra de reconhecimento {N:f:s + Adj:f:s} o *ExtractTerm* extrai a sequência *informação [N:f:s] obtida [Pp:f:s] [Adj:f:s]*, que não corresponde a nenhuma realidade linguística que nos interesse reter.

Após a aplicação da regra de aprendizagem, o *EtiqueLex* opta por marcar a forma *obtida* com a etiqueta [Pp], uma vez que a sequência N:\* + Pp:\* não é contemplada no dicionário de tipologias, levando o *ExtractTerm* a ignorar todas as sequências que obedecem a esta estrutura tipo.

Com base no que expusemos, criámos as seguintes regras de aprendizagem:

### Regra 1

- [N:\*] + [Adj:\*] [Pp:\*] + [Prep1:a] → [N:\*] + [Pp:\*] + [Prep1:a]
- [N:\*] + [Adj:\*] [Pp:\*] + [Prep1:após] → [N:\*] + [Pp:\*] + [Prep1:após]
- [N:\*] + [Adj:\*] [Pp:\*] + [Prep1:até] → [N:\*] + [Pp:\*] + [Prep1:até]
- [N:\*] + [Adj:\*] [Pp:\*] + [Prep1:conforme] → [N:\*] + [Pp:\*] + [Prep1:conforme]
- [N:\*] + [Adj:\*] [Pp:\*] + [Prep1:consoante] → [N:\*] + [Pp:\*] + [Prep1:consoante]
- [N:\*] + [Adj:\*] [Pp:\*] + [Prep1:contra] → [N:\*] + [Pp:\*] + [Prep1:contra]





- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $[N] \rightarrow [N]$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $[N] \rightarrow [N]$
- $\rightarrow$

*Leitura da Regra 2:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $[N]$  e a classe gramatical que se encontra à sua esquerda é um  $N$  e à esquerda deste um  $N$  e a classe gramatical que se encontra à sua direita é um  $N$ .

então o `ExtractTerm` selecciona a etiqueta `NP` em detrimento da outra.

Exemplo:

[...]

### Regra 3

- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`
- `NP` → `NP`

#### *Leitura da Regra 3:*

Sempre que uma forma é etiquetada com a dupla etiqueta `NP NP` e a classe gramatical que se encontra à sua esquerda é um `NP` e a classe gramatical que se encontra à sua direita é um `NP`, então o `ExtractTerm` selecciona a etiqueta `NP` em detrimento da outra.

Exemplo (1):

Exemplo (2):

#### Regra 4

- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$
- $\rightarrow$

- $\text{NP} \rightarrow \text{NP} \text{ VP}$

- $\text{VP} \rightarrow \text{VP} \text{ NP}$

*Leitura da Regra 4:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\text{NP} \text{ VP}$  e a classe gramatical que se encontra à sua esquerda é um  $\text{VP}$  e à esquerda deste um  $\text{NP}$  e a classe gramatical que se encontra à sua direita é um  $\text{NP}$  então o `ExtracTerm` selecciona a etiqueta  $\text{NP}$  em detrimento da outra.

Exemplo:

$\text{NP}$   
 $\text{VP}$   
 $[\dots]$

**Regra 5**

- $\text{NP} \rightarrow \text{NP} \text{ VP}$

*Leitura da Regra 5:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\text{NP} \text{ VP}$  e a classe gramatical que se encontra à sua esquerda é um  $\text{NP}$  e a classe gramatical que se encontra à sua direita é um  $\text{VP}$ , então o `ExtracTerm` selecciona a etiqueta  $\text{VP}$  em detrimento da outra

Exemplo: $[\dots]$   $\text{NP}$   $\text{VP}$   $[\dots]$

## Regra 6

- $\text{NP} \rightarrow \text{NP} \text{ PP}$

*Leitura da Regra 6:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\text{NP}$  e a classe gramatical que se encontra à sua direita é uma  $\text{PP}$  e à direita desta se encontre um  $\text{NP}$ , então o `ExtractTerm` selecciona a etiqueta  $\text{NP}$  em detrimento da outra.

Exemplo: [...]

[...].

## Regra 7

- $\text{NP} \rightarrow \text{PP} \text{ NP}$

*Leitura da Regra 7:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\text{NP}$  e a classe gramatical que se encontra à sua esquerda é um  $\text{PP}$ , então o `ExtractTerm` selecciona a etiqueta  $\text{NP}$  em detrimento da outra.

Exemplo: [...].

## Regra 8

- $\text{NP} \rightarrow \text{NP} \text{ NP}$

*Leitura da Regra 8 :*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\text{NP}$  e a classe gramatical que se encontra à esquerda é um  $\text{NP}$ , então o `ExtractTerm` selecciona a etiqueta  $\text{NP}$  em detrimento da outra.

Exemplo:

### Regra 9

- 

*Leitura da Regra 9:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\langle \text{N} \text{ NP} \rangle$  e a classe gramatical que se encontra à esquerda é um  $\langle \text{NP} \rangle$ , e a classe gramatical que se encontra à sua direita é um  $\langle \text{NP} \rangle$ , então o `ExtractTerm` selecciona a etiqueta  $\langle \text{NP} \rangle$  em detrimento da outra.

Exemplo:

`NP <NP> <NP> <NP>`  
`<NP> <NP> <NP> <NP>`

### Regra 10

- 

*Leitura da Regra 10:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\langle \text{N} \text{ VP} \rangle$  e a classe gramatical que se encontra à sua esquerda é um  $\langle \text{VP} \rangle$ , e a classe gramatical que se encontra à sua direita é um  $\langle \text{VP} \rangle$ , então o `ExtractTerm` selecciona a etiqueta  $\langle \text{VP} \rangle$  em detrimento da outra.

Exemplo:

`VP <VP> <VP> <VP>`  
`<VP> <VP> <VP> <VP>`

### 2.3. Nome e adjectivo

O nome e o adjectivo são duas classes gramaticais que mantêm relações muito estreitas entre si, sendo frequente a concentração numa única forma das duas classes de palavras. Só a observação do contexto lexical permite a identificação da classe gramatical do lexema, tendo em conta a sua função sintáctica no enunciado. Desde modo, o nome pode desempenhar a função de adjectivo e o adjectivo a função de nome.

Vejamos o caso do lexema *estatístico* que, no dicionário de formas flexionadas, foi marcado com duas etiquetas *estatístico* e *estatística*.

O nome *estatístico* é um termo que dá conta «*da ciência que tem por objecto elaborar tais estatísticas, determinar correlações que apresentem os seus dados e tirar delas as suas consequências para descrição e explicação do que passou e previsão e organização do futuro*» (Porto Editora, 8ª Edição). O adjectivo correspondente, na sua forma feminina, é grafado da mesma maneira, sendo, do ponto de vista morfológico, um homónimo: *estatística*. A paridade das duas formas faz com que só o contexto lexical nos permita identificar claramente a respectiva classe gramatical, mediante a intervenção de uma análise sintáctica. No exemplo *estatística*, o lexema *estatístico* desempenha, inequivocamente, a função de um adjectivo epíteto relacional, por se encontrar em situação pós-nominal, indicando uma relação com o referente do nome do qual deriva.

Esta questão não pode, no entanto, ser confundida com as seqüências cuja estruturas apresentam ordem gramatical *estatística*. Em *estatística*

o substantivo `adjectivo` é um substantivo epíteto, que tem o estatuto de adjectivo relacional, até porque, neste contexto, o substantivo `adjectivo` pode ser substituído pelo adjectivo `relacional`.

Assim, `adjectivo` tem uma variante ortográfica `adjetivo` e uma variante gráfica `adjetivo`, sendo `adjetivo` um sinónimo das três variantes.

No entanto, para o `ExtracTerm` esta situação não é ambígua, uma vez que à forma `adjetivo` corresponde exclusivamente a uma única etiqueta `adjetivo`, apesar de tal forma poder assumir a função de `adjectivo`.

Outras formas há que estão lexicograficamente registadas com duas etiquetas metalinguísticas distintas, por corresponderem a definições distintas.

É o caso, por exemplo, do lexema `variavel`, adjectivo que significa «*que varia ou pode variar*» e do substantivo `letra`, termo da Matemática que se define como: «*(mat.) símbolo - geralmente uma letra - com que se designa qualquer dos elementos de um conjunto, (domínio da variável)*» (Porto Editora, 8ª Edição). No seio de cada uma das classes gramaticais, o lexema é polissémico.

Para o funcionamento do `ExtracTerm`, só construímos regras de aprendizagem que nos permitem distinguir, em alguns contextos, o adjectivo do substantivo, não nos detendo na distinção das polissemias no seio de cada classe gramatical.

### 2.3.1. Regras linguísticas de aprendizagem

Tanto o nome como o adjectivo pós-nominal são as duas classes gramaticais plenas, que constam da tipologia de regras de reconhecimento. Como tal, é importante que elas sejam, tanto quanto possível, monoetiquetadas para que o ExtractTerm actue correctamente sobre o maior número possível de lexemas.

#### Regra 1

- $\langle \text{NP} \text{ NP} \rangle \rightarrow \langle \text{Prep} \text{ NP} \rangle$

*Leitura da Regra 1:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\langle \text{NP} \text{ NP} \rangle$  e a classe gramatical que se encontra à sua esquerda é uma  $\langle \text{Prep} \rangle$ , então o Extracterm selecciona a etiqueta  $\langle \text{Prep} \rangle$  em detrimento da outra.

Exemplo: [...]  $\langle \text{NP} \text{ NP} \rangle$

$\langle \text{Prep} \rangle$   $\langle \text{NP} \text{ NP} \rangle$  [...]  $\langle \text{NP} \text{ NP} \rangle$

#### Regra 2

- $\langle \text{NP} \text{ NP} \rangle \rightarrow \langle \text{NP} \text{ NP} \rangle$

*Leitura da Regra 2:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\langle \text{NP} \text{ NP} \rangle$  e a classe gramatical que se encontra à sua esquerda é uma  $\langle \text{NP} \rangle$ , então o Extracterm selecciona a etiqueta  $\langle \text{NP} \text{ NP} \rangle$  em detrimento da outra.

Exemplo:  $\langle \text{NP} \rangle$   $\langle \text{NP} \text{ NP} \rangle$  [...]  $\langle \text{NP} \text{ NP} \rangle$

### Regra 3

- $\langle \text{NP} \text{ VP} \rangle \rightarrow \langle \text{NP} \text{ VP} \rangle$

*Leitura da Regra 3:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\langle \text{NP} \text{ VP} \rangle$  e a classe gramatical que se encontra à sua esquerda é uma  $\langle \text{NP} \rangle$ , então o Extracterm selecciona a etiqueta  $\langle \text{NP} \rangle$  em detrimento da outra.

Exemplo:

### Regra 4

- $\langle \text{NP} \text{ VP} \rangle \rightarrow \langle \text{NP} \text{ VP} \rangle$

*Leitura da Regra 3:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\langle \text{NP} \text{ VP} \rangle$  e a classe gramatical que se encontra à sua esquerda é uma  $\langle \text{NP} \rangle$ , então o Extracterm selecciona a etiqueta  $\langle \text{NP} \rangle$  em detrimento da outra.

Exemplo:

### Regra 5

- $\langle \text{NP} \text{ VP} \rangle \rightarrow \langle \text{NP} \text{ VP} \rangle$

*Leitura da Regra 5:*

Sempre que uma forma é etiquetada com a dupla etiqueta  $\langle \text{NP} \text{ VP} \rangle$  e a classe gramatical que se encontra à sua esquerda é um  $\langle \text{NP} \rangle$  e a classe que se encontra à sua direita é um  $\langle \text{VP} \rangle$ , então o Extracterm selecciona a etiqueta  $\langle \text{VP} \rangle$  em detrimento da outra.

Exemplo:

### 3. **ExtracTerm: utilização das regras de aprendizagem**

A partir das regras de aprendizagem constituímos um dicionário de regras de desambiguação, que o ExtracTerm aplica logo após a primeira multietiquetagem. De ora em diante, deixamos de nos referir ao EtiqueLex, para só referirmo-nos ao ExtracTerm, sendo a etiquetagem, assumidamente, uma fase prévia ao funcionamento do ExtracTerm.

Assim, a ordem de aplicação das regras altera-se:

- (i) etiquetagem do *corpus*: corresponde à atribuição de todas as etiquetas metalinguísticas que uma dada forma pode assumir, independentemente do contexto em que ocorre;
- (ii) aplicação das regras de desambiguação: o ExtracTerm aplica as regras de desambiguação, cuja finalidade é a de anular as multietiquetas, para poder proceder à aplicação das regras seguintes (cf. anexo 6);
- (iii) aplicação das regras de reconhecimento: a aplicação destas regras consiste na identificação das estruturas pré-definidas que podem assumir as unidades terminológicas multilexémicas;
- (iv) levantamento das unidades terminológicas multilexémicas (cf. anexo 7).

Só acedemos aos resultados que decorrem da aplicação das regras de desambiguação (ii) e do levantamento das unidades terminológicas multlexémicas (iv), ambos apresentados em ficheiros HTML.

As regras de aprendizagem estão registadas no dicionário de regras de desambiguação, que estruturámos do modo que se segue:

[Art:def:][Pron:pess:][Pron:dem:][Prep1:a]>[N:f:s]@[Art:def:]

[Art:def:][Pron:pess:][Pron:dem:][Prep1:a]<[Ger]@[Art:def:]

[Art:def:][Pron:pess:][Pron:dem:][Prep1:a]<[Pp]>[Pron:poss:]@[Art:def:]

[Art:def:][Pron:pess:][Pron:dem:][Prep1:a]<[N]>[V:inf]@[Art:def:]

[Art:def:][Pron:pess:][Pron:dem:][Prep1:a]>[Art:indef:]@[Prep1:a]

[Adj:][Pp:]<[N:]>[Prep1:a]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:após]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:até]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:conforme]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:consoante]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:contra]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:desde]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:durante]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:entre]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:excepto]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:fora]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:menos]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:para]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:perante]@[Pp:]

[Adj:][Pp:]<[N:]>[Prep1:por]@[Pp:]

O algoritmo correspondente a aplicação das regras de desambiguação apresenta o seguinte código:

```

bool CAmbiguidades::condicao_aplicavel(CAmbiguidades&
ambiguidade, std::vector<CEntrada>::iterator& it)
{
    std::vector<CEntrada>::iterator    it_tmp = it;
    std::vector<std::string>::iterator it_rgrs_esq, it_rgrs_dir;

    // nas mesmas posições dos vectores de etiquetas da regra

    //std::string    dump_entrada = dump_strVector_2_string(it_tmp-
>etiqueta);

    for ( it_rgrs_esq = ambiguidade.m_condicao_esquerda.begin();
it_rgrs_esq != ambiguidade.m_condicao_esquerda.end(); it_rgrs_esq++)
    {
        it_tmp--;

        //std::string dump_curr = dump_strVector_2_string(it_tmp-
>etiqueta);
        //std::string          dump_esq
        =
dump_strVector_2_string(ambiguidade.m_condicao_esquerda);

        if ( !it_tmp->Contem_Etiqueta( (*it_rgrs_esq) ) )
            return false;
    }

    it_tmp = it;

    for ( it_rgrs_dir = ambiguidade.m_condicao_direita.begin();
it_rgrs_dir != ambiguidade.m_condicao_direita.end(); it_rgrs_dir++)
    {
        it_tmp++;

        //std::string dump_curr = dump_strVector_2_string(it_tmp-
>etiqueta);
        //std::string          dump_dir
        =
dump_strVector_2_string(ambiguidade.m_condicao_direita);

        if ( !it_tmp->Contem_Etiqueta( (*it_rgrs_dir) ) )
            return false;
    }

    // Se cheguei aqui é pq a condição se aplica à esquerda e à
direita, vou trocar etiqueta

    if ( ambiguidade.m_sinal_resolucao )
    {
        std::ofstream          out;
        std::vector<std::string>::iterator itEtiquetas;
        std::vector<std::string>::iterator iter;
        CEntrada
        copia_entrada;
    }

```

```

        copia_entrada = (*it);

        out.open( "AmbiOut.txt", std::ios_base::app);
        out << "Ambiguidade: " << copia_entrada.entrada;

        for ( itEtiquetas = copia_entrada.etiqueta.begin();
itEtiquetas != copia_entrada.etiqueta.end(); itEtiquetas++ )
        {
            out << "[";
            out << (*itEtiquetas);
            out << "]" ";
        }

        it->etiqueta.clear();

        for ( iter = ambiguidade.m_resolucao.begin(); iter !=
ambiguidade.m_resolucao.end(); iter++ )
        {
            std::string tmp =
copia_entrada.Devolve_Etiqueta_Semelhante((*iter));
            it->etiqueta.push_back(tmp);
        }

        out << " \tResolução: " << (*it).entrada;
        for ( itEtiquetas = (*it).etiqueta.begin(); itEtiquetas !=
(*it).etiqueta.end(); itEtiquetas++ )
        {
            out << "[";
            out << (*itEtiquetas);
            out << "]" ";
        }
        out << std::endl;
        out.close();
    }
    return true;
}

std::vector<std::string>
CAmbiguidades::Resolve_Ambiguidade(std::vector<CEntrada>::iterat
or& it)
{
    std::vector<std::string>          novas_etiquetas;
    std::vector<CAmbiguidade>        ambiguidades;
    std::vector<CAmbiguidade>::iterator itAmbiguidade;

    ambiguidades = Existe_Ambiguidade( it->etiqueta );

    if ( ambiguidades.size() > 0 )
    {
        for ( itAmbiguidade = ambiguidades.begin(); itAmbiguidade
!= ambiguidades.end(); itAmbiguidade++)
            if ( condicao_aplicavel( (*itAmbiguidade) , it) )
                return it->etiqueta;
    }

    return it->etiqueta;
}

```

A extracção obtida com a aplicação destas regras permite-nos avaliar a eficácia da metodologia adoptada.

### 3.1. Análise de resultados

Para avaliar com rigor o campo de actuação do extractor, comparamos um exemplo de *corpus* etiquetado antes da aplicação de regras de ambiguidades e o mesmo exemplo depois da aplicação das regras de ambiguidades.

Assinalamos a vermelho, as etiquetas metalinguísticas sobre as quais as regras de desambiguação irão actuar:

A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] figura [N:f:s] [V:cj] 2  
mostra [Pp:m:pl] [Adj:m:pl] os [N:m:pl] [V:cj] resultados [N:m:pl]  
obtidos [Pp:m:pl] [Adj:m:pl] para [N:m:pl] a [Art:def:f:s] [Pron:pess:3p:f:s]  
[Pron:dem:3p:f:s] [Prep1:a] zona [N:f:s] da [Prep2:em] área [N:f:s] em  
[N:m:pl] estudo [N:m:pl]. A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s]  
[Prep1:a] preto [Adj:m:s] estão representadas [Pp:f:pl] [Adj:f:pl] as  
[N:f:pl] áreas [N:f:pl] homogéneas [N:f:pl] do [N:m:pl] ponto  
de [N:f:s] vista [N:f:s] [Pp:f:s] [Adj:f:s] radiométrico [N:f:s], a [Art:def:f:s]  
[Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] branco [N:f:s] as [N:f:pl]  
transições [N:f:pl] radiométricas [N:f:pl] fortes [N:f:pl].

O mesmo texto etiquetado com o *ExtracTerm*, que agora contempla as regras de desambiguação, apresenta o seguinte formato de saída:

A [Art:def:f:s] figura [N:f:s] 2 mostra os resultados obtidos [Pp:m:pl] para a [Art:def:f:s] zona da área em estudo. A [Prep1:a] preto [Adj:m:s] estão representadas [Pp:f:pl] as áreas homogéneas do ponto de vista [N:f:s] radiométrico, a [Prep1:a] branco as transições ] radiométricas fortes.

As regras permitiram levantar a ambiguidade de 9 formas, reduzindo consideravelmente, a margem de erro na extracção de unidades terminológicas.

Para o levantamento das 9 formas, foram aplicadas as seguintes regras de desambiguação:

[Art:def:][Pron:pess:][Pron:dem:][Prep1:a]>[N:]@[Art:def:]

Ex: a figura; a zona; a cor

[N:][V:cj]<[Art:def:]@[N:]

Ex: a figura

[Pp:][Adj:]<[N:]>[Prep1:para]@[Pp:]

Ex: os resultados obtidos para

[Art:def:][Pron:pess:][Pron:dem:][Prep1:a]>[Adj:]@[Prep1:a]

Ex: a preto estão ... ; a branco estão ...

[Pp:][Adj:]@<[V:cj]@[Pp:]

Ex: estão representadas

[N:][Pp:][Adj:]<[Prep1:]@[N:]

Ex: ponto de vista

Apesar de os resultados serem encorajadores, consideramos que as regras de desambiguação carecem de aperfeiçoamento. Tal melhoramento só será atingido com o estudo mais aprofundado do comportamento de cada uma das entidades que constituem as classes gramaticais.

As regras que concebemos não abarcam todas as situações de ambiguidade em que ocorrem cada uma das classes gramaticais que fazem parte das regras de desambiguação. A título de exemplo, a identificação das diferenças entre `seguida` e `seguido` não é exaustiva, o que tem por consequência o não levantamento de ambiguidades em situações como as que passamos a apresentar:

`seguida` [Adj:f:s] `da` [Prep2:da:f:s] `Erosão` [N:f:s] `por` [Prep1:por] `B` `do` [Prep2:do:m:s] `conjunto` [N:m:s] `obtido` [Pp:m:s] [Adj:m:s].

ou ainda

`seguido` [Adj:m:s] `da` [Prep2:da:f:s] `Dilatação` [N:f:s] `por` [Prep1:por] `B` `do` [Prep2:do:m:s] `conjunto` [N:m:s] `obtido` [Pp:m:s] [Adj:m:s] `e` [Conj:e] `cuja` [Pron:rel:f:s] `notação` `é` ...

Apesar de as regras de desambiguação não estarem esgotadas, o `ExtracTerm` actua sobre todas as sequências que contêm ambiguidades, correspondentes às regras identificadas.

A comparação entre os resultados obtidos com `ExtracTerm` sem aplicação das regras de ambiguidades:

resultados	obtidos	{N:m:pl + Adj:m:pl}
área	em estudo	{N:f:s + Prep1:em + N:m:s}
áreas	homogéneas	{N:f:pl + Adj:f:pl}
ponto	de vista	{N:m:s + Prep1:de + N:f:s}
transições	radiométricas	{N:f:pl + Adj:f:pl}

e os resultados obtidos com o *ExtracTerm* com as regras de desambiguação:

área	em estudo	{N:f:s + Prep1:em + N:m:s}
áreas	homogéneas	{N:f:pl + Adj:f:pl}
transições	radiométricas	{N:f:pl + Adj:f:pl}

permitem-nos corroborar a vantagem desta metodologia. De seis propostas iniciais de unidades terminológicas multilexémicas, passamos a três, o que demonstra ser uma redução notável no esforço de análise e de validação dos dados por parte do especialista.

### 3.2. Teste do *ExtracTerm* sobre um *corpus* alargado

Atingindo este ponto, é necessário fazer incidir o *ExtracTerm* sobre novos textos, por forma a verificar a sua exequibilidade.

Assim, optámos, por fazer correr o *ExtracTerm* a partir de uma tese para obtenção de grau de especialista à categoria de Investigador Auxiliar do Laboratório Nacional de Engenharia Civil (Fonseca, 1998). Este texto é constituído por 49834 ocorrências para 4330 formas.

### 3.2.1. Análise dos resultados

Este texto é então varrido pelo ExtractTerm. Verificámos que, das 49834 ocorrências, 3040 não foram etiquetadas, o que corresponde, aproximadamente, a 937 formas, após a eliminação de números, alíneas, palavras estrangeiras, erros gráficos e ortográficos, entre outros.

As palavras não etiquetadas correspondem à ausência de registo no dicionário de formas flexionadas. Após verificação de tal facto, adicionámos uma função ao etiquetador. Optámos por extrair para um ficheiro todas as formas não etiquetadas. Desta forma, podemos atribuir etiquetas metalinguísticas a estas formas e, posteriormente, adicioná-las ao dicionário de formas flexionadas. Deste modo, o EtiqueLex, numa fase final, ou o ExtractTerm numa fase inicial, permite a auto-alimentação do dicionário de formas flexionadas.

Exemplifiquemos com o *corpus* teste, a qualidade da etiquetagem:

A aquisição de imagens por meio de  
Detecção Remota ( DR ) baseia-se no facto  
de os objectos  
da superfície terrestre reflectirem  
radiação\_electromagnética ( REM ) proveniente do  
Sol e emitirem eles a sua própria  
própria REM. A intensidade e a  
composição espectral da REM emitida  
assim como a percentagem de  
REM reflectida dependem das  
características físicas e químicas  
do objecto. Assim, é possível  
inferir das características físicas  
e químicas de objectos à  
superfície da Terra através da  
REM proveniente desses

objectos e registada por sensores  
a bordo de satélites artificiais  
da Terra, designados  
satélites de Detecção Remota

A extracção obtida a partir deste parágrafo é a seguinte:

aquisição de imagens {N:f:s + Prep1:de + N:f:pl}  
Detecção Remota {N:f:s + Adj:f:s}  
objectos da superfície {N:m:pl + Prep2:da:f:s + N:f:s}  
superfície terrestre {N:f:s + Adj:2gen:s}  
composição espectral {N:f:s + Adj:2gen:s}  
superfície da Terra {N:f:s + Prep2:da:f:s + N:f:s}  
bordo de satélites {N:m:s + Prep1:de + N:m:pl}  
satélites artificiais {N:m:pl + Adj:2gen:pl}  
satélites de Detecção {N:m:pl + Prep1:de + N:f:s}  
satélites de Detecção Remota {N:m:pl +  
Prep1:de + N:f:s + Adj:f:s}  
Detecção Remota {N:f:s + Adj:f:s}

Com cada *corpus* novo podemos enriquecer os dicionários de base, o que nos garante, à partida, a qualidade da etiquetagem, facto que tem consequências na qualidade da extracção. Quanto mais alargado for o nosso *corpus*, menores são as hipóteses de encontrar formas não etiquetadas e maiores as hipóteses de identificar neologismos terminológicos.

Com esta amostra, podemos verificar que, na globalidade, a quantidade de formas não etiquetadas não é muito elevada. Tal facto permite-nos concluir, de forma empírica, que o *corpus* de partida é representativo no que concerne a terminologia utilizada pelo especialistas em Detecção Remota.

## **Conclusão**

---

## Conclusão

Na dissertação que ora apresentamos debatemos um conjunto de pressupostos teóricos cuja reformulação se revelou de importância fundamental no desenvolvimento da metodologia que esteve na base da concepção e da construção de dois programas informatizados para o tratamento automático da língua de especialidade: um etiquetador, o EtiqueLex, e um extractor, o ExtracTerm.

Os resultados, que consideramos bastante satisfatórios, obtidos a partir do funcionamento destes dois produtos, são, em parte, consequência do estabelecimento de critérios rigorosos no que diz respeito à selecção dos textos. Por outro lado, a identificação, do ponto de vista teórico, das entidades linguísticas sobre as quais incidiu o nosso estudo, as unidades terminológicas multilexémicas, permitiu-nos isolar as estruturas terminogénicas mais frequentes no nosso *corpus*.

A partir da observação deste mesmo *corpus*, elaborámos as regras de reconhecimento que permitiram ao ExtracTerm reconhecer todas as estruturas previamente contempladas no dicionário de regras de reconhecimento por nós concebido.

Também as regras de desambiguação foram elaboradas a partir da observação dos primeiros resultados obtidos com o ExtracTerm.

O ExtracTerm é um programa dinâmico. Os dicionários que estão na base do seu funcionamento, isto é, o dicionário de formas flexionadas, o dicionário de

locuções e o dicionário de Detecção Remota, são dicionários abertos, aos quais é possível a qualquer momento adicionar informação, sempre que tal se revele necessário.

A funcionalidade que permite ao ExtracTerm extrair todas as formas não etiquetadas para um ficheiro é de uma enorme utilidade, uma vez que possibilita acrescentar etiquetas metalinguísticas a todas as formas presentes nesse ficheiro e adicioná-las posteriormente ao dicionário de formas flexionadas. Deste modo, o ExtracTerm proporciona uma actualização permanente do dicionário, garantindo a qualidade da etiquetagem dos textos de uma determinada área de especialidade.

Não se verifica qualquer impedimento do ExtracTerm ser aplicado a outras áreas de especialidades. No caso de, por exemplo, recorrermos ao ExtracTerm para a etiquetagem de um *corpus* em Medicina, empregamos a mesma metodologia, sabendo *a priori* que as formas não etiquetadas serão, no início, em número superior.

Todos os dicionários que suportam o ExtracTerm encontram-se actualmente em ficheiros TXT, sendo a sua capacidade virtualmente infinita, limitados, exclusivamente, ao espaço do disco rígido em que se encontram instalados.

O acréscimo de textos ao *corpus*, bem como o enriquecimento gradual do dicionário de formas flexionadas, são mecanismos que permitem uma actualização constante da terminologia empregue pelo especialista. Perante tais

circunstâncias. o *ExtracTerm* irá encontrar cada vez menos formas não etiquetadas, sendo a probabilidade de vir a encontrar neónimos considerável.

Não podemos, no entanto, descurar uma observação minuciosa e regular aos *corpora*, a fim de detectar eventuais inovações sintácticas ao nível da estrutura das unidades terminológicas multilexémicas. Caso tais inovações se verifiquem, acrescentar-se-á, no dicionários de regras de reconhecimento, a regra que dá conta desses processos inovadores, para que o *ExtracTerm* possa proceder ao levantamento das unidades que encontram correspondência nas novas matrizes terminogénicas. Os bons resultados obtidos com a aplicação desta metodologia, não nos impedem, porém, de apontar algumas lacunas que, conscientemente, não foram preenchidas, por ultrapassarem claramente o âmbito desta dissertação. Reconhecemos, não obstante, a necessidade de continuarmos a desenvolver uma reflexão apurada em torno de tais lacunas como um factor imprescindível para o futuro aperfeiçoamento da metodologia que ora utilizámos e, conseqüentemente, para um melhoramento das potencialidades deste programa.

Neste estudo, privilegiámos uma aproximação lexical às unidades terminológicas multilexémicas cuja base é um nome, por serem as mais representativas de qualquer terminologia. Não podemos, contudo, ignorar outras classes gramaticais que, embora menos frequentes, são de extrema importância para o tratamento automática da língua natural. Estamos, por exemplo, a referir-nos aos verbos que, embora com propriedades distintas, também

denominam conceitos fundamentais para o entendimento de uma língua de especialidade.

Assim, o estudo das unidades terminológicas multilexémicas não ficou esgotado. Consideramos de uma grande utilidade o aprofundamento do comportamento linguístico assumido isoladamente pelas classes gramaticais mais frequentes: os nomes, os adjectivos e as preposições.

As regras de desambiguação que concebemos foram criadas a partir da observação que efectuámos ao contexto lexical em que ocorrem as unidades terminológicas multilexémicas. Não fomos exaustivos no levantamento das ambiguidades para cada classe gramatical atribuída a uma mesma forma: por exemplo, não identificámos todas as situações lexicais em que um se confunde com um . O levantamento dos contextos lexicais em que ocorrem as ambiguidades é primordial para a concepção de regras de desambiguação mais rigorosas e abrangentes.

Também não fazia parte dos objectivos desta trabalho estabelecer critérios e elaborar regras que permitissem distinguir o comportamento de um do comportamento de outro . A equipa de G. Gross visa obter a descrição exaustiva de todas as unidades lexicais, procedendo a uma classificação semântica e sintáctica, tendo por base o conceito de classes de objectos. Tal descrição linguística tem por finalidade a construção de dicionários electrónicos. De acordo com esta metodologia, considera-se que um nome pertence, por exemplo, à classe dos meios de transporte <mt>, que por sua vez se divide em meios de transporte em comum <mt-a>, meios de transporte aéreos <mt-a>,

etc.. Todas as unidades lexicais estão associadas, desta forma, a classes e subclasses (cf. Blanco, 1999).

Passamos a ilustrar com algumas entradas lexicográficas, a construção de um dicionário electrónico bilingue, francês - espanhol:

auto à huit cylindres/G:nf/T:inc/C:mt-tmi/D:transp.ter/Es: coche de ocho cilindros

auto aérodynamique/G:inf/T:inc/C:mt-tmi/D:transp.ter/Es:coche aerodinámico

auto de course/G:inf/T:inc/C:mt-tmi/D:trasnp:ter/S:voiture de course/Es:coche de carreras

Embora esta metodologia tenha sido concebida para a construção de dicionários electrónicos monolingues e bilingues, julgamos que esta descrição sintáctico-semântica pode ser utilizada para descrever as unidades terminológicas e elaborar regras de desambiguação. Esta descrição é de tal forma minuciosa que nos permite distinguir o comportamento linguístico das unidades de classe de objecto para classe de objecto.

Por outro lado, a nossa metodologia pode funcionar como uma metodologia complementar à da de G. Gross. As unidades lexicais - muitas delas terminológicas - descritas pela equipa de G. Gross são retiradas de dicionários já existentes. O levantamento automático de unidades a partir de *corpora* permite um aumento das unidades a descrever para uma determinada área do conhecimento.

Não obstante, uma inevitável da metodologia, consideramos os resultados obtidos muito encorajadores: a forma de funcionamento do programa

está estabelecida, dependendo da qualidade da sua actuação exclusivamente do melhoramento da descrição linguística.

Resta-nos, ainda, fazer uma referência à utilização diminuta que fizemos da estatística. Tal opção metodológica, não invalida a futura utilização de uma abordagem estatística mais desenvolvida, não em substituição da aproximação linguística, mas em complementaridade a esta.

Acreditamos que a estatística pode validar os resultados obtidos através da análise linguística, na medida em que permite quantificar a taxa de sucesso do etiquetador ou de uma determinada regra. Com o recurso à estatística, podemos saber em percentagem o valor que assume uma determinada classe gramatical no *corpus* ou ainda se uma determinada estrutura merece ser melhor observada com vista à criação de regras de desambiguação.

Este tipo de informação permite-nos gerir a investigação em função da repercussão que pode ter o estudo de uma entidade linguística num determinado *corpus* e deste modo estabelecer prioridades de investigação.

Com esta dissertação, pretendemos demonstrar que o tratamento automático da língua natural é a consequência de uma acção concertada transdisciplinar, que resulta da fusão do saber linguístico com o saber informático. Do saber linguístico resulta a análise linguística, do saber informático a construção dos algoritmos que permitem a aplicação das regras linguísticas.

## **Bibliografia**

---

## **Bibliografia de referência**

**AARTS**, Jan (1990), «Corpus Linguistics: An Appraisal», *Computers in Literary and Linguistics Research*, Proceedings of the International Conferences of the Association for Literary and Linguistic Computing, Paris-Genève, Hamesse, J., Zampolli, A., pp. 13-28.

**AARTS**, Jan (1996), «Grammatical Annotation», *ICAME Journal* No. 20, ICAME, The Norwegian Computing Centre for the Humanities, Norway, Bergen University, pp. 104-107.

**AARTS**, Jan (1991), «Intuition-based and observation-based grammars», *English Corpus Linguistic*, Studies in Honour of Jan Svartvik, ed. by Karin Aijmer & Bengt Altenberg, London and New York, Longman, pp. 44 - 62.

**ADAM**, Jean-Michel (1990), *Éléments de linguistique textuelle. Théorie et pratique de l'analyse textuelle*, Liège, Mardaga, 265 p.

**ADAM**, Jean-Michel (1997), *Les textes: types et prototypes. Récit, description, argumentation, explication et dialogue*, Paris, Nathan, 223 p.

**ADAM**, Jean-Michel (1999), *Linguistique textuelle, des genres de discours aux textes*, Paris, Nathan, 208 p.

**AMOSSY**, Ruth; **PIERROT**, Anne Herschberg (1997), *Stéréotypes et clichés. Langue, discours, société*, Paris, Nathan, 128 p.

**ASCOMBRE**, Jean-Claude (1999), «Le jeu de la prédication dans certains composés nominaux», *Langages*, n° 122, Le groupe nominal: contraintes distributionnelles et hypothèses de descriptions, sous la direction d'Anne Daladier, Paris, Larousse, pp. 52 - 69.

**ASSADI**, Houssem; **BOURIGAULT**, Didier (2000), «Analyses syntaxique et statistique pour la construction d'ontologies à partir de textes», *Ingénierie des connaissances, Évolutions récentes et nouveaux défis*, Ed. Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault, Paris, Eyrolles, pp. 243 - 256.

**AUGER**, Pierre; **DROUIN**, Patrick; **L'HOMME**, Marie-Claude (1991), «Automatisation des Procédures de Travail en Terminographie», *META*, Vol. 36, n° 1, Montréal, PUM, pp. 121 - 127.

**BACELAR**, Feranda; **PEREIRA**, Luísa; **SARAMAGO**, João (2000), «Portuguese corpora at CLUL», *Proceedings of the Second International Conference on Language Resources and Evaluation*, Vol. III, Athens, Greece, pp. 1603 - 1607.

**BAHNS**, Jens (1996), *Kollokationen als lexikographisches Problem. Eine Analyse allgemeiner und spezieller Lernerwörterbücher des Englischen*, Tübingen, Max Niemeyer Verlag, 135 p.

**BAKHTINE**, Mikhaïl (1984), *Esthétique de la création verbale*, Paris, Gallimard, 400 p.

**BAKHTINE**, Mikhaïl (2000) «The problem of the speech genres», *The Discourse Reader*, Ed. Adam Jaworski and Nikolas Coupland, New York, Routledge, pp. 121 - 132.

**BALDINGER**, Kurt (1984), *Vers une semantique moderne*, Paris, Klincksieck, 255 p.

**BARNBROOK**, Geoff (1996), *Language and Computers, A Pratical Introduction to the Computer Analysis of Language*, Edinburgh, Edinburgh University Press, 209 p.

**BARREIRO**, Anabela Marques (1998), *Propriedades Sintáctico-Semânticas dos Participios Passados em Português Europeu*, Dissertação de Mestrado em Linguística, Universidade Nova de Lisboa, 149 p.

**BEACCO**, Jean-Claude (1995), «À propos de la structuration des communautés discursives: beaux-arts et appréciatif», *Les enjeux des discours spécialisés*, Coordination éditoriale, Jean-Claude Beacco, Sophie Moirand, Paris, Presses de la Sorbonne Nouvelle, pp. 135 - 157.

**BEACCO**, Jean-Claude, **MOIRAND**, Sophie (1995), «Autour des discours de transmission de connaissances», *Langages* n° 117, Paris, Larousse, pp. 32 - 53.

**BECHARA**, Evanildo (2001), *Moderna Gramática Portuguesa*, Rio de Janeiro, Ed. Lucerna, 672 p.

**BÉJOINT**, Henri (1997), «Regards sur la définition en terminographie», in *Cahiers de Lexicologie*, Vol. LXX 1997 - I, Paris, Didier Erudition, pp. 19 - 26.

**BÉJOINT**, Henri (1993), «La définition en terminographie», *Aspects du vocabulaire*, Arnaud, P. et Thoiron (éds), Lyon, PUL, pp. 19 - 26.

**BÉJOINT**, Henri, **THOIRON**, Philippe (1997), «Modèle relationnel, définition et dénomination», *Autour de la Dénomination*, Boisson, H., Thoiron, Ph. (éds) Lyon, PUL, p. 187-204.

**BENNETT**, Paul (1993), «Noms composés, termes, dénominations complexes: problématiques linguistiques et traitements automatiques», *Traitement Automatique des Langues*, 34, 2, Paris, Revue de l'ATALA, pp. 43 - 58.

**BENSON**, Morton; **BENSON**, Evelyn, **ILSON**, Robert (1997), *The BBI Dictionary of English Word Combinations*, Amsterdam / Philadelphia, John Benjamins, pp. VII - XXXIX.

**BESSÉ**, Bruno de (1990), «La définition terminologique», Actes du Colloque *La Définition*, Paris, Larousse, pp. 252 - 261.

**BIBER**, Douglas; **CONRAD**, Susan; **REPPEN**, Randi (1998), *Corpus linguistics: investigating language structure and use*, Cambridge, Cambridge University Press, 300 p.

**BLANCO**, Xavier (199), *Lexicographie bilingue français-espagnol et classes d'objets*, Barcelona, Universitat Autònoma de Barcelona, 87 p.

**BONNAFOUS**, Simone, **TOURNIER**, Maurice (1995), «Analyse su discours, lexicométrie, communication et politique», *Langages* n° 117, Paris, Larousse, pp. 67 - 81.

**BOUILLON**, Pierrette (1998), *Traitement automatique des langues naturelles*, Paris, Bruxelles, Co-édition AUPELF- UREF, 245 p.

**BOURIGAULT**, Didier (1993), «Analyse syntaxique locale pour le repérage de termes complexes dans un texte», *Traitement Automatique des Langues*, 34, 2, Paris, Revue l'ATALA, pp. 105 - 118.

**BOURIGAULT**, Didier (1994), *LEXTER un Logiciel d'Extraction de TERminologie. Application à l'extraction des connaissances à partir de textes*, Paris, Thèse en mathématiques, informatique appliquée aux sciences de l'homme, École des Hautes Études en Sciences Sociales

**BOURIGAULT**, Didier (2001), «Introduction», *Recent Advances in Computational Terminology*, Ed. by Didier Bourigault, Christian Jacquemin,

Marie-Claude l'Homme, Amsterdam / Philadelphia, John Benjamins, pp. VIII - XVIII.

**BOURIGAULT**, Didier; **SLOZDIAN**, Monique (1999), «Pour une terminologie textuelle», *Terminologies Nouvelles*, N° 19, Bruxelles, Revue coédité par l'Agence de la Francophonie et Communauté française de Belgique, pp. 29 - 32.

**BOUTET**, Josiane; **GARDIN**, Bernard; **LACOSTE**, Michèle (1995), «Discours en situation de travail», *Langages*, n°117, Paris, Larousse, pp. 12 - 31.

**BRONCKART**, Jean-Paul (1994) - *Le fonctionnement des discours. Un modèle psychologique et une méthode d'analyse*, Lausanne, Delachaux & Niestlé, 175p.

**BRONCKART**, Jean-Paul (1997), *Activité langagière, textes et discours. Pour un interactionisme socio-discursif*, Lausanne, Delachaux & Niestlé, 351 p.

**BUDIN**, Gerhard (1990), «Terminological Analysis of LSP Phraseology», in *Journal of the International Institute for Terminology Research – IIITF – Vol. 1*, no° 1-2, Wien, International Network for Terminology (TermNet), pp. 64-69.

**BUDIN**, Gerhard; **BÜHLER**, Hildegund (1999), «Grundsätze und Methoden der neuen Terminographie», Hoffmann, L.; Kalver-Kämper. H.; Wiegand, H. E. [eds.]. *Fachsprachen. / Languages for Special Purposes. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin / New York, deGruyter, pp. 2097 - 2108.

**BUVET**, P.- A. (1998), «Détermination et classes d'objets», *Langages*, N° 131, Les classes d'objets, sous la direction de Denis Le Pesant et Michel Mathieu-Colas, Paris, Larousse, pp. 91 - 102.

**CABRÉ**, Teresa (1993), *La Terminología. Teoría, metodología, aplicaciones*, Barcelona, Editorial Antártida/Emprúries, 529 p.

**CADIOT**, Pierre (1992), «À entre deux noms: vers la composition nominale». *Lexique*, n° 11, *Les prépositions. Méthodes d'analyse*, Lille, Presses Universitaires de Lille, pp. 193 - 240.

**CADIOT**, Pierre (1997), *Les prépositions en français*, Paris, Armand Colin, 295 p.

**CARAÇA**, João (1997), *O que é a ciência*, Lisboa, Difusão Cultural, 111 p.

**CARNAP**, Rudolf (1997), *Signification et nécessité*, Paris, Gallimard, 382 p.

**CARRÉ**, René; **DÉGREMONT**, J.-F.; **GROSS**, Maurice et al. (1991), *Language Humain et Machine*, Paris, Presses du CNRS, 298 p.

**CHARAUDEAU**, Patrick (1995), «Une analyse sémiolinguistique du discours», *Langages* n° 117, Paris, Larousse, pp. 96 - 111.

**CHARAUDEAU**, Patrick (1997), *Les discours d'information médiatique. La construction du miroir social*, Paris, Nathan, 286 p.

**CHARLET**, Jean, **ZACKLAD**, Manuel, **KASSEL**, Gilles, **Bourigault**, Didier (2000) «Ingénierie des connaissances: recherches et perspectives... », *Ingénierie des connaissances, Évolutions récentes et nouveaux défis*, Ed. Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault, Paris, Eyrolles, pp. 1 - 24.

**CHOMSKY**, Noam (1980), *Estruturas Sintáticas*, Lisboa, Ed. 70, 126.

**CINTRA**, Lindely, **CUNHA**, Celso (1994), *Nova Gramática do Português Contemporâneo*, Lisboa, Edições João Sá da Costa, 734 p.

**CONCEIÇÃO**, M. Célio (1996), «Terminologie et transmission du savoir: (re)construction(s) de concepts», *Daylang, Sématique des termes et construction des connaissances*, dir. Delavigne V., e Bouveret M., Rouen, Université de Rouen, pp. 33 - 42.

**CONCEIÇÃO**, M. Célio (1997), «La mémoire des termes: analyse conceptuelle», *Mémoire des mots, Actes des Vèmes Journées Scientifiques du Réseau LTT de l'AUELF-UREF*, Tunes, AUELF-UREF, Serviced, pp. 369 - 376.

**CONCEIÇÃO**, M. Célio (2000), «Terminologie, connaissance et industrie», *Actas da Conferência sobre cooperação no domínio da terminologia na Europa*, dir S. Lervad, Paris, União Latina, AET, pp. 91 - 96.

**CONDAMINES**, Anne, **REBEYROLLE**, Josette (2000), «Construction d'une base de connaissances terminologiques à partir de textes: expérimentation et définition d'une méthode», *Ingénierie des connaissances, Évolutions récentes et nouveaux défis*, Ed. Jean Charlet, Manuel Zacklad, Gilles Kassel, Didier Bourigault, Paris, Eyrolles, pp. 225 - 242.

**CORBIN**, Danielle (1992), «Hypothèses sur les frontières de la composition nominale», *Cahiers de Grammaire*, n° 17, pp. 26 - 55.

**CORBIN**, Danielle; **TEMPLE**, Martine (1994), «Le monde des mots et des sens construits: catégories sémantiques, catégories référentielles», *Cahiers de Lexicologie*, n° 65, Paris, Didier, 5 -28.

**CORBIN**, Danielle (1997) «Locutions, composés, unités polylexématiques: lexicalisation et mode de construction», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 53 - 102.

**CORMIER**, Monique; **RIOUX**, Louis-Paul (1991), «Procédés de Formation et Matrices Terminogéniques en Terminologie des Systèmes Experts», *META*, Vol. 36, nº 1, Montréal, PUM. pp.248 - 268.

**COSTA**, M. Rute V. (1990) - "A coesão interna das lexias complexas enquanto unidades terminológicas", *Actas do Colóquio de Lexicologia e Lexicografia*, Lisboa, Centro de Estudos Comparados, Linha de Acção 2 - INIC, pp. 167 - 122.

**COSTA**, M. Rute V. (1993), *Terminologia da Economia Monetária. Relações conceptuais e semânticas numa sistemática terminológica e lexicográfica*, Dissertação de Mestrado em Linguística, Universidade Nova de Lisboa, 191 p.

**COSTA**, M. Rute V. (1994), «Bases de données textuelles et automatisation en terminographie», *Terminologie et Phraséologie. Acteurs et aménageurs. Actes de la Deuxième Université d'Automne en Terminologie*, Paris, La Maison du Dictionnaire, pp. 83 - 92.

**COSTA**, M. Rute V. (2001) - "O termo como veículo de especificidades conceptuais e semânticas", *Polifonia*, UNIL, Nº 4, Lisboa, Colibri, pp. 199 - 204.

**COSTA**, M. Rute V. (2001), «O conceito, a denominação e a definição terminológica», *Terminologias 12*, Termip, Lisboa, Colibri (no prelo).

**COSTA**, M. Rute V. (2001) - "Contribuição para a delimitação do texto de especialidade", *L'éloge de la différence: la voix de l'autre*, *Actas das VI<sup>es</sup> Journées scientifiques du Réseau Lexicologie, Terminologie et Traduction*, organizado pela Agence francophone pour l'enseignement supérieur et la recherche, l'École de traducteurs et d'interprètes de Beyrouth (ETIB) de l'Université Saint-Joseph, Beyrouth, 11 - 13 novembre 1999, (no prelo).

**COSTA**, M. Rute V., **LINO**, M. Teresa, **CONCEIÇÃO**, M. Célio (1996) - (em colaboração) "Terminologia, Informática e Multimédia", *Jornada Panllatina de Terminologia. Perspectives i camps d'aplicació*, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 73 - 81.

**COUTINHO**, M. Antónia (1999), *Texto(s) e competência lexical*, Tese de Doutoramento em Linguística, Universidade Nova de Lisboa, 413 p.

**COWIE**, A. P. (1998), «Introduction», *Phraseology, Theory, Analysis, and Applications*, Oxford, Oxford University Press, pp. 1 - 20.

**COWIE**, A. P. (1998), «Phraseological Dictionaries: Some East-West Comparisons», *Phraseology, Theory, Analysis, and Applications*, Oxford, Oxford University Press, pp. 209 - 228.

**CUSIN-BERCHE**, Fabienne (1999), «La notion d'«unité lexicale» en linguistique et son usage en lexicologie», *Statut de l'unité lexicale, Revue LINX*, n° 40, sous la direction de Gérard Petit, Nanterre, Université Paris X, pp. 11 - 31.

**CRISPIM**, Maria de Lourdes; **XAVIER**, Maria Francisca (1999), «Constitution (et utilisation d'un corpus de portugais médiéval)», *Revue française de linguistique appliquée*, IV-1, Amsterdam, Ed. 'De Werelt', pp. 41 - 45.

**CZAP**, H. (1989), «Le concept de concept», in *Terminologie Diachronique*, Actes du Colloque organisé à Bruxelles les 25 et 26 Mars 1988, Centre de Terminologie de Bruxelles, Institut Marie Haps, Conseil International de la Langue Française, pp. 69 - 70.

**DAHLBERG**, I. (1981), «Les objets, les notions, les définitions et les termes», *Textes choisis de terminologie, I. Fondements théoriques de la terminologie*, dir. V.I. Siforov, Québec, GIRSTERM, pp. 221 - 282.

**DAILLE**, Béatrice (1994), *Approche mixte pour l'extraction de terminologie: statistique lexicale et filtres linguistiques*, Thèse de doctorat en informatique fondamentale, Université Paris VII, 228 p.

**DAILLE**, Béatrice (1995), «Repérage et extraction de terminologie par une approche mixte statistique et linguistique», *Traitement Automatique des langues*, 36, 1 - 2, Paris, pp. 101 - 118.

**DAILLE**, Beatrice (1996), «Study and Implementation of Combined Techniques for Automatic Extraction of Terminology», *The Balancing Act. Combining Symbolic and Statistical Approaches to Language*, ed. by Judith L. Klavans and Philip Resnik, Cambridge, Massachusetts, London, England, The MIT Press, pp. 49 - 66.

**DAILLE**, Béatrice (1996), «Study and Implementation of Combined Techniques for Automomatic Extraction of Terminology», *The Balance Act: Combining Symbolic ans Statistical Approaches to Language*, ed. Judith L.Klavans, Philip Resnik, London, The MIT Press, pp. 49 - 66.

**DANISH STANDARDS ASSOCIATION** (1991), *SGML - ODA. Présentation des concepts et comparaisons fonctionnelle*, Paris, AFNOR, 87 p.

**DANLOS**, Laurence (1988), «Introduction: lexique-grammaire des expressions figées», *Langages*, n° 90, Les expressions figées, sous la direction de Laurence Danlos, Paris, Larousse, pp. 5 - 6.

**DANLOS**, Laurence (1994), «Coder les informations monolingues sur les noms pour éviter des règles bilingues sensibles du contexte», *Le traducteur et l'ordinateur*, *Langages*, n° 116, sous la direction de Jean-René Lamiral, Paris, Larousse, pp. 95 - 110.

**DARMSTETER**, Arsène (1967), [1874], *Traité de la formation des noms composés*, Paris, Champion, 365 p.

**DAVID**, Sophie; **PLANTE**, Pierre (1990), «Le la nécessité d'une approche morpho-syntaxique en analyse de textes», *Revue de liaison de la recherche en informatique cognitive des organisations*, Vol 2, n° 3, Québec, ICO, pp. 140 - 165.

**DAVID**, Sophie; **DUMAS**, Lucie; **MARANDIN**, Jean-Marie; **PLANTE**, André; **PLANTE** Pierre (1990) , *Termino. Version 1.0 – Rapport de Recherche* par le Groupe R.D.L.C. – Recherche et Développement en Linguistique Computationnelle, Centre d'Analyse de Textes par Ordinateur, Université du Québec à Montréal, 57 p.

**DAVID**, Sophie (1993), «Remarques à propos du mode de construction des unités de forme NN», *Traitement Automatique des Langues*, 34, 2, Paris, Revue l'ATALA, pp. 59 - 74.

**DESCLÉS**, Jean-Pierre (1997), «Systèmes d'exploration contextuelle», *Co-texte et calcul du sens*, (dir. Claude Guimier), Caen, Presses Universitaire de Caen, pp. 215 - 231.

**DROZD**, L. (1981), «Science terminologique: objet et méthode», *Textes choisis de terminologie, I. Fondements théoriques de la terminologie*, dir. V.I. Siforov, Québec, GIRSTERM, pp. 115 - 132.

**DUBOIS**, Danièle (dir.), *Sémantique Cognitive. Catégorie, prototypes, typicalité*, Coll. Sciences du Langage, Paris, CNRS Édition, pp. 342.

**DUBOIS**, Danièle (1997), « Catégories, prototypes et figements. Constructions d'invariants et systèmes symboliques», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 103 - 130.

**DUCROT**, Oswald (1980), *Les mots du discours*. Les Éditions de Minuit, Paris, 241 p.

**ECO**, Umberto (1993), *Leitura do Texto Literário. Lector in Fabula*, Lisboa, Presença, 263 p.

**EL-BÈZE**, Marc, **SPRIET**, Thierry (1995), «Intégration et contraintes syntaxiques dans un système d'étiquetage probabiliste», *Traitement Automatique des Langues*, 36, 1 - 2, Paris, Revue l'ATALA, pp. 47 - 66.

**FELBER**, Helmut (1987), *Manuel de Terminologie*, Paris, Unesco / Infoterm, 447 p.

**FELBER**, Helmut, **BUDIN**, Gerhard (1989), *Terminologie in Theorie und Praxis*, Tübingen, Gunter Narr Verlag, 315 p.

**FELBER**, Helmut (1994/1995), «Terminology Research: Its Relation to the Theory of Science», in *Terminologie et Linguistique de Spécialité. Études de vocabulaires et textes spécialisés*, ALFA, Volume 7/8, 1994 /1995, Actes de langue française et de linguistique, Universitas Dalhousiana, Nova Scotia, Canada, pp. 163 - 172.

**FÓNAGY**, Ivan (1997), «Figement et changement sémantiques», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 131 - 164.

**FOUCAULT**, Michel (1969), *L'archéologie du savoir*, Coll. Bibliothèque des Sciences Humaines, Paris, Gallimard, 275 p.

**FOUCAULT**, Michel (1971), *L'ordre du discours*, Paris, Gallimard, 82 p.

**FRADIN**, Bernard (1996), «L'identification des unités lexicales», *Sémiotiques*, n° 11, Savoirs lexicaux et savoirs encyclopédiques, t.1 Légitimer la description

lexicale, n° dirigé par David Piotrowski et Peter Stockinger, Paris, Didier Érudition, pp. 55 -93.

**FREY, Claude; LATIN, Danièle (1997),** *Le corpus lexicographique. Méthodes de constitution et de gestion*, Actes des troisièmes journées scientifiques du réseau thématique de recherche «Etude du français en francophonie», Actualité scientifique, Louvain-la-Neuve, Duculot, AUPELF UREF, Recueils, 424 p.

**FUCHS, Catherine; DANLOS, Laurence; Lacheret-DUJOUR, Anne et ali.** (1993), *Linguistiques et traitements automatiques des langues*, Paris, Hachette, 303 p.

**FUCHS, Catherine (1996),** *Les ambiguïtés du français*, Paris, Ophrys 184 p.

**GALINSKI, Christian (1990),** «Terminology and Phraseology», in *Journal of the International Institute for Terminology Research – IIITF – Vol. 1, no° 1-2*, Wien, International Network for Terminology (TermNet), pp.70-86.

**GAUDIN, François (1996),** «L'ombre du concept», in *La Dénomination, Méta*, Numéro spécial, vol. 41, n°4, Montréal, Presses Universitaire de Montréal, pp. 604 - 621.

**GLÄSER, Rosemarie (1994/95),** «Relations between Phraseology and Terminology with Special Reference to English», *Terminologie et Linguistique de Spécialité. Étude de vocabulaires et textes spécialisés*, Halifax, ALFA, pp 41- 60.

**GOFFIN, Robert (1989),** «La terminologie des sciences et des techniques nucléaires. Un cas de diachronie récente», *Terminologie Diachronique*, Actes du Colloque organisé à Bruxelles les 25 et 26 Mars 1988, Centre de Terminologie de Bruxelles, Institut Marie Haps, Conseil International de la Langue Française, pp. 94 - 107.

**GOUADEC**, Daniel (1994), «Nature et traitement des entités phraséologiques», *Terminologie et phraséologie. Acteurs et aménageurs*, Paris, Maison du Dictionnaire, pp.167 - 193.

**GRECIANO**, Gertrud (1993), «Vers une modélisation phraséologique: Acquis et projets d'EUROPHRAS», *Terminologies Nouvelles*, N° 10, Bruxelles, RINT, pp. 16 - 22.

**GRECIANO**, Gertrud (1997), «La phraséogénèse du discours», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 179 - 200.

**GREIMAS**, Algirdas Julien (1966), *Sémantique structurale*, Paris, PUF, 216 p.

**GREIMAS**, Algirdas Julien (1976), *Sémiotiques et Sciences Sociales*, Paris, Éditions du Seuil, 262 p.

**GREIMAS**, Algirdas Julien; **COURTÉS**, Joseph (1979), *Sémiotique. Dictionnaire raisonné de la théorie du langage*, Tome 1, Paris, Hachette, 443 p.

**GREIMAS**, Algirdas Julien; **COURTÉS**, Joseph (1979), *Sémiotique. Dictionnaire raisonné de la théorie du langage*, Tome 2, Paris, Hachette, 270 p.

**GRFENSTETTE**, Gregory (1996), «Evaluation techniques for Automatic Semantic Extraction: Comparing Syntactic And Window Based Approches», *Corpus Processing for Lexical Acquisition*, ed. by Branimir Boguraev and James Pustejovsky, Cambridge, Massachusetts, London, England, The MIT Press, pp. 205 - 216.

**GROSS**, Gaston; **CHARAND**, Jacques; **VIVÉS**, Robert; **MATHIEU-COLAS**, Michel; **BILLY**, Pierre (1986), «Typologie des noms composés», Rapport technique A.T.P. du C.N.R.S., - *Nouvelles Recherches sur le Langage*. Paris, Université de Paris 13.

**GROSS**, Gaston (1988), «Degré de figement des noms composés», *Langages*, n° 90, Les expressions figées, sous la direction de Laurence Danlos, Paris, Larousse, pp. 57 - 72.

**GROSS**, Gaston (1988), «Structure des noms composés», *Actes du Colloque Informatique et Langue Naturelle*, Nantes, Université de Nantes, pp. 357 - 381.

**GROSS**, Gaston (1990), «Définition des noms composés dans un lexique-grammaire», *Dictionnaires électronique du français, Langue Française*, n° 87, sous la direction de Blandine Courtois, Max Silberstein, Paris, Larousse, pp. 83-90.

**GROSS**, Gaston (1991), «Des noms composés dans un dictionnaire électronique», *Les industries de la langue, Perspectives des années 1990, Tome II*, Montréal, Office de la Langue Française, Société des traducteurs du Québec, pp. 681 - 702.

**GROSS**, Gaston (1996), *Les expressions figées en français. Noms composés et autres locutions*, Coll. L'Essentiel Français, Paris, Ophrys, 61 p.

**GROSS**, Gaston (1997), «Du bon usage de la notion de locution», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 201 - 223.

**GUERRA**, Javier Pérez (1998), *Análisis computarizado de textos. Una introducción a TACT*, Vigo, Universidade de Vigo, Servicio de Publicacións, 274 p.

**HABERT**, Benoît, **JACQUEMIN**, Christian (1993), «Noms composés, termes, dénominations complexes: problématiques linguistiques et traitements automatiques», *Traitement Automatique des langues*, 34, 2, Revue de l'ATALA, Paris, pp. 5 - 42.

**HABERT**, Benoît, **SALEM**, André (1995), «L'utilisation de catégorisations multiples pour l'analyse quantitative de données textuelles», *Traitement Automatique des Langues*, 36, 1 -2, Paris, *Revue l'ATALA*, pp. 249 - 276.

**HABERT**, Benoît; **NAZARENKO**, Adeline; **SALEM**, André (1997), *Les linguistiques de corpus*, Paris, Armand Colin, 240 p.

**HABERT**, Benoît; **FABRE**, Cécile; **ISSAC**, Fabrice (1998), *De l'écrit au numérique, Constituer, normaliser et exploiter les corpus électroniques*, Paris, Masson, 320 p.

**HALLIDAY**, M.A.K.; **HASAN**, Ruqaiya (1976), *Cohesion in English*, London and New York, Longman, 374 p.

**HALLIDAY**, M.A.K. (1991), «Corpus Studies and probabilistic grammar», *English Corpus Linguistic*, Studies in Honour of Jan Svartvik, ed. by Karin Aijmer & Bengt Altenberg, London and New York, Longman, pp. 30 - 43.

**HAUSMANN**, Franz Josef (1989), «Le dictionnaire des collocations», *Wörterbücher, Ein internationales Handbuch für Lexicographie*, Herausgegeben von Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta, Berlin, New York, Walter de Gruyter, pp. 1010 -1019.

**HAUSMANN**, Franz Josef (1997), «Tout est idiomatique dans la langue», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp.277 -290.

**HEID**, Ulrich (1992), «Décrire les collocations - deux approches lexicographiques et leur application dans un outil informatisé», *Actes du colloque Anniversaire de l'ETI*, Genève, Université de Genève, 24 p.

**HEID**, Ulrich; **FREIBOTT**, Gerhard (1991), «Collocations dans une base de données terminologique et lexicale», *META*, Vol. 36, N° 1. mars, PUM, pp.77 - 91.

**HEID**, Ulrich (1992), «Notes on the use of lexical functions for the description of collocations in an NLP lexicon.», DRAFT, 12 p.

**HEID**, Ulrich (2001), «Collocations in Sublanguage Texts: Extraction from Corpora», *Handbook of Terminology Management, Application-Oriented terminology Management*, Volume 2, compiled by Sue Ellen Whright, Gerhard Budin, John Benjamins publishing Company, Amsterdam, Philadelphia, pp. 788 - 808.

**HERWIJNEN**, Eric van (1994), *Practical SGML*, Boston / Dordrecht / London, Kluwer Academic Publishers, 288 p.

**HYMES**, D. (1980, tradução), «Modèles pour l'interaction du langage et de la vie sociale», *Études de Linguistique Appliquée* 37, Paris, Didier Erudition, pp. 127 - 153.

**JEANDILLOU**, Jean-François (1997), *L'Analyse textuelle*, Armand Collin, Coll. Cursus, Paris, 192 p.

**JUNG**, René (1990), «Remarques sur la constitution du lexique des noms composés» *Dictionnaires électronique du français, Langue Française*, n° 87, sous la direction de Blandine Courtois, Max Silberstein, Paris, Larousse, pp. 71 - 83.

**KANDELAKI**, T.L. (1981), «Les sens des termes et les systèmes de sens en terminologies scientifiques et techniques», *Textes choisis de terminologie, I. Fondements théoriques de la terminologie*, dir. V.I. Siforov, Québec, GIRSTERM, pp. 135 - 184.

**KENNEDY**, Graeme (1998), *An introduction to Corpus Linguistics*, Longman, London and New York, 315 p.

**KJAER**, Anne Lise (1990), «Context-Conditioned Word Combinations in Legal Language», *Journal of the International Institute for Terminology Research – IIITF* – Vol. 1, no° 1-2, Wien, International Network for Terminology (TermNet), pp.21-32.

**KJAER**, Anne Lise (1990), «Phraseology Research – State-Of-The-Art», *Journal of the International Institute for Terminology Research – IIITF* – Vol. 1, no° 1-2, Wien, International Network for Terminology (TermNet), pp.3-20.

**KLEIBER**, Georges (1990), *La sémantique du Prototype. Catégories et sens lexical*, Paris, PUF, 199 p.

**KLEIBER**, Georges (1991), «Prototype et prototypes: encore une affaire de famille», *Sémantique et cognition. Catégories, prototypes, typicalité*, sous la direction de Danièle Dubois, Paris, CNRS édition, 130 p.

**KOCOUREK**, Rostislav (1991), «Textes et termes», *META*, Vol, 36, N° 1, mars 1991, PUM, Montréal, pp. 71 - 76.

**KOCOUREK**, Rostislav (1994/95), «Les textes spécialisés et la terminologie en tant qu'objet de l'analyse linguistique», *Terminologie et Linguistique de Spécialité. Étude de vocabulaires et textes spécialisés*, Halifax, ALFA, pp.9-16.

**KOCOUREK**, Rostislav; **ROUSSEAU**, Jean-Louis et collectif (1993), «Terminologie, discours et textes spécialisés», *Actes du XV<sup>e</sup> Congrès international des linguistes*, Vol. 1, Québec, Les Presses de l'Université Laval, pp. 187 - 212

**KUHN**, Thomas S. (1970), *La structure des révolutions scientifiques*, Flammarion, Paris, 284 p.

**KWON-PAK**, Song-Nim (1997), «Les Prépositions, révélateurs de polysémie nominale», *Unités lexicales et identité sémantique*, sous la direction de Pierre Cadiot, *Sémiotiques*, N° 13, Paris, Didier-Erudition, CNRS-INALF, pp. 31 - 40.

**LAMIZET**, Bernard (1992), *Les lieux de la communication*, Mardaga, Liège, 347 p.

**LAPORTE**, Éric (1988), «Reconnaissance des expressions figées lors de l'analyse automatique», *Langages*, n° 90, Les expressions figées, sous la direction de Laurence Danlos, Paris, Larousse, pp. 5 - 6.

**LAPORTE**, Éric (1997), «Les mots. Un demi-siècle de traitements», *Traitement Automatique des Langues*, 38, 2, Paris, Revue l'ATALA, pp. 47 - 68.

**LAPORTE**, Éric (2001), «Resolução de ambiguidades», *Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações*, Org. Elisabete Marques Ranchhod, Lisboa, Caminho, pp. 49 - 89.

**LASZLO**, Pierre (1995), *A Palavra das Coisas. A Linguagem da Química*, Coll. Ciência Aberta, Lisboa, Gradiva, 283.

**LE PESANT**, Denis (1996), «Principes d'organisation des données lexicales dans un dictionnaire électronique», in *Sémiotiques*, n° 11, Savoirs lexicaux et savoirs encyclopédiques, t.1 Légitimer la description lexicale, n° dirigé par David Piotrowski et Peter Stockinger, Paris, Didier Érudition, pp. 35 - 54

**LE PESANT**, Denis (1998), «Introduction aux classes d'objets», in *Langages*, n° 131, Les classes d'objets, Paris, Larousse, pp. 6 - 33.

**LEBART**, L.; **SALEM**, A. (1994), *Statistique Textuelle*, Paris, Dunod, 34 p.

**LEACOCK**, Claudia; **TOWELL**, Geoffrey; **VOORHEES**, Ellen V. (1996), «Towards building Contextual Representations of Word Senses Using

Statistical Models», *Corpus Processing for Lexical Acquisition*, ed. by Branimir Boguraev and James Pustejovsky, Cambridge, Massachusetts, London, England, The MIT Press, pp. 97 - 112.

LEECH, Geoffrey (1991), «The state of the arte», *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, London & New York, Longman, pp. 8 - 29.

LEECH, Geoffrey (1997), «Introducing corpus annotation», *Corpus Annotation, Linguistics Information from Computer Text Corpora*, ed. Roger Garside, Geoffrey Leech, Tony MacEnery, London & New York, Longman, pp.1 - 18.

LEECH, Geoffrey (1997), «Grammatical tagging», *Corpus Annotation, Linguistics Information from Computer Text Corpora*, ed. Roger Garside, Geoffrey Leech, Tony MacEnery, London & New York, Longman, pp.19 - 33.

LEECH, Geoffrey; GARSIDE, R.; BRYANT, M. (1994), «CLAWS4: The tagging of the British National Corpus», *Proceedings of the 15<sup>th</sup> International Conference on Computational Linguistics (COLING 94)*, Kyoto, Japan, pp.662 - 628.

LERAT, Pierre (1983), *Sémantique descriptive*, Paris, Hachette, 126 p.

LERAT, Pierre (1995), *Les langues spécialisées*, Paris, PUF, 201 p.

LERAT, Pierre (1996), «La notion de connaissance est-elle un concept opérationnel en linguistique?», Kleiber, G., Riegel, M. (éds) *Les formes du sens*, Études de linguistique française médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans, Louvain-la-Neuve, Editions Duculot, pp.241 - 248.

LINO, M. Teresa (1994), «Base de données textuelles ne terminographiques», *Meta* 39 / 4, Montréal, PUM, pp. 786 - 789.

**LINO**, M. Teresa (1994/95), «Corpus informatisés et travail en terminographie: expérience portugaise», *Terminologie et Linguistique de Spécialité. Étude de vocabulaires et textes spécialisés*, Halifax, ALFA, pp. 421 - 425.

**LOTTE**, D.S. (1981), « Principes d'établissement de la terminologie scientifique et technique», *Textes choisis de terminologie, I. Fondements théoriques de la terminologie*, dir. V.I. Siforov, Québec, GIRSTERM, pp. 1 - 54.

**LYONS**, John (1978), *Sémantique Linguistique*, Paris, Larousse, 495 p.

**MAINGUENEAU**, Dominique (1984), *Genèse du discours*, Mardaga, Coll. Philosophie et Langage, Liège, 209 p.

**MAINGUENEAU**, Dominique (1991), *L'analyse du discours. Introduction aux lectures de l'archive*, Paris, Hachette, 268 p.

**MAINGUENEAU**, Dominique (1992), «Le 'tour' ethnolinguistique de l'analyse du discours», *Langages* 105, Larousse, Paris, pp. 114 - 125.

**MAINGUENEAU**, Dominique (1996), *Les termes clés de l'analyse du discours*, Editions su Seuil, Coll. Mémo, Paris, 94 p.

**MAINGUENEAU**, Dominique (1998), *Analyser les textes de communication*, Paris, Dunod, 211 p.

**MAINGUENEAU**, Dominique, **COSSUTTA**, Frédéric (1995), «L'analyse des discours constituants», *Langages* n°117, Larousse, Paris, pp. 112-125.

**MALACA**, Jean Malaca (19981), *Sintaxe transformacional do adjectivo*, Lisboa, INIC, 561 p.

**MANUELITO**, Helena (1995), *A terminologia da franchia, franchise ou franchising - a problemática da ambiguidade num vocabulário não harmonizado*, Dissertação de Mestrado, Lisboa, Universidade Nova de Lisboa, 316 p.

**MARQUEZ**, Eduardo (1998), «Classification des adjectifs: étude exploratoire sur l'organisation sémantique-pragmatique des adjectifs», *Langages* n° 132, Larousse, Paris, pp. 87-107.

**MARTIN**, Éveline (1993), *Reconnaissance de contextes thématiques dans un corpus textuel. Éléments de lexico-sémantique*, Paris, Hachette, 283 p.

**MARTIN**, Robert (1983), *Pour une logique du sens*, Paris, PUF, 268 p.

**MARTIN**, Robert (1990), «La Définition Naturelle», in Actes du Colloque *La Définition*, Paris, Larousse, pp. 86 - 95.

**MARTIN**, Robert (1997), « Sur les facteurs du figement lexical », *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 291 - 306.

**MARTIN**, Willy, **HEID**, Ulrich (2001), «Frame-based definitions and the selection of multiword term candidates in DOT», *Actes des quatrièmes rencontres "Terminologie et Intelligence Artificielle*, Nancy, Unité de Recherche et Innovation. INIST - CNRS, pp. 55 - 65.

**MATEUS**, M. Helena; **BRITO**, Ana Maria; **DUARTE**, Inês; **FARIA**, Isabel Hub (1989), *Gramática da Língua Portuguesa*, Lisboa, Caminho, 417 p.

**MATHIEU-COLAS**, Michel (1994), *Les mots à traits d'union. Problèmes de Lexicographie informatique*, Paris, Hachette, 351 p.

**MATHIEU-COLAS**, Michel (1996), «Typologie des noms composés français», *Cahiers de Lexicologie*, 69, 1996-2, Paris, Didier Erudition, pp. 71 - 125.

**McENERY**, Tony, **WILSON**, Andrew (1996), *Corpus Linguistics*, Ed. Tony McEnery And Andrew Wilson, Manchester, Edinburgh University Press, 207 p.

**McENERY**, Tony, **LANGÉ**, Jean-Marc, **OAKES**, Micheal, **VÉRONIS**, Jean (1997), «The exploitation of multilingual annotated corpora for term extraction», *Corpus Annotation, Linguistics Information from Computer Text Corpora*, ed. Roger Garside, Geoffrey Leech, Tony MacEnery, London & New York, Longman, pp. 220 - 230.

**MEJRI**, Salah (1997), «Binarisme, dualité et séquences figées», Kleiber, G., Riegel, M. (éds) *Les formes du sens*, Études de linguistique française médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans, Louvain-la-Neuve, Editions Duculot, pp. 249 - 256.

**MEJRI**, Salah (1998), «Le figement et la linéarité du signe linguistique», *L'information grammaticale*, Numéro spécial Tunisie, sous la direction de Taïeb Baccouche et Salah Mejri, Paris, Société pour l'information grammaticale, pp. 17-21.

**MEJRI**, Salah (1998), «La conceptualisation dans les séquences figées», *L'information grammaticale*, Numéro spécial Tunisie, sous la direction de Taïeb Baccouche et Salah Mejri, Paris, Société pour l'information grammaticale, pp. 41 - 48.

**MEL'CUK**, Igor A. (1993), *Cours de Morphologie Générale (Théorique et Descriptive)*, Vol.1.: *Introduction et Première partie: Le mot*, Montréal, PUM, CNRS Éditions, 412 p.

**MEL'CUK**, Igor A. (1994). *Cours de Morphologie Générale (Théorique et Descriptive)*, Vol.2.: *Deuxième partie: Significations morphologiques*, Montréal, PUM, CNRS Éditions, 456.

**MEL'CUK**, Igor A.; CLAS, André; POLGUÈRE, Alain (1995), *Introduction à la Lexicologie explicative et combinatoire*, Coll. Champs Linguistiques, Manuels, Louvain-la-Neuve, Duculot, 256.

**MEL'CUK**, Igor A (1996), «Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon,» *Lexical Functions in Lexicography and Natural Language Processing*, Ed. by Leo Wanner, Amsterdam / Philadelphia, John Benjamins, pp. 37 - 102.

**MEL'CUK**, Igor A (1998), «Collocations and Lexical Functions», *Phraseology, Theory, Analysis, and Applications*, Ed. A. P. Cowie, Oxford, Oxford University Press, pp. 23 - 54.

**MENÉNDEZ**, Fernanda Miranda (1997), *A «Construção do Discurso Setecentista. Dos Processos Discursivos à História da Língua*, Dissertação de Doutorado, Universidade Nova de Lisboa, 482 p.

**MERIALDO**, Bernard (1995), «Modèles probabilistes et étiquetage automatique», *Traitement Automatique des Langues*, 36, 1- 2, Paris, Revue l'ATALA, pp. 7 - 22.

**MIKKELSEN**, Hans Kristian (1990), «Two New Bibliographies about Special Language Research», in *Journal of the International Institute for Terminology Research – IIITF – Vol. 1, no<sup>o</sup> 1-2*, Wien, International Network for Terminology (TermNet), pp. 90-93.

**MOIRAND**, Sophie (1995), «L'évaluation dans les discours scientifiques et professionnels», *Les enjeux di discours spécialisés*, coordination éditorale, Jean-Claude Beacco, Sophie Moirand, Paris, Presses de la Sorbonne Nouvelle, pp. 81 - 94.

**MOLES**, Abraham A. (1995), *Les sciences de l'imprécis*, Paris, Éditions du Seuil, 360 p.

**MORTUREUX**, Marie-Françoise (1995), «Les vocabulaires scientifiques et techniques», *Les enjeux du discours spécialisés*, coordination éditoriale, Jean-Claude Beacco, Sophie Moirand, Paris, Presses de la Sorbonne Nouvelle, pp. 13-26.

**MULLER**, Charles (1992), *Initiation aux Méthodes de la Statistique Linguistique*, Réimpression de l'édition Hachette de 1973, Paris, Honoré Champion, 185 p.

**MULLER**, Charles (1992), *Principes et Méthodes de Statistique Lexicale*, Réimpression de l'édition Hachette de 1977, Paris, Honoré Champion, 206 p.

**NORME internationale** (2000), *Travaux terminologiques - Vocabulaire, Partie 1: Théorie et application*, Numéro de référence ISO/FDIS 1087 - 1: 2000 (E/F), Genève, ISO, 30 p.

**NORME internationale** (2000), *Travaux terminologiques - Vocabulaire, Partie 2: Applications logicielles*, Numéro de référence ISO/FDIS 1087 - 2: 2000 (E/F), Genève, ISO, 26 p.

**NEDOBITY**, Wolfgang (1990) , «Simple Phrase Structure Grammars and their application in Terminology», in *Journal of the International Institute for Terminology Research – IIITF – Vol. 1, no° 1-2*, Wien, International Network for Terminology (TermNet), pp. 59-63.

**NKWENTI-AZEH**, Blaise (1994), «Positional and combinational characteristics of terms: consequences for corpus-based terminography», *TERMINOLOGY*, vol 1 (1), Amsterdam / Philadelphia, John Benjamins Publishing Company, pp. 61 - 95.

**OESER**, Erhard; **BUDIN**, Gerhard (1999), «Grundlagen der Terminologiewissenschaft», in Hoffmann, L.; Kalver-Kämper, H.; Wiegand, H. E. [eds.]. *Fachsprachen. / Languages for Special Purposes. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin / New York, deGruyter, pp. 2172 -2183.

**PEARSON**, Jennifer (1998), *Terms in Context*, Coll. Studies in Corpus Linguistics, Amsterdam / Philadelphia, John Benjamins Publishing Company, 242 p.

**PÉRY-WOODLEY**, Marie-Paule (1995), «Quels corpus pour quels traitements automatiques», *Traitement Automatique des Langues*, 36, 1 -2, Paris, Revue l'ATALA, pp. 213 - 232.

**PETITJEAN**, André (1989), «Les typologies textuelles», *Pratiques* 62, pp.86 - 125.

**PHAL**, André. (1971), *Vocabulaire général d'orientation scientifique (V.G.O.S.). Part du lexique commun dans l'expression scientifique*, École normale Supérieure de Saint-Cloud, Centre de Recherche et d'Étude pour la Diffusion du Français, Paris, CRÉDIF, 90 p.

**PICHT**, Herbert (1990), «A Study of LSP Phraseological Elements in Spanish Technical Texts», in *Journal of the International Institute for Terminology Research – IIITF – Vol. 1, no° 1-2*, Wien, International Network for Terminology (TermNet), pp.48–58.

**PICHT**, Herbert (1990), «LSP Phraseology from the Terminological Point of view», in *Journal of the International Institute for Terminology Research – IIITF – Vol. 1, no° 1-2*, Wien, International Network for Terminology (TermNet), pp.32-48.

**PIERREL**, Jean-Marie; **ROMARY** Laurent (1996), «Lexique, langue et tâche dans le dialogue homme-machine finalisé», in *Sémiotiques*, n° 11, Savoirs

lexicaux et savoirs encyclopédiques, t.1 Légitimer la description lexicale, n° dirigé par David Piotrowski et Peter Stockinger, Paris, Didier Érudition, pp. 95-117.

**PLANTE, Pierre, DAVID, Sophie** (1991), «Le progiciel Termino: de la nécessité d'une analyse morphosyntaxique pour le dépouillement terminologique des textes», *Les industries de la langue, Perspectives des années 1990*, Office de la langue française, Société des traducteurs du Québec, Montréal, 21-24 novembre 1990, tome I, pp 71 - 88.

**PLUNGER, Christine** (1990), «Synoptic Analysis of the Workshop on Phraseology, LSP and Terminology, November 24 and 25, 1989, Vienna», in *Journal of the International Institute for Terminology Research – IIITF – Vol. 1*, no° 1-2, Wien, International Network for Terminology (TermNet), pp.87-89.

**POTTIER, B.** (1967), *Présentation de la linguistique. Fondements d'une théorie*, Paris, Éditions Klincksieck, 78 p.

**POTTIER, B.** (1987), *Théorie et analyse en linguistique*, Coll. Langue Linguistique Communication, Paris, Hachette, 224.

**POTTIER, Bernard** (1992), *Sémantique Générale*, Coll. Linguistique Nouvelle, Paris, PUF, 237.

**QUEMADA, Bernard** (1992), «Avant-propos», *Dictionnaire et Lexicographie. FRANTEXT: Autour d'une base de données textuelles. Témoignages d'utilisateurs et voies nouvelles*, Vol 2, coll. dirigée par Bernard Quemada, Paris, Didier Erudition, pp. 9 - 15.

**RABATEL, Alain** (1991), *La construction textuelle du point de vue*, Lausanne, Delachaux et Niestlé, 202 p.

**RANCHOD**, Elisabete Marques (2001), «O uso de dicionários e de autómatos finitos na representação lexical das línguas naturais», *Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações*, Org. Elisabete Marques Ranchhod, Lisboa, Caminho, pp.13 - 47.

**RASTIER**, François (1987), *Sémantique interprétative*, Paris, PUF, 284 p.

**RASTIER**, François (1989), *Sens et textualité*, Paris, Hachette, 286 p.

**RASTIER**, François (1991), *Sémantique et recherches cognitives*, Paris, PUF, 262 p.

**RASTIER**, François (1997), «Défigements sémantiques en contexte», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 307 - 323.

**RASTIER**, François (1998), «Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage», *Langages* 129, Larousse, pp. 97 - 111.

**RASTIER**, François; **CAVAZZA**, Marc; **ABEILLÉ**, Anne (1994), *Sémantique pour l'analyse. De la linguistique à l'informatique*, Paris, Masson, 240 p.

**REBOUL**, Anna; **MOESCHLER**, Jacques (1998), *Pragmatique du discours. De l'interprétation de l'énoncé à l'interprétation du discours*, Paris, Armand Colin, 220 p.

**RENOUF**, Antoinette; **SINCLAIR**, John (1991), «Collocational frameworks in English», *English Corpus Linguistic*, Studies in Honour of Jan Svartvik, ed. by Karin Aijmer & Bengt Altenberg, London and New York, Longman, pp. 128 - 143.

**REY**, Alain (1976), *Théories du signe et du sens, Lectures II*, Éditions Klincksieck, Paris, 408 p.

**REY**, Alain (1979), *La terminologie: noms et notions*, Paris, PUF 127 p.

REY, Alain (1997), «Phraséologie et pragmatique», *La Locution entre langue et usages*, Fontenay Saint-Cloud, ENS Éditions, pp. 333 -346.

REY, Alain (1997), «Préface», *Dictionnaires des Expressions et Locutions*, Le Robert, Coll. "les usuels", Le Robert, Paris, 888 p.

RICOEUR, Paul (1986), *Du texte à l'action. Essais d'herméneutique*, II, Seuil, Coll. Esprit / Seuil, Paris, 409 p.

RICOEUR, Paul (1990), *Soi-même comme un autre*, Éditions du Seuil, Coll. Points, Série Essais, Paris, 425 p.

RITCHIE, Graeme D.; RUSSELL, Graham, J.; BLACK, Alan W. et al. (1992), *Computational Morphology. Pratical Mechanisms for the English Lexicon*, London, The MIT Press, 291 p.

ROBERTS, Roda P (1993), «La phraséologie: état des recherches», *Terminologies Nouvelles*, N° 10, Bruxelles, RINT, pp. 36 - 42.

ROBERTS, Roda P. (1994/95), «Identifying the Phraseology of Language for Special Purposes», *Terminologie et Linguistique de Spécialité, Études des vocabulaires et textes spécialisés*, ALFA, volume 7/8, Halifax, pp, 61-74

ROUSSEAU, Louis-Jean (1993), «Terminologie et phraséologie, deux composantes indissociables des langues de spécialités.», *Terminologies Nouvelles*, N° 10, Bruxelles, RINT, pp. 9 - 11.

ROUSSEAU, Louis-Jean (1999), «De nouvelles avenues pour la terminologie», N° 19, Bruxelles, Revue coédité par l'Agence de la Francophonie et Communauté française de Belgique, pp. 3 - 5.

**SABAH**, Gérard (1997), «Le sens dans le Traitement Automatique des Langues», *Traitement Automatique des Langues*, 38, 2, Paris, Revue l'ATALA, pp. 91 - 134.

**SAGER**, Juan (1990), *A Pratical Course in Terminology Processing*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 254 p.

**SAGER**, Juan (2000), «Pour une approche fonctionnelle de la terminologie», *Le sens en terminologie*, sous la dir. d'Henri Béjoint et Philippe Thoiron, Lyon, PUL, pp. 40 - 53.

**SANROMÁN**, Álvaro Iriarte (2001), *A unidade lexicográfica. Palavras, Colocações, Frasemas, Pragmatemas*, Braga, Centro de Estudos Humanísticos, Colecção Polierdo 5, Universidade do Minho, 394 p.

**SANTOS**, Diana (2001), «Processamento da linguagem natural: uma apresentação através das aplicações», *Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações*, Org. Elisabete Marques Ranchhod, Lisboa, Caminho, pp. 229 - 260.

**SCHAEFFER**, Jean-Marie (1986), «Du texte au genre», *Théorie des genres*, Paris, Seuil, pp. 179 - 205.

**SCHEMANN**, Hans (1989), «Das phraseologische Wörterbuch» *Ein internationales Handbuch für Lexicographie*, Herausgegeben von Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta, Berlin, New York, Walter de Gruyter, pp. 1019 -1032.

**SCHMITZ**, Klaus-Dirk (1999), «Computergestützte Terminographie: Systeme und Anwendungen», in Hoffmann, L.; Kalver-Kämper. H.; Wiegand, H. E. [eds.]. *Fachsprachen. / Languages for Special Purposes. Ein internationales*

*Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin / New York, deGruyter, pp. 2164 -2170.

**SCHÖNEFELD**, Doris (1999), «Corpus Linguistics and Cognitivism», *International Journal of Corpus Linguistics*, Vol 4, Amsterdam / Philadelphia, John Benjamins Publishing, pp. 137 - 171.

**SEEWALD**, Uta (1993), «Traitement morphosémantique des noms composés de structure N - N à partir de l'analyse morphosémantique des mots simples», *Traitement Automatique des Langues*, 34, 2, Paris, Revue l'ATALA, pp. 75 - 88.

**SIFOROV**, V.I. (1981), «Problèmes de terminologie scientifique et technique», *Textes choisis de terminologie, I. Fondements théoriques de la terminologie*, dir. V.I. Siforov, Québec, GIRSTERM, pp. 301 - 316.

**SILBERSTEIN**, Max (1990), «Le dictionnaire électronique des mots composés», *Dictionnaires électronique du français, Langue Française*, n° 87, sous la direction de Blandine Courtois, Max Silberstein, Paris, Larousse, pp. 71 - 83.

**SILBERSTEIN**, Max (1993), *Dictionnaires électroniques et analyse automatique des textes. Le système INTEX*, Paris, Masson, 233 p.

**SINCLAIR**, John (1991), *Corpus, concordance, collocation*, Oxford University Press, Oxford, 179 p.

**SINCLAIR**, John (1996), *EAGLES: Preliminary Recommendations on Corpus Typology (EAG -- TCWG --CTYP/P)*, Version of May, pp. 25, disponible em <http://www.ilc.pi.cnr.it>.

**SINCLAIR**, John; **BALL**, J, (1996), *EAGLES: Preliminary Recommendations on Text Typology (EAG -- TCWG -- TYP/P)*, Version of Jun, pp. 71, disponible em <http://www.ilc.pi.cnr.it>.

**SOUSA SANTOS**, Boaventura de (1987), *Um Discurso sobre as Ciências*, Lisboa, Afrontamento, 58 p.

**SOUSA SANTOS**, Boaventura de (1993), *Introdução a uma Ciência Pós-Moderna*, Lisboa, Afrontamento, 199 p.

**SPERBERG**, C.M. e **BURNARD**, Lou (1994), *Guidelines for Electronic Text Encoding and Interchange, TEI P3, Text Encoding Initiative*, ACH, ACL, ALLC, Chicago, Oxford, 1290 p.

**STA**, Jean-David (1995), «Comportement statistique des termes et acquisition terminologique à partir de corpus», *Traitement Automatique des Langues*, 36, 1 -2, Paris, Revue l'ATALA, pp. 119 - 132.

**STEIN**, Achim, **SCHMID**, Helmut (1995), «Étiquetage morphologique de textes français avec un arbre de décisions», *Traitement Automatique des Langues*, 36, 1 -2, Paris, Revue l'ATALA, pp. 23 - 36.

**THOIRON**, Philippe, **ARNAUD**, P., **BÉJOINT**, Henri & **BOISSON**, C.P. (1996), «Notion d' «archi-concept» et dénomination», *La Dénomination, Méta*, Numéro spécial, vol. 41, n° 4, Montréal, PUM, pp. 512-524.

**VARELA**, Fernando, **KUBARTH**, Hugo (1994), *Diccionario Fraseológico del Español Moderno*, Madrid, Gredos, 296 p.

**VÉRONIS**, Jean; **KHOURI**, Liliane (1995), «Étiquetage grammatical multilingue: le projet MULTEXT», *Traitement Automatique des Langues*, 36, 1 -2, Paris, Revue l'ATALA, pp. 233-248.

**VIVÉS**, Robert (1990), «Les composés nominaux par juxtaposition», *Dictionnaires électronique du français, Langue Française*, n° 87, sous la direction de Blandine Courtois, Max Silberstein, Paris, Larousse, pp. 98 - 103..

**WEHRLI**, Éric (1997), *L'Analyse Syntaxique des Langues Naturelles. Problèmes et Méthodes*, Paris, Masson, 249 p.

**WÜSTER**, Eugen (1981), «L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses», *Textes choisis de terminologie, I. Fondements théoriques de la terminologie*, dir. V.I. Siforov, Québec, GIRSTERM, pp.55- 114.

**WÜSTER**, Eugen (1985), *Einführung in die allgemeine Terminologielehre und terminologische Lexicographie*, Copenhagen, The Copenhagen School of Economics, 213 p.

**WYNNE**, Martin (1994), *A Post-Editor's Guide to CLAWS 7 Tagging*, disponible em <http://www.comp.lancs.ac.uk/computing/users/eiamjw/claws7>

## **Bibliografia da Detecção Remota**

**ATKINSON**, Peter M., **TATE**, Nicholas J., dir., (1999), *Advances in Remote Sensing and GIS Analysis*, Chichester, John Wiley & Sons, 273 p.

**BONN**, Ferdinand, dir., (1996), *Précis de Télédétection. Applications Thématiques*, vol.2, Québec, Presses de l'Université du Québec / AUPELF-UREF, 633 p.

**BONN**, Ferdinand, **ROCHON**, Guy (1992), *Précis de Télédétection. Principes et Méthodes*, vol. 1, Québec, Presses de l'Université du Québec / AUPELF-UREF, 485 p.

**CAMPBELL**, James B. (1996), *Introduction to Remote Sensing*, 2<sup>a</sup> ed., London, Taylor & Francis, 622 p.

**CHUVIECO**, Emilio (1996), *Fundamentos de Teledetección Espacial*, 3<sup>a</sup> ed., Madrid, RIALP, 568 p.

**GIRARD**, Michel-Claude, **GIRARD**, Colette M. (1999), *Traitement des Données de Télédétection*, Paris, Dunod, 529 p.

**DONNAY**, Jean-Paul, **BARNSELY**, Mike J., **LONGLEY**, Paul A., dir. (2001), *Remote Sensing and Urban Analysis*, London, Taylor & Francis, 268 p.

**JÄHNE**, Bernd (1997), *Digital Image Processing. Concepts, Algorithms, and Scientific Applications*, 4<sup>a</sup> ed., Berlin, Springer, 555 p.

**JENSEN**, J. R. (1996), *Introductory Digital Image Processing: A Remote Sensing Perspective*, 2<sup>a</sup> ed., Englewood Cliffs, Prentice-Hall, 330 p.

**LILLESAND**, Thomas M., **KIEFER**, Ralph W. (2000), *Remote Sensing and Image Interpretation*, 4<sup>a</sup> ed., New York, John Wiley & Sons, 724 p.

**LLIBOUTRY**, Louis (1992), *Sciences Géométriques et Télédétection*, Paris, Masson, 289 p.

**LUNETTA**, Ross S., **ELVIDGE**, Christopher D., dir., (1999), *Remote Sensing Change Detection. Environmental Monitoring, Methods and Applications*, London, Taylor & Francis, 318 p.

**MARION**, André (1987), *Introduction aux Techniques de Traitement d'Images*, Paris, Eyrolles, 278 p.

**MATHER**, Paul M. (1996), *Computer Processing of Remotely-Sensed Images. An Introduction*, Chichester, John Wiley & Sons, 352 p.

**PAUL**, Serge, **DEPECKER**, Loïc, **GOILLOT**, Charles, **LENCO**, Michel (1991), *Introduction à l'Étude de la Télédétection Aérospatiale et de son Vocabulaire*, Paris, La Documentation Française, 316 p.

**PAUL**, S. et al. (1997), *Terminologie de Télédétection et Photogrammétrie*, Paris, Conseil International de la Langue Française, 455 p.

**PINILLA**, Carlos R. (1995), *Elementos de Teledetección*, Madrid, Ra-ma, 313p.

**RICHARDS**, John A. (1995), *Remote Sensing Digital Image Analysis. An Introduction*, 2<sup>e</sup> ed., Berlin, Springer-Verlag, 340 p.

**ROBIN**, Marc (1995), *La Télédétection. Des Satellites aux Systèmes d'Information Géographiques*, Paris, Nathan, 318 p.

**SABINS**, Floyd F. (1997), *Remote Sensing. Principles and Interpretation*, 3<sup>e</sup> ed., New York, W. H. Freeman and Company, 494 p.

## **Corpus**

**ANDRADE**, E. Linhares de (1992), «Inventário de terreno, contacto com a área teste de Benavente (Coruche)», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 16 p.

**BARATA**, Teresa (1995), «Tratamento Digital de Imagens de Detecção Remota» in *Geostatística, Detecção Remota e Sistemas de Informação Geográfica Aplicados à Gestão do Ambiente e Recursos Naturais*, Seminário COMMETT, Lisboa, Instituto Superior Técnico, 27 p.

**CAETANO**, Mário, **SANTOS**, João, **P. NAVARRO**, Ana (1997), «Uma Metodologia Integrada para a Produção de Cartas de Uso do Solo utilizando Imagens de Satélite e Informação Geo-Referenciada não Espectral», *Cartografia e Cadastro*, Lisboa, Publicação do Instituto Português de Cartografia e Cadastro, pp. 71 - 78.

**CASACA**, João (1993), *Movimento orbital de satélites artificiais da terra*, Lisboa, LNEC, 21 p.

**CASACA**, João (1994), «A observação da terra com imagens numéricas multiespectrais», *Ingenium, Revista da Ordem os Engenheiros*, Lisboa, pp. 13-25.

**CATALÃO**, J., **REBORDÃO**, J. M. (1989), «Utilização dos modelos digitais do terreno no contexto das aplicações da detecção remota», *Congresso 89, Pela Valorização Profissional do Engenheiro, Cartografia Automática e Sistemas de Informação Geográfica*, Coimbra, pp. 1 - 14.

**CORREIA**, Maria dos Santos, **BESSA**, Maria João, **FERNANDES**, Delfim (s.d.), «Contributo da Imagem de Satélite na Cartografia Temática», CING,

Lisboa, Ministério do Planeamento do Território, Secretaria de Estado da Administração Local e de Ordenamento do Território, 5 p.

**DIRECÇÃO de Serviços de Estatísticas Agrícolas** (1992), «Projecto Piloto da Teledetecção às Estatísticas Agrícolas. Inventários de Terreno. Preparação do Inventário de um Segmento», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 20 pp.

**DAVEAU**, Suzanne (1982), «A Utilização possível dos registos de satélites em infravermelho térmico na climatologia regional», *II Colóquio Ibérico de Geografia*, Lisboa, pp. 253 - 263.

**FERNANDES**, João Cordeiro (1994), «A Detecção Remota na avaliação do uso e ocupação do solo em meio rural», *Ingenium, Revista da Ordem os Engenheiros*, Lisboa, pp. 114 - 118.

**FERNANDES**, João Manuel Cordeiro (1989), «Elaboração duma carta temática a partir de imagens de satélite», *Congresso 89, Pela Valorização Profissional do Engenheiro, Cartografia Automática e Sistemas de Informação Geográfica*, Coimbra, 8 p.

**FONSECA**, Ana Maria (1989), *Aquisição de Informação Geográfica por sensores remotos orbitais, seu processamento e exploração*, Trabalho de síntese apresentado a concurso de acesso à categoria de assistente de investigação do LNEC, Departamento de Barragens, Núcleo de Medidas Geodésicas, Ministério das Obras Públicas, Transportes e Comunicações, Lisboa, LNEC, 94 p.

**FONSECA**, Ana Maria (1994), *Utilização de imagens numéricas multiespectrais no controlo da reserva nacional*, Lisboa, LNEC, 17 p.

**FONSECA**, Ana Maria (1997), *Melhoramento de uma classificação de ocupação do solo urbano utilizando informação textural*, Lisboa, LNEC, 9 p.

**FONSECA**, Ana Maria (1998), *Análise quantitativa de imagens: utilização de imagens numéricas multiespectrais no controlo do desenvolvimento urbano*, Lisboa, LNEC, 278 p.

**FONSECA**, Ana Maria, **RIBEIRO**, Helena Cristina, **LEAL**, João (1992), *Contribuição das imagens orbitadas por sensores remotos orbitais para actualização de cartografia à escala 1: 50 000*, Lisboa, LNEC, 23 p.

**HENRIQUES**, Rui Gonçalves (1982), *Técnicas de Detecção Remota*, Ministério da Habitação, Obras Públicas e Transportes, Lisboa, LNEC, 35 p.

**HENRIQUES**, Rui Gonçalves (1985), *Aplicação de Técnicas de Detecção Remota a problemas de recursos hídricos*, Lisboa, LNEC, 65 p.

**HENRIQUES**, Rui Gonçalves, **BESSA**, Maria João Souto (1992), «A Cartografia Corine Land Cover e o Método de Abordagem da Estratificação e Amostragem Areolar no âmbito dos Inventários Regionais», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 16 p.

**LE GORGEU**, J. P. (1992), «Sistemas de Observação da Terra, Satélites características das imagens, processos de aquisição», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 7 p.

**MADUREIRA**, Maria Ofélia; **ALBUQUERQUE**, Maira Ragina (1991), «Cartografia biofísica de Portugal com base na interpretação de imagens de satélite "Landast"», *Comunicações do Instituto de Investigação Científica*

*Tropical*, Lisboa, Ministério do Planamento e de Administração do Território, Secretaria de Estado da Ciência e Tecnologia, pp. 5 - 18.

**MEYER-ROUX**, Jean (1992), «Introdução geral, Apresentação do Projecto MARS», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 4 p.

**MINISTÉRIO DO EQUIPAMENTO SOCIAL**, (1984), «Caracertização fisiográfica, estudo hidrogeológico das regiões do Alentejo e Ordenamento do Território utilizando técnicas de detecção remota», *Aplicação de Técnicas de Detecção Remota nas regiões do Algarve e do Alentejo*, Lisboa, LNEC, 23 p.

**MOREIRA**, M. Eugénia S. A. (1982), «Ensaio de Cartografia Automática da Foz do Guadiana a partir de registos Landast 1», *II Colóquio Ibérico de Geografia*, Lisboa, pp. 291 - 302.

**MOREIRA**, M. Eugénia S. A. (1989), *Cartografia Temática do Arrozal nos cursos inferiores de alguns rios portugueses*, Lisboa, LNEC, 43 p.

**MOREIRA**, M. Eugénia S. A.; **COELHO**, Valentina; **SILVESTRE**, Sabino; **DIAS**, Cristina; **OLIVEIRA** Eduardo; **GONÇALVES**, Henriques e alii (1993), *Litoral do Algarve: Interpretação de imagens digitalizadas obtidas por satélites LANDSAT*, Lisboa, LNEC, pp. 53 - 68.

**MOREIRA**, M. Eugénia S. A., **OLIVEIRA**, Eduardo M. (1988), *Estuário do Sado: identificação de unidades fisiográficas a partir de imagens LANDSAT*, Lisboa, LNEC, 47 p.

**MORGADO**, Ana Maria (1997), «Fotogrametria Digital - Automatização da Orientação de Imagens Digitais», *Cartografia e Cadastro*, Lisboa, Publicação do Instituto Português de Cartografia e Cadastro, pp. 3 - 11.

**MUTHMANN, R.** (1992), «Integração da Teledetecção nos Modelos do Eurostat», in *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 6 p.

**NETO, Francelina A.** (1997), «Projecto de um Sistema Automático de Integração de Imagens e Dados Cartográficos na Detecção de Alterações da Superfície Terrestre», *Cartografia e Cadastro*, Lisboa, Publicação do Instituto Português de Cartografia e Cadastro, pp. 63 - 70.

**OLIVEIRA, Eduardo** (s.d.), «Técnicas numéricas de processamento de imagens», Lisboa, LNEC. 17 p.

**OLIVEIRA, Tiago, PAÚL, José C., PEREIRA, José** (1994), «Cartografia de formações arbustivas empregando sistemas de informação geográfica e detecção remota», *BIG, Boletim de Informação Geográfica*, Lisboa, USIG, pp. 10 - 12.

**PERDIGÃO, António** (1992), «Inventários de terreno, metodologia, documentos de terreno, instruções, nomenclatura» in *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 13 p.

**PERDIGÃO, Vanda** (1992), «Estratificação, Etapa Preparatória da Constituição da Amostra», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 8 p.

**PINA, Pedro** (1995), «Contribuição da Morfologia Matemática no Tratamento de Imagens de Detecção Remota», *Geostatística, Detecção Remota e Sistemas de Informação Geográfica Aplicados à Gestão do Ambiente e Recursos Naturais*, Seminário COMMETT, Lisboa, Instituto Superior Técnico, 52 p.

**PORTUGAL**, Jorge Marini (1984/1986), *Notas para um Curso sobre Cartografia e Teledetecção no Ordenamento do Território*, Lisboa, FCSH / UNL, 82 p.

**PORTUGAL**, Jorge Marini (1992), *Introdução às Tecnologias de levantamento de Informação Geográfica Física*, Lisboa, Edição do autor, 171 p.

**RAMALHO**, José; **JAN**, Lina; **PERDIGÃO**, Maria Vanda; **OLIVEIRA**, Eduardo; **HENRIQUES**, Rui Gonçalves (1988), *Caracterização Fisiográfica da Região do Alentejo. Classificação do Uso do Solo na Zona de Estremoz, Borba e Vila Viçosa*, Lisboa, LNEC, 27 p.

**RAMALHO**, José; **JAN**, Lina; **PERDIGÃO**, Maria Vanda; **OLIVEIRA**, Eduardo; **HENRIQUES**, Rui Gonçalves (1988), *Caracterização Fisiográfica da Região do Alentejo. Classificação do Uso do Solo na Sub-região de Mértola*, Lisboa, LNEC, 13 p.

**REBORDÃO**, J.M. (1986), «Dificuldades na Interpretação de Dados de Detecção Remota em Sistemas de Informação Geográfica», *Simpósio Ordenamento do Território Assistido por Computador*, Relatório Final, Lisboa, Ministério do Plano e da Administração do Território, pp. 82 - 94.

**REBORDÃO**, J.M. (1988), «Foto-interpretação assistida por computador de dados digitalizados: aplicação a contagem de espécies quase-regularmente distribuída», in *Photo-Interprétation*, Paris, Éditions Technip, pp. 33 - 40.

**REBORDÃO**, J.M., **HENRIQUES**, R. G. (s.d.), *Recursos Naturais - Detecção Remota*, CNIG, Ministério do Planeamento e de Administração do Território, Lisboa, Ministério do Planeamento e da Administração do Território, 15 p.

**ROMANA**, J.M., **REBORDÃO**, J.M., **MELO**, F. (1989), «Imagens SPOT e Índices temáticos aplicados à Cartografia da região do rio Tejo (Portugal)», *Photo-Interprétation*, Paris, Éditions Technip, pp. 31 - 32.

**ROSA**, Manuel (1992), «Programa Português de Utilização da Teledetecção», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 19 p.

**SILVA**, Erivaldo António de (1997), «Extracção de Feições Cartográficas de Imagens Multiespectrais Fundidas Usando Morfologia Matemática», *Cartografia e Cadastro*, Lisboa, Publicação do Instituto Português de Cartografia e Cadastro, pp. 63 - 70.

**SILVA**, João Manuel Romana da (1987), «Influência da Atmosfera na estrutura local da imagem satélite. (Calibração e correcção atmosférica de imagens)», Lisboa, Relatório de Estágio científico do curso de Física da Faculdade de Ciências de Lisboa, 132 p.

**STAKENBORG**, J. (1992), «Processamento de Dados para estatísticas agrícolas», *Teledetecção Aplicada às Estatísticas Agrícolas. Inquéritos por áreas de Sondagem*, Lisboa, Commission of the European Communities , Joint Research Centre, Ispra-Italy, 6 p.

**TENEDÓRIO**, José António (1989), *Concepção de uso e de evolução de uso do solo por interpretação de fotografia aérea vertical. Almada: exemplo metodológico*, Centro de Estudos de Geografia e Planeamento Regional, Lisboa, FCSH / UNL, INIC, 79 p.

**TENEDÓRIO**, José António (1995), «Teledetecção em Geografia: extracção automática do contorno da mancha construída a partir do processamento digital de imagens SPOT HRV», *Actas do II Congresso da Geografia Portuguesa*, Vol. I, Lisboa, APG, pp. 41 - 48.

**TENEDÓRIO**, José António (1995), *Introdução à Teledetecção. Manual teórico*, Lisboa, Geovisão / CEFA, 34 p.

**TENEDÓRIO**, José António (1995), «Teledetecção em áreas periurbanas: combinação de índices temáticos para localizar a mudança de uso do solo com recurso às imagens digitais», *Actas do VI Colóquio Ibérico de Geografia*, Vol. III, Porto, Universidade do Porto, pp. 1241 - 1249

**TENEDÓRIO**, José António (1998), «Sobre o ensino da Teledetecção em Geografia (e Planeamento Regional)» *Inforgeo*, nº 12 / 13, Lisboa, Edições Colibri, APG, pp. 627 - 635.

**TENEDÓRIO**, José António (2000), «Tecnologias de Informação Geográfica e Identificação de Corredores Verdes», in MACHADO, Reis (coord.), *Corredores Verdes: Contributo para um Ordenamento Sustentável Regional e Local*, Lisboa, FFCT / UNL, (no prelo).

## **Anexos**

---

**Anexo 1** - Dicionário de formas flexionadas

**Anexo 2** - Dicionário de locuções

**Anexo 3** - Dicionário de unidades terminológicas da detecção remota

**Anexo 4** - Texto etiquetado com o EtiqueLex, sem a aplicação de regras de desambiguação

**Anexo 5** - Sequências extraídas com ExtracTerm, a partir de texto etiquetado não desambiguado

**Anexo 6** - Texto etiquetado com o EtiqueTerm, com a aplicação de regras de desambiguação

**Anexo 7** - Sequências extraídas com ExtracTerm, a partir de texto etiquetado desambiguado

**Anexo 8** - Texto etiquetado com ExtracTerm, a partir de novo *corpus*, com aplicação de regras de desambiguação

**Anexo 9** - Sequências extraídas com ExtracTerm, a partir de texto etiquetado desambiguado

# **Anexo 1**

.....

Dicionário de formas flexionadas

a Art: def: f: s  
a Pron: pess: 3p: f: s  
a Pron: dem: 3p: f: s  
a Prep1: a  
à Prep2: à: f: s  
abaixo Adv  
abandonada Adj: f: s  
abandonada Pp: f: s  
abandonadas Pp: f: pl  
abandonadas Adj: f: pl  
abandonado Adj: m: s  
abandonado Pp: m: s  
abandonados Adj: m: pl  
abandonados Pp: m: pl  
abandonar V: inf  
abandono N: m: s  
abarca V: cj  
abarcando Ger  
abarcar V: inf  
abas N: f: pl  
abastecer V: inf  
abastecida Pp: f: s  
abastecida Adj: f: s  
abastecidas Adj: f: pl  
abastecido Pp: m: s  
abastecido Adj: m: s  
abastecidos Pp: m: pl  
abastecidos Adj: m: pl  
abcissas N: f: pl  
aberrantes Adj: m: s: f: pl  
aberta Adj: f: s  
aberta Pp: f: s  
abertas Adj: f: pl  
abertas Pp: f: pl  
aberto Adj: m: s  
aberto Pp: m: s  
abertos Adj: m: pl  
abertos Pp: m: pl  
abertura N: f: s  
aberturas N: f: pl  
aborda V: cj  
abordada Adj: f: s  
abordada Pp: f: s  
abordadas Adj: f: pl  
abordadas Pp: f: pl  
abordado Adj: m: pl  
abordado Pp: m: s  
abordados Adj: m: pl  
abordados Pp: m: pl  
abordagem N: f: s  
abordagens N: f: pl  
abordam V: cj  
abordar V: inf  
abordaremos V: cj  
abrange V: cj  
abrangem V: cj  
abrangendo Ger  
abrangente Adj: 2gen: s  
abrangentes Adj: 2gen: pl

abranger V:inf  
 abrangerá V:cj  
 abrangerão V:cj  
 abrangida Adj:f:s  
 abrangida Pp:f:s  
 abrangidas Adj:f:pl  
 abrangidas Pp:f:pl  
 abrangido Adj:m:s  
 abrangido Pp:m:s  
 abrangidos Adj:m:pl  
 abrangidos Pp:m:pl  
 abrasão N:m:s  
 abrem V:cj  
 abreviadamente Adv  
 abreviatura N:f:s  
 abriga V:cj  
 abrigada Adj:f:s  
 abrigada Pp:f:pl  
 abrigadas Adj:f:pl  
 abrigadas Pp:f:s  
 abrigado Adj:m:s  
 abrigado Pp:m:s  
 abrigados Adj:m:pl  
 abrigados Pp:m:pl  
 abrigar V:inf  
 abrigo N:m:s  
 abrigo V:cj  
 abril N:m:s  
 abrir V:inf  
 abriram V:cj  
 abrupta Adj:f:s  
 abrupto Adj:m:s  
 abruptamente Adv  
 absolutamente Adv  
 absolutas Adj:f:pl  
 absoluto Adj:m:s  
 absolutos Adj:m:pl  
 absorção N:f:s  
 absorções N:f:pl  
 absorva V:cj  
 absorvam V:cj  
 absorve V:cj  
 absorvem V:cj  
 absorvendo Ger  
 absorvente Adj:m:s:f:s  
 absorventes Adj:m:s:f:pl  
 absorver V:inf  
 absorverem V:cj  
 absorvida Adj:f:pl  
 absorvida Pp:f:s  
 absorvidade N:f:s  
 absorvidas Adj:f:pl  
 absorvidas Pp:f:pl  
 absorvido Adj:m:s  
 absorvido Pp:m:s  
 absorvidos Adj:m:s:ps  
 absorvidos Pp:m:s:ps  
 abstraindo Ger  
 abundância N:f:s

abundante Adj:2gen:s  
abundantemente Adv  
abundantes Adj:2gen:pl  
acabada Adj:f:s  
acabada Pp:f:s  
acabadas Adj:f:pl  
acabadas Pp:f:pl  
acabado Adj:m:s  
acabado Pp:m:s  
acabados Adj:m:pl  
acabados Pp:m:pl  
acabámos V:cj  
acabar V:inf  
acabara V:cj  
acabaram V:cj  
acabou V:cj  
acácia N:f:s  
académica Adj:f:s  
académicas Adj:f:pl  
académico Adj:m:s  
académicos Adj:m:pl  
acarretar V:inf  
acarretaria V:cj  
acarretou V:cj  
acaso Adv  
acaso N:m:s  
acastanhada Adj:f:s  
acautelada Adj:f:s  
acautelada Pp:f:s  
acauteladas Adj:f:pl  
acauteladas Pp:f:pl  
acautelado Adj:m:s  
acautelado Pp:m:s  
acautelados Adj:m:pl  
acautelados Pp:m:pl  
acautelando Ger  
acção N:f:s  
acciona V:cj  
accionada Adj:f:s  
accionada Pp:f:s  
accionadas Adj:f:pl  
accionadas Pp:f:pl  
accionado Adj:m:s  
accionado Pp:m:s  
accionados Adj:m:pl  
accionados Pp:m:pl  
accionar V:inf  
acções N:f:pl  
acedendo Ger  
aceder V:inf  
acdermos V:cj  
aceita V:cj  
aceitação N:f:s  
aceitando Ger  
aceitar V:inf  
aceitaria V:cj  
aceitáveis Adj:2gen:pl  
aceitável Adj:2gen:s  
aceite Adj:2gen:s

aceite Pp:2gen:s  
 aceites Adj:2gen:pl  
 aceites Pp:2gen:pl  
 aceitou V:cj  
 aceleração N:f:s  
 acelerada Adj:f:pl  
 acelerada Pp:f:s  
 aceleradas Adj:f:pl  
 aceleradas Pp:f:pl  
 acelerado Adj:m:s  
 acelerado Pp:m:s  
 acelerados Adj:m:s  
 acelerados Pp:m:pl  
 acelerar V:inf  
 acento N:m:s  
 acentuada Adj:f:s  
 acentuada Pp:f:s  
 acentuadamente Adv  
 acentuadas Adj:f:pl  
 acentuadas Pp:f:pl  
 acentuado Adj:m:s  
 acentuado Pp:m:s  
 acentuados Adj:m:pl  
 acentuados Pp:m:pl  
 acentuar V:inf  
 aconselháveis Adj:2gen:pl  
 aconselhável Adj:2gen:s  
 acepções N:f:pl  
 acertada Adj:f:s  
 acertada Pp:f:s  
 acertadas Adj:f:pl  
 acertadas Pp:f:pl  
 acertado Adj:m:s  
 acertado Pp:m:s  
 acertados Adj:m:pl  
 acertados Pp:m:pl  
 acertar V:inf  
 acessibilidade N:f:pl  
 acessíveis Adj:2gen:pl  
 acessível Adj:2gen:f:s  
 acesso N:m:s  
 acetato N:m:s  
 achatamento N:m:s  
 acidentada Adj:f:s  
 acidentada Pp:f:s  
 acidentadas Adj:f:pl  
 acidentadas Pp:f:pl  
 acidentado Adj:m:s  
 acidentado Pp:m:s  
 acidentados Adj:m:pl  
 acidentados Pp:m:pl  
 acidentais Adj:2gen:pl  
 acidental Adj:2gen:s  
 acidentiar V:inf  
 acidentes N:m:pl  
 acima Adv  
 acinzentados Adj:m:pl  
 aclarar V:inf  
 acompanhar V:cj

acompanhada Adj:f:s  
acompanhada Pp:f:s  
acompanhadas Adj:f:pl  
acompanhadas Pp:f:pl  
acompanhado Adj:m:s  
acompanhado Pp:m:s  
acompanhados Adj:m:pl  
acompanhados Pp:m:pl  
acompanhamento N:m:s  
acompanhar V:inf  
aconselha V:cj  
aconselhada Adj:f:s  
aconselhada Pp:f:s  
aconselhadas Adj:f:pl  
aconselhadas Pp:f:pl  
aconselhado Adj:m:s  
aconselhado Pp:m:s  
aconselhados Adj:f:pl  
aconselhados Pp:f:pl  
aconselhamos V:cj  
aconselhar V:inf  
aconselhável Adj:2gen:s  
aconteça V:cj  
acontece V:cj  
acontecendo Ger  
acontecer V:inf  
acontecerá V:cj  
aconteceu V:intra:cj  
acontecia V:intra:cj  
acontecimentos N:m:pl  
acoplada Adj:f:s  
acoplada Pp:f:s  
acopladas Adj:f:pl  
acopladas Pp:f:pl  
acoplado Adj:m:s  
acoplado Pp:m:s  
acoplados Adj:m:pl  
acoplados Pp:m:pl  
acoplar V:inf  
acordada Adj:f:s  
acordada Pp:f:s  
acordadas Adj:f:pl  
acordadas Pp:f:pl  
acordado Adj:m:s  
acordado Pp:m:s  
acordados Pp:m:pl  
acordados Adj:m:pl  
acordar V:inf  
acordo N:m:s  
acordos N:m:pl  
açores Np  
acostagem N:f:s  
acp Sigla  
acreditamos V:cj  
acreditar V:inf  
acredite V:cj  
acrescentada Adj:f:s  
acrescentada Pp:f:s  
acrescentadas Adj:f:pl

acrescentadas Pp:f:pl  
 acrescentado Adj:m:s  
 acrescentado Pp:m:s  
 acrescentados Adj:m:pl  
 acrescentados Pp:m:pl  
 acrescentam V:cj  
 crescida Adj:f:s  
 crescida Pp:f:s  
 crescidas Adj:f:pl  
 crescidas Pp:f:pl  
 crescido Adj:m:s  
 crescido Pp:m:s  
 crescidos Adj:m:pl  
 crescidos Pp:m:pl  
 acréscimo N:m:s  
 acréscimos N:m:pl  
 acrescentar V:inf  
 acrónimo N:m:s  
 acrónimos N:m:pl  
 acta N:f:s  
 actas N:f:pl  
 activa V:cj  
 activa Adj:f:s  
 activação N:f:s  
 activamente Adv  
 activar V:inf  
 activas Adj:f:pl  
 actividade N:f:s  
 actividades N:f:pl  
 activo V:cj  
 activo Adj:m:s  
 activos Adj:m:pl  
 acto N:m:s  
 actos N:m:pl  
 actua V:cj  
 actuação N:f:s  
 actuações N:f:pl  
 actuais Adj:2gen:pl  
 actual Adj:2gen:s  
 actualidade N:f:s  
 actualidades N:f:pl  
 actualização N:f:s  
 actualizações N:f:pl  
 actualizada Adj:f:s  
 actualizada Pp:f:s  
 actualizadas Adj:f:pl  
 actualizadas Pp:f:pl  
 actualizado Adj:m:s  
 actualizado Pp:m:s  
 actualizados Adj:m:pl  
 actualizados Pp:m:pl  
 actualizar V:inf  
 actualizarem V:cj  
 actualizáveis Adj:2gen:pl  
 actualizável Adj:2gen:s  
 actualmente Adv  
 actuam V:cj  
 actuando Ger  
 actuar V:inf

actuou V:cj  
 acautelar V:inf  
 acuidade N:f:s  
 acumulação N:f:s  
 acumulações N:f:pl  
 acumulada Adj:f:s  
 acumulada Pp:f:s  
 acumuladas Adj:f:pl  
 acumuladas Pp:f:pl  
 acumulado Adj:m:s  
 acumulado Pp:m:s  
 acumulados Adj:m:pl  
 acumulados Pp:m:pl  
 acumulando Ger  
 acumular V:inf  
 acusando Ger  
 adapta V:cj  
 adaptação N:f:s  
 adaptações N:f:pl  
 adaptada Adj:f:s  
 adaptada Pp:f:s  
 adaptado Adj:m:s  
 adaptado Adj:m:s  
 adaptado Pp:m:s  
 adaptado Pp:m:s  
 adaptados Adj:m:pl  
 adaptados Pp:m:pl  
 adaptar V:inf  
 adaptou V:cj  
 adelgaçamento N:f:s  
 adelgaçamentos N:f:pl  
 adenda N:f:s  
 adendas N:f:pl  
 adequa V:cj  
 adequabilidade N:f:s  
 adequação N:f:s  
 adequada Adj:f:s  
 adequada Pp:f:s  
 adequadamente Adv  
 adequadas Adj:f:pl  
 adequadas Pp:f:pl  
 adequado Adj:f:s  
 adequado Pp:f:s  
 adequados Adj:f:pl  
 adequados Pp:f:pl  
 adequar V:inf  
 adesão N:f:s  
 adiada Adj:f:s  
 adiada Pp:f:s  
 adiadas Adj:f:pl  
 adiadas Pp:f:pl  
 adiado Adj:m:s  
 adiado Pp:m:s  
 adiados Adj:m:pl  
 adiados Pp:m:pl  
 adiantar V:inf  
 adiante Adv  
 adição N:f:s  
 adiciona V:cj

adicionada Adj:f:s  
 adicionada Pp:f:s  
 adicionadas Adj:f:pl  
 adicionadas Pp:f:pl  
 adicionado Adj:m:s  
 adicionado Pp:m:s  
 adicionados Adj:m:pl  
 adicionados Pp:m:pl  
 adicionais Adj:2gen:pl  
 adicional Adj:2gen:s  
 adicionam V:cj  
 adicionar V:inf  
 adições N:f:pl  
 adimensionais Adj:2gen:pl  
 adimensional Adj:2gen:s  
 aditividade N:f:s  
 aditivo N:m:s  
 aditivos N:m:pl  
 adjacente Adj:2gen:s  
 adjacentes Adj:2gen:pl  
 adjudicou V:cj  
 administração N:f:s  
 administrativa Adj:f:s  
 administrativamente Adv  
 administrativas Adj:f:pl  
 administrativos Adj:m:pl  
 admissíveis Adj:2gen:pl  
 admissível Adj:2gen:s  
 admite V:cj  
 admitido V:cj  
 admitindo Ger  
 admitir V:inf  
 adoção N:f:s  
 adotada Adj:f:s  
 adotada Pp:f:s  
 adotadas Adj:f:pl  
 adotadas Pp:f:pl  
 adotado Adj:m:s  
 adotado Pp:m:s  
 adotados Adj:m:pl  
 adotados Pp:m:pl  
 adotamos V:cj  
 adotando Ger  
 adotar V:inf  
 adotou V:cj  
 adquire V:cj  
 adquirem V:cj  
 adquirida Adj:f:s  
 adquirida Pp:f:s  
 adquiridas Adj:f:pl  
 adquiridas Pp:f:pl  
 adquirido Adj:m:s  
 adquirido Pp:m:s  
 adquiridos Adj:m:pl  
 adquiridos Pp:m:pl  
 adquirindo Ger  
 adquirir V:inf  
 adquiriram V:cj  
 adquiriu V:cj

aduzimos V:cj  
 adversa Adj:f:s  
 adversas Adj:f:pl  
 adverso Adj:m:s  
 adversos Adj:m:pl  
 aérea Adj:f:s  
 aéreas Adj:f:pl  
 aéreo Adj:m:s  
 aéreos Adj:m:pl  
 aerodinâmica Adj:f:s  
 aerodinâmicas Adj:f:pl  
 aerodinâmico Adj:m:s  
 aerodinâmicos Adj:m:pl  
 aeródromo N:m:s  
 aeródromos N:m:pl  
 aeroespacial Adj:2gen:s  
 aerofotográfica Adj:f:s  
 aerofotográficas Adj:f:pl  
 aerofotográfico Adj:m:s  
 aerofotográficos Adj:m:pl  
 aeromotor N:m:s  
 aeromotores N:m:pl  
 aeronave N:m:s  
 aeronaves N:m:pl  
 aeroplano N:m:s  
 aeroplanos N:m:pl  
 aeroporto N:m:s  
 aeroportos N:m:pl  
 aerosóis N:m:pl  
 aerosóis N:m:pl  
 aerosol N:m:s  
 aerossóis N:m:pl  
 aerotransportada Adj:m:s  
 aerotransportada Pp:f:s  
 aerotransportadas Adj:f:pl  
 aerotransportadas Pp:f:pl  
 aerotransportado Adj:m:s  
 aerotransportado Pp:m:s  
 aerotransportados Adj:m:pl  
 aerotransportados Pp:m:pl  
 aerotransportáveis Adj:2gen:pl  
 aerotransportável Adj:2gen:s  
 aerotriangulação N:f:s  
 afasta V:cj  
 afastada Adj:f:s  
 afastada Pp:f:s  
 afastadas Adj:f:pl  
 afastadas Pp:f:pl  
 afastado Adj:m:s  
 afastado Pp:m:s  
 afastados Adj:m:pl  
 afastados Pp:m:pl  
 afastamento N:m:s  
 afastamentos N:m:pl  
 afastar V:inf  
 afecta Adj:f:s  
 afecta Pp:f:s  
 afecta V:cj  
 afectação N:f:s

afetações	N:f:pl
afectada	Adj:f:s
afectada	Pp:f:s
afectadas	Adj:f:pl
afectadas	Pp:f:pl
afectado	Adj:m:s
afectado	Pp:m:s
afectados	Adj:m:pl
afectados	Pp:m:pl

## **Anexo 2**

Dicionário de locuções

aquém de@Loc  
 à beira@Loc  
 à beira de@Loc  
 à certa@Loc  
 a custo@Loc  
 a despeito de@Loc  
 a fim de@Loc  
 à frente da@Loc  
 à frente das@Loc  
 à frente de@Loc  
 à frente do@Loc  
 à frente dos@Loc  
 a jusante@Loc  
 à medida de@Loc  
 à medida que@Loc  
 a par com@Loc  
 a par de@Loc  
 à parte@Loc  
 a ponto de@Loc  
 a princípio@Loc  
 a propósito@Loc  
 a propósito de@Loc  
 a respeito da@Loc  
 a respeito das@Loc  
 a respeito de@Loc  
 a respeito do@Loc  
 a respeito dos@Loc  
 a rigor@Loc  
 à risca@Loc  
 à roda da@Loc  
 à roda das@Loc  
 à roda de@Loc  
 à roda do@Loc  
 à roda dos@Loc  
 a seguir@Loc  
 a seguir a@Loc  
 a título de@Loc  
 à vista de@Loc  
 à volta de@Loc  
 abaixo da@Loc  
 abaixo das@Loc  
 abaixo de@Loc  
 abaixo do@Loc  
 abaixo dos@Loc  
 acerca da@Loc  
 acerca das@Loc  
 acerca de@Loc  
 acerca do@Loc  
 acerca dos@Loc  
 acima da@Loc  
 acima das@Loc  
 acima de@Loc  
 acima do@Loc  
 acima dos@Loc  
 adiante da@Loc  
 adiante das@Loc  
 adiante de@Loc  
 adiante do@Loc  
 adiante dos@Loc

afim de@Loc  
ainda agora@Loc  
ainda assim@Loc  
ainda que@Loc  
além da@Loc  
além das@Loc  
além de@Loc  
além do@Loc  
além dos@Loc  
antes da@Loc  
antes das@Loc  
antes de@Loc  
antes do@Loc  
antes dos@Loc  
antes que@Loc  
ao acaso@Loc  
ao certo@Loc  
ao envés@Loc  
ao invés de@Loc  
ao lado da@Loc  
ao lado das@Loc  
ao lado de@Loc  
ao lado do@Loc  
ao lado dos@Loc  
ao longo da@Loc  
ao longo das@Loc  
ao longo de@Loc  
ao longo do@Loc  
ao longo dos@Loc  
ao passo que@Loc  
ao passo que@Loc  
ao principio@Loc  
ao redor da@Loc  
ao redor das@Loc  
ao redor de@Loc  
ao redor do@Loc  
ao redor dos@Loc  
ao todo@Loc  
apesar da@Loc  
apesar das@Loc  
apesar de@Loc  
apesar do@Loc  
apesar dos@Loc  
após da@Loc  
após das@Loc  
após de@Loc  
após do@Loc  
após dos@Loc  
aquando de@Loc  
às avessas@Loc  
às vezes@Loc  
assim como@Loc  
assim que@Loc  
até à@Loc  
até ao@Loc  
até aos@Loc  
até às@Loc  
até que@Loc  
atrás da@Loc

atrás das@Loc  
atrás de@Loc  
atrás do@Loc  
atrás dos@Loc  
através da@Loc  
através das@Loc  
através de@Loc  
através do@Loc  
através dos@Loc  
bem assim@Loc  
cada vez@Loc  
cada vez que@Loc  
cerca de@Loc  
com certeza@Loc  
com efeito@Loc  
com propósito@Loc  
com recurso a@Loc  
com referência a@Loc  
com respeito a@Loc  
com respeito a@Loc  
com respeito à@Loc  
com respeito ao@Loc  
com respeito aos@Loc  
com respeito às@Loc  
da mesma forma@Loc  
dado que@Loc  
de acordo@Loc  
de acordo com@Loc  
de antemão@Loc  
de assento@Loc  
de certeza@Loc  
de cima da@Loc  
de cima das@Loc  
de cima de@Loc  
de cima do@Loc  
de cima dos@Loc  
de conformidade com@Loc  
de facto@Loc  
de feito@Loc  
de forma a@Loc  
de forma que@Loc  
de frente@Loc  
de graça@Loc  
de modo a@Loc  
de modo algum@Loc  
de modo que@Loc  
de nenhum modo@Loc  
de propósito@Loc  
de que@Loc  
de resto@Loc  
de rigor@Loc  
de seguida@Loc  
de súbito@Loc  
de todo@Loc  
debaixo da@Loc  
debaixo das@Loc  
debaixo de@Loc  
debaixo do@Loc  
debaixo dos@Loc

em lugar da@Loc  
em lugar das@Loc  
em lugar de@Loc  
em lugar do@Loc  
em lugar dos@Loc  
em prol da@Loc  
em prol das@Loc  
em prol de@Loc  
em prol do@Loc  
em prol dos@Loc  
em razão da@Loc  
em razão das@Loc  
em razão de@Loc  
em razão do@Loc  
em razão dos@Loc  
em redor da@Loc  
em redor das@Loc  
em redor de@Loc  
em redor do@Loc  
em redor dos@Loc  
em rigor@Loc  
em seguida@Loc  
em seguida a@Loc  
em termos de@Loc  
em torno da@Loc  
em torno das@Loc  
em torno de@Loc  
em torno do@Loc  
em torno dos@Loc  
em troco da@Loc  
em troco das@Loc  
em troco de@Loc  
em troco do@Loc  
em troco dos@Loc  
em vão@Loc  
em vez da@Loc  
em vez das@Loc  
em vez de@Loc  
em vez do@Loc  
em vez dos@Loc  
em virtude de@Loc  
enquanto que@Loc  
fora da@Loc  
fora das@Loc  
fora de@Loc  
fora do@Loc  
fora dos@Loc  
graças a@Loc  
graças à@Loc  
graças ao@Loc  
graças aos@Loc  
graças às@Loc  
já agora@Loc  
já que@Loc  
junto a@Loc  
junto à@Loc  
junto ao@Loc  
junto aos@Loc  
junto às@Loc

junto da@Loc  
junto das@Loc  
junto de@Loc  
junto do@Loc  
junto dos@Loc  
logo que@Loc  
mal que@Loc  
mesmo assim@Loc  
não obstante@Loc  
nem que@Loc  
no entanto@Loc  
no tocante a@Loc  
onde quer que@Loc  
ou seja@Loc  
para além da@Loc  
para além das@Loc  
para além de@Loc  
para além do@Loc  
para além dos@Loc  
para baixo da@Loc  
para baixo das@Loc  
para baixo de@Loc  
para baixo do@Loc  
para baixo dos@Loc  
para cima da@Loc  
para cima das@Loc  
para cima de@Loc  
para cima do@Loc  
para cima dos@Loc  
para com@Loc  
para diante@Loc  
para já@Loc  
pela certa@Loc  
perto da@Loc  
perto das@Loc  
perto de@Loc  
perto do@Loc  
perto dos@Loc  
por acaso@Loc  
por alto@Loc  
por baixo da@Loc  
por baixo das@Loc  
por baixo de@Loc  
por baixo do@Loc  
por baixo dos@Loc  
por causa da@Loc  
por causa das@Loc  
por causa de@Loc  
por causa do@Loc  
por causa dos@Loc  
por cima da@Loc  
por cima das@Loc  
por cima de@Loc  
por cima do@Loc  
por cima dos@Loc  
por conseguinte@Loc  
por consequência@Loc  
por defronte da@Loc  
por defronte das@Loc

por defronte de@Loc  
 por defronte do@Loc  
 por defronte dos@Loc  
 por dentro da@Loc  
 por dentro das@Loc  
 por dentro de@Loc  
 por dentro do@Loc  
 por dentro dos@Loc  
 por detrás da@Loc  
 por detrás das@Loc  
 por detrás de@Loc  
 por detrás do@Loc  
 por detrás dos@Loc  
 por diante da@Loc  
 por diante das@Loc  
 por diante de@Loc  
 por diante do@Loc  
 por diante dos@Loc  
 por enquanto@Loc  
 por entre@Loc  
 por forma a@Loc  
 por isso@Loc  
 por meio da@Loc  
 por meio das@Loc  
 por meio de@Loc  
 por meio do@Loc  
 por meio dos@Loc  
 por oposição a@Loc  
 por oposição à@Loc  
 por oposição ao@Loc  
 por oposição aos@Loc  
 por oposição às@Loc  
 por outro lado@Loc  
 por parte da@Loc  
 por parte das@Loc  
 por parte de@Loc  
 por parte do@Loc  
 por parte dos@Loc  
 por seu turno@Loc  
 por sua vez@Loc  
 por trás da@Loc  
 por trás das@Loc  
 por trás de@Loc  
 por trás do@Loc  
 por trás dos@Loc  
 por último@Loc  
 por um lado@Loc  
 por vezes@Loc  
 por via de@Loc  
 posto que@Loc  
 primeiro que@Loc  
 quando de@Loc  
 quanto a@Loc  
 quanto à@Loc  
 quanto ao@Loc  
 quanto aos@Loc  
 quanto às@Loc  
 quanto mais@Loc  
 que nem@Loc

defronte da@Loc  
 defronte das@Loc  
 defronte de@Loc  
 defronte do@Loc  
 defronte dos@Loc  
 dentre de@Loc  
 dentro da@Loc  
 dentro das@Loc  
 dentro de@Loc  
 dentro do@Loc  
 dentro dos@Loc  
 dentro em@Loc  
 depois da@Loc  
 depois das@Loc  
 depois de@Loc  
 depois do@Loc  
 depois dos@Loc  
 desde já@Loc  
 desde logo@Loc  
 desde que@Loc  
 detrás de@Loc  
 diante da@Loc  
 diante das@Loc  
 diante de@Loc  
 diante do@Loc  
 diante dos@Loc  
 em baixo da@Loc  
 em baixo das@Loc  
 em baixo de@Loc  
 em baixo do@Loc  
 em baixo dos@Loc  
 em breve@Loc  
 em cima da@Loc  
 em cima das@Loc  
 em cima de@Loc  
 em cima do@Loc  
 em cima dos@Loc  
 em comum@Loc  
 em conformidade com@Loc  
 em detrimento da@Loc  
 em detrimento das@Loc  
 em detrimento de@Loc  
 em detrimento do@Loc  
 em detrimento dos@Loc  
 em favor da@Loc  
 em favor das@Loc  
 em favor de@Loc  
 em favor do@Loc  
 em favor dos@Loc  
 em frente a@Loc  
 em frente à@Loc  
 em frente ao@Loc  
 em frente aos@Loc  
 em frente às@Loc  
 em frente da@Loc  
 em frente das@Loc  
 em frente de@Loc  
 em frente do@Loc  
 em frente dos@Loc

se bem que@Loc  
sem demora@Loc  
sem dúvida@Loc  
sem embargo de@Loc  
sem que@Loc  
sempre que@Loc  
sempre que@Loc  
sempre que@Loc  
suposto que@Loc  
tal que@Loc  
tanto mais que@Loc  
tanto que@Loc  
uma vez que@Loc  
visto que@Loc

## **Anexo 3**

---

Dicionário de unidades terminológicas da detecção remota

absorção atmosférica@N:f:s  
 agregação de objectos@N:f:s  
 agregação espacial@N:f:s  
 água líquida@N:f:s  
 água profunda@N:f:s  
 amplitude térmica@N:f:s  
 ângulo azimutal@N:m:s  
 ângulo de iluminação@N:m:s  
 ângulo de varrimento@N:m:s  
 ângulo zenital de iluminação@N:m:s  
 ângulo zenital@N:m:s  
 assinatura espectral@N:f:s  
 assinaturas espectrais@N:f:pl  
 detecção remota@N:f:s  
 filmes ortocromáticos@N:m:pl  
 filmes pancromáticos@N:m:pl  
 filtragem de imagens@N:m:s  
 fluxo de calor@N:m:s  
 fotointerpretação assistida por computador@N:f:s  
 foto-interpretção assistida por computador@N:f:s  
 imagem assitida por computador@N:f:s  
 imagens assistidas por compitador@N:f:pl  
 imagem de textura@N:f:s  
 imagem de satélite@N:f:s  
 imagens de satélite@N:f:pl  
 imagem filme@N:f:s  
 imagem filtrada@N:f:s  
 imagem LANDSAT TM@N:f:s  
 imagem MSS@N:f:s  
 imagem multibanda@N:f:s  
 imagem não corrigida@N:f:s  
 imagem numérico hiperespectral@N:f:s  
 imagem oblíqua@N:f:s  
 imagem orbital multiespectral@N:f:s  
 imagem termal aérea@N:f:s  
 imagem termal@N:f:s  
 imagem textura@N:f:s  
 imagem-textura@N:f:s  
 inclinação orbital@N:f:s  
 inclinação solar@N:f:s  
 índice de biomassa@N:m:s  
 índice de brilho@N:m:s  
 índice de refração@N:m:s  
 índice de textura@N:m:s  
 lei de difusão@N:f:s  
 lei de Kepler@N:f:s  
 lei de Planck@N:f:s  
 lei dos corpos negros@N:f:s  
 lente convergente@N:f:s  
 lente divergente@N:f:s  
 luminância hiperfrequência@N:f:s  
 métodos de correcção@N:f:s  
 modelo de reflectância@N:m:s  
 modo imagem@N:m:s  
 ondas electromagnéticas@N:f:pl  
 ondas longitudinais@N:f:pl  
 ondas sonoras@N:f:pl  
 polarização HV@N:f:s  
 polarização vertical@N:f:s

potência espectral@N:f:s  
radar altimétrico@N:m:s  
radar lateral@N:m:s  
radiação electromagnética@N:f:s  
radiação solar incidente@N:f:s  
raio X@N:m:s  
reflectância absoluta@N:f:s  
reflexão difusa@N:f:s  
reflexão especular@N:f:s  
resolução espacial@N:f:s  
resolução espectral@N:f:s  
resolução transversal@N:f:s  
rugosidade de superfície@N:f:s  
satélite de observação da terra@N:m:s  
satélite de teledeteção@N:m:s  
satélite geoestacionário@N:m:s  
satélite LANDSAT@N:m:s  
satélite militar@N:m:s  
satélite RADARSAT@N:m:s  
satélite SPOT@N:m:s  
satélite TIROS@N:m:s  
segmentação de imagem@N:f:s  
segmentação de imagem@N:f:s  
sinal analógico@N:m:s  
sinal digital@N:m:s  
síntese de abertura@N:f:s  
sistema de Munsell@N:m:s  
sistema de varrimento espectral@N:m:s  
sistema de varrimento óptico@N:m:s  
sistema fotográfico@N:m:s  
SPOT IMAGE@N:f:s  
stress vegetal@N:m:s  
suporte fotográfico@N:m:s  
taxa de polarização@N:f:s  
técnicas de compressão@N:f:s  
temperatura de brilho@N:f:s  
temperatura radiativa@N:f:s  
tratamento de imagens@N:m:s  
variação de radiodifusão@N:m:s  
variação de temperatura@N:m:s  
variação espectral@N:f:s  
variação temporal@N:f:s  
varrimento espectral@N:m:s  
varrimento multidimensional@N:m:s  
varrimento temporal@N:m:s  
visão estereoscópica@N:f:s  
zona espectral@N:f:s  
zonas espectrais@N:f:pl

## **Anexo 4**

---

---

Texto etiquetado com o EtiqueLex,  
sem a aplicação de regras de desambiguação

Teledeteção [N:f:s] em [Prepl:em] áreas [N:f:pl] periurbanas [Adj:f:pl] : combinação [N:f:s] de [Prepl:de] índices [N:m:pl] temáticos [Adj:m:pl] para [Prepl:para] localizar [V:inf] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] mudança [N:f:s] de [Prepl:de] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] com [Prepl:com] recurso [N:m:s] às [Prep2:às:s:pl] imagens [N:f:pl] digitais [Adj:2gen:pl] Spot [Sigla] HRV [Sigla] .

As [Art:def:f:pl] manifestações [N:f:pl] territoriais [Adj:2gen:pl] do [Prep2:do:m:s] crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] urbanas [Adj:f:pl] têm [V:cj] sido [Pp] abundantemente [Adv] tratadas [Pp:f:pl] [Adj:f:pl] nas [Prep2:nas:f:pl] ópticas [Adj:f:pl] [N:f:pl] económica [Adj:f:s] e [Conj:e] financeira [Adj:f:s] , social [Adj:2gen:s] e [Conj:e] política [N:f:s] , demográfica [Adj:f:s] e [Conj:e] habitacional [Adj:2gen:s] , fundiária [Adj:f:s] e [Conj:e] funcional [Adj:2gen:s] . As [Art:def:f:pl] técnicas [N:f:pl] e [Conj:e] os [Art:def:m:pl] [Pron:poss] [Pron:dem] métodos [N:m:pl] para [Prepl:para] localizar [V:inf] , descrever [V:inf] e [Conj:e] analisar [V:inf] esse [Pron:dem:3p:m:s] crescimento [N:m:s] baseiam-se , sobretudo [Adv] , na [Prep2:na:f:s] estatística [Adj:f:s] [N:f:s] descritiva [Adj:f:s] e [Conj:e] na [Prep2:na:f:s] análise [N:f:s] de [Prepl:de] dados [Pp:m:pl] [N:m:pl] ( análise [N:f:s] individual [Adj:2gen:s] de [Prepl:de] variáveis [Adj:2gen:pl] [N:f:pl] , análise [N:f:s] de [Prepl:de] variáveis [Adj:2gen:pl] [N:f:pl] duas [Num:card] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] duas [Num:card] , análise [N:f:s] em [Prepl:em] componentes [N:f:pl] principais [Adj:2gen:pl] , análise [N:f:s] factorial [Adj:2gen:s] de [Prepl:de] correspondências [N:f:pl] , classificação [N:f:s] ascendente [Adj:2gen:s] hierárquica [Adj:f:s] , etc . ) . A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] informação [N:f:s] é [V:cj] tratada [Pp:f:s] [Adj:f:s] diacronicamente [Adv] para [Prepl:para] deduzir [V:inf] , a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] partir [V:inf] de [Prepl:de] escalas [N:f:pl] muito [Adv] [Pron:indef:m:s] variadas [Adj:f:pl] , as [Art:def:f:pl] transformações [N:f:pl] que [Adv] [Pron:rel] vão [V:cj] ocorrendo [Ger] em [Prepl:em] unidades [N:f:pl] territoriais [Adj:2gen:pl] definidas [Adj:f:pl] [Pp:f:pl] administrativamente [Adv] ( freguesia [N:f:s] , concelho [N:m:s] , commune , département , län [N:] , etc ) .

A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] abordagem [N:f:s] estatística [Adj:f:s] [N:f:s] do [Prep2:do:m:s] fenómeno [N:m:s] é [V:cj] delicada [Adj:f:s] . Em [Prepl:em] primeiro [Adj:m:s] [Adv] [Num:ord] lugar [N:m:s] porque [Adv] [Conj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] consideração [N:f:s] da [Prep2:da:f:s] unidade [N:f:s] administrativa [Adj:f:s] como [Adv] [Conj] unidade [N:f:s] mínima [Adj:f:s] de [Prepl:de] referência [N:f:s] espacial [Adj:2gen:s] esconde [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] diversidade [N:f:s] de [Prepl:de] situações [N:f:pl] no [Prep2:no:m:s] seu [Pron:poss:3p:m:s] interior [Adj:2gen:s] [N:m:s] . Em [Prepl:em] segundo [Adv] [N:m:s] [Num:ord] lugar [N:m:s] , porque [Adv] [Conj] dificilmente [Adv] se [Conj] [Pron:poss] obtém [V:cj] um [Art:indef] conjunto [N:m:s] completo [Adj:m:s] de [Prepl:de] variáveis [Adj:2gen:pl] [N:f:pl] ( demográficas [Adj:f:pl] , socio-económicas , de [Prepl:de] uso [N:m:s] [V:cj] e [Conj:e] ocupação [N:f:s] física [Adj:f:s] [N:f:s] do [Prep2:do:m:s] território [N:m:s] ) que [Adv] [Pron:rel] responda [V:cj] em [Prepl:em] simultâneo [Adj:m:s] à [Prep2:à:f:s] definição [N:f:s] de [Prepl:de] coroa [N:f:pl] , sectores [N:m:pl] ou [Conj:ou] zonas [N:f:pl] periurbanas [Adj:f:pl] e [Conj:e] à

[Prep2:à:f:s] sua [Pron:poss:3p:f:s] caracterização [N:f:s] detalhada [Adj:f:s] [Pp:f:s] . Os [Art:def:m:pl] [Pron:poss] [Pron:dem] dados [Pp:m:pl] [N:m:pl] , quando [Adv] [Conj] existem [V:cj] , são [V:cj] de [Prep1:de] difícil [Adj:2gen:s] similaridade [N:f:s] e [Conj:e] raramente [Adv] permitem [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] detecção [N:f:s] de [Prep1:de] descontinuidades [N:f:pl] e [Conj:e] de [Prep1:de] fronteiras [N:f:pl] de [Prep1:de] forma [N:f:s] sistemática [Adj:f:s] , como [Adv] [Conj] exige [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] reflexão [N:f:s] sobre [Prep1:sobre] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] fenómeno [N:m:s] de [Prep1:de] periurbanização [N:f:s] .

A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] classificação [N:f:s] e [Conj:e] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] representação [N:f:s] cartográfica [Adj:f:s] do [Prep2:do:m:s] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] e [Conj:e] da [Prep2:da:f:s] sua [Pron:poss:3p:f:s] evolução [N:f:s] permitem [V:cj] abordar [V:inf] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] crescimento [N:m:s] urbano [Adj:m:s] segundo [Adv] [N:m:s] [Num:ord] uma [Art:indef:f:s] óptica [Adj:f:s] [N:f:s] fisico-funcional . O [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] cálculo [N:m:s] [N:m:s] , a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] partir [V:inf] de [Prep1:de] cartas [N:f:pl] , das [Prep2:das:f:pl] mutações [N:f:pl] de [Prep1:de] uso [N:m:s] [V:cj] e [Conj:e] da [Prep2:da:f:s] área [N:f:s] ocupada [Pp:f:s] [Adj:f:s] por [Prep1:por] cada [Pron:indef] uso [N:m:s] [V:cj] , possibilita [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] introdução [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] física [Adj:f:s] [N:f:s] na [Prep2:na:f:s] análise [N:f:s] quantitativa [Adj:f:s] do [Prep2:do:m:s] problema [N:f:s] de [Prep1:de] crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] urbanas [Adj:f:pl] .

Partindo [Ger] destas [Prep2:destas:f:pl] duas [Num:card] hipóteses [N:f:pl] concluímos [V:cj] em [Prep1:em] 1989 a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] primeira [Adj:f:s] [Adv] etapa [N:f:s] de [Prep1:de] um [Art:indef] trabalho [N:m:s] que [Adv] [Pron:rel] teve [V:cj] por [Prep1:por] objectivo [Adj:m:s] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] definição [N:f:s] duma [Prep2:duma:f:s] metodologia [N:f:s] para [Prep1:para] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] realização [N:f:s] de [Prep1:de] cartas [N:f:pl] de [Prep1:de] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] e [Conj:e] de [Prep1:de] evolução [N:f:s] de [Prep1:de] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] , por [Prep1:por] interpretação [N:f:s] estereoscópica [Adj:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] . O [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] resultado [N:m:s] final [Adj] é [V:cj] constituído [Adj:m:s] [Pp:m:s] por [Prep1:por] duas [Num:card] cartas [N:f:pl] temáticas [Adj:f:pl] do [Prep2:do:m:s] Concelho [N:m:s] de [Prep1:de] Almada , onde [Adv] está [V:cj] representado [Pp:m:s] [Adj:m:s] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] em [Prep1:em] 1986 ( Carta [N:f:s] de [Prep1:de] Uso [N:m:s] [V:cj] do [Prep2:do:m:s] Solo [N:m:s] ) e [Conj:e] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] sua [Pron:poss:3p:f:s] evolução [N:f:s] entre [Prep1:entre] 1967 e [Conj:e] 1986 ( Carta [N:f:s] de [Prep1:de] Evolução [N:f:s] de [Prep1:de] Uso [N:m:s] [V:cj] do [Prep2:do:m:s] Solo [N:m:s] ) , segundo [Adv] [N:m:s] [Num:ord] uma [Art:indef:f:s] legenda [N:f:s] em [Prep1:em] 12 classes

[N:f:pl] . A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] quantificação [N:f:s] das [Prep2:das:f:pl] cartas [N:f:pl] permitiu [V:cj] apurar [N:inf] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] área [N:f:s] ocupada [Pp:f:s] [Adj:f:s] por [Prep1:por] cada [Pron:indef] uso [N:m:s] [V:cj] em [Prep1:em] 1986 e [Conj:e] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] área [N:f:s] das [Prep2:das:f:pl] zonas [N:f:pl] que [Adv] [Pron:rel] sofreram [V:cj] modificação [N:f:s] entre [Prep1:entre] 1967 e [Conj:e] 1986 . Posteriormente [Adv] procedeu-se à [Prep2:à:f:s] análise [N:f:s] quantitativa [Adj:f:s] das [Prep2:das:f:pl] transformações [N:f:pl] ocorridas [Pp:f:pl] [Adj:f:pl] em [Prep1:em] 20 anos [N:m:pl] .

A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] informação [N:f:s] obtida [Pp:f:s] [Adj:f:s] sobre [Prep1:sobre] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] , por [Prep1:por] segmentação [N:f:s] e [Conj:e] interpretação [N:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] , é [V:cj] extraordinariamente [Adv] precisa [V:cj] e [Conj:e] exaustiva [Adj:f:s] e [Conj:e] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] unidade [N:f:s] de [Prep1:de] análise [N:f:s] , a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] zona [N:f:s] urbana [Adj:f:s] homogénea [Adj:f:s] , responde [V:cj] satisfatoriamente [Adv] à [Prep2:à:f:s] abordagem [N:f:s] do [Prep2:do:m:s] crescimento [N:m:s] urbano [Adj:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] . Contudo [Conj] , a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] fase [N:f:s] de [Prep1:de] fotointerpretação [N:f:s] combinada [Adj:f:s] [Pp:f:s] ( interpretação [N:f:s] comparativa [Adj:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] de [Prep1:de] datas [N:f:pl] diferentes [Adj:2gen:pl] ) é [V:cj] muito [Adv] [Pron:indef:m:s] morosa [Adj:f:s] e [Conj:e] não [Adv] evita [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] introdução [N:f:s] de [Prep1:de] subjectividade [N:f:s] na [Prep2:na:f:s] classificação [N:f:s] das [Prep2:das:f:pl] mutações [N:f:pl] de [Prep1:de] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] . A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] utilização [N:f:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] e [Conj:e] do [Prep2:do:m:s] tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] anula [V:cj] parcialmente [Adv] os [Art:def:m:pl] [Pron:poss] [Pron:dem] inconvenientes [Adj:2gen:pl] enunciados [Adj:m:pl] [Pp:m:pl] e [Conj:e] permite [V:cj] respeitar [V:inf] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] noção [N:f:s] de [Prep1:de] análise [N:f:s] por [Prep1:por] zona [N:f:s] .

1 . Problema O problema [N:f:s] que [Adv] [Pron:rel] se [Conj] [Pron:poss] põe [V:cj] é [V:cj] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] da [Prep2:da:f:s] construção [N:f:s] de [Prep1:de] um [Art:indef] método [N:m:s] que [Adv] [Pron:rel] permita [V:cj] " simular [V:inf] " o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] processo [N:m:s] [V:cj] de [Prep1:de] fotointerpretação [N:f:s] analógica [Adj:f:s] utilizando [Ger] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] teledeteccção [N:f:s] por [Prep1:por] satélite [N:m:s] . Por [Prep1:por] outro [Pron:dem:m:s] [Pron:indef:m:s] lado [N:m:s] , trata-se de [Prep1:de] saber [V:inf] se [Conj] [Pron:poss] os [Art:def:m:pl] [Pron:poss] [Pron:dem] dados [Pp:m:pl] [N:m:pl] registados [Pp:m:pl] [Adj:m:pl] por [Prep1:por] satélite [N:m:s] possibilitam [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] classificação [N:f:s] do [Prep2:do:m:s] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] em [Prep1:em] meio [Adj:m:s] [Adv] densamente [Adv] edificado [Adj:m:s] [Pp:m:s] e [Conj:e] o [Art:def:s] [Pron:pess:3p:s]

[Pron:dem:3p:s] estudo [N:m:s] diacrónico [Adj:m:s] das [Prep2:das:f:pl] modificações [N:f:pl] ocorridas [Pp:f:pl] [Adj:f:pl] .

A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] aplicação [N:f:s] dos [Prep2:dos:m:pl] métodos [N:m:pl] de [Prep1:de] fotointerpretação [N:f:s] ao [Prep2:ao:m:s] tratamento [N:m:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] é [V:cj] possível [Adj:2gen:s] e [Conj:e] frequentemente [Adv] eficaz [Adj:2gen:s] através\_da [Loc] utilização [N:f:s] de [Prep1:de] suporte [N:m:s] [V:cj] em [Prep1:em] papel [N:m:s] , de [Prep1:de] filme [N:m:s] ou [Conj:ou] directamente [Adv] do [Prep2:do:m:s] monitor [N:m:s] da [Prep2:da:f:s] estação [N:f:s] de [Prep1:de] tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] , quer [V:cj] em [Prep1:em] modo [N:m:s] pancromático [Adj:m:s] quer [V:cj] em [Prep1:em] composição [N:f:s] colorida [Adj:f:s] [Pp:f:s] . É importante [Adj:2gen:s] referir [V:inf] que [Adv] [Pron:rel] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] resolução [N:f:s] de [Prep1:de] 10 metros [N:m:pl] das [Prep2:das:f:pl] imagens [N:f:pl] SPOT [Sigla] , em [Prep1:em] modo [N:m:s] pancromático [Adj:m:s] , permite [V:cj] obter [V:inf] uma [Art:indef:f:s] imagem [N:f:s] próxima [Adj:f:s] duma [Prep2:duma:f:s] fotografia [N:f:s] aérea [Adj:f:s] à [Prep2:à:f:s] escala [N:f:s] aproximada [Pp:f:s] [Pp:f:s] de [Prep1:de] 1 : 50 . 000 e [Conj:e] até [Adv] [Prep1:até] de [Prep1:de] 1 : 30 . 000 . A [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] fotointerpretação [N:f:s] é [V:cj] desde\_logo [Loc] possível [Adj:2gen:s] ( segmentação [N:f:s] do [Prep2:do:m:s] território [N:m:s] em [Prep1:em] entidades [N:f:pl] diferentes [Adj:2gen:pl] , identificação [N:f:s] das [Prep2:das:f:pl] grandes [Adj:2gen:pl] unidades [N:f:pl] de [Prep1:de] paisagem [N:m:s] , . . . ) mas [Adv] [Conj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] subjectividade [N:f:s] é [V:cj] inevitável [Adj:2gen:s] .

O [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] que [Adv] [Pron:rel] vamos [V:cj] expor [V:inf] é [V:cj] um [Art:indef] procedimento [N:m:s] , ainda [Adv] experimental [Adj:2gen:s] , que [Adv] [Pron:rel] permite [V:cj] reconhecer [V:inf] automaticamente [Adv] as [Art:def:f:pl] zonas [N:f:pl] que [Adv] [Pron:rel] mudaram [V:cj] de [Prep1:de] afectação [N:f:s] utilizando [Ger] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] combinação [N:f:s] automática [Adj:f:s] de [Prep1:de] índices [N:m:pl] temáticos [Adj:m:pl] calculados [Adj:m:pl] [Pp:m:pl] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] partir [V:inf] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] de [Prep1:de] alta [Adj:f:s] resolução\_espacial [N:f:s] . Consideramos [V:cj] , pois [Adv] [Conj] , a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] combinação [N:f:s] de [Prep1:de] índices [N:m:pl] temáticos [Adj:m:pl] como [Adv] [Conj] um [Art:indef] ensaio [N:m:s] preliminar [Adj:2gen:s] de [Prep1:de] " simulação [N:f:s] " do [Prep2:do:m:s] processo [N:m:s] [V:cj] de [Prep1:de] fotointerpretação [N:f:s] . Exclui-se desta [Prep2:desta:f:s] fase [N:f:s] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] determinação [N:f:s] do [Prep2:do:m:s] sentido [N:m:s] [Pp:m:s] [Adj:m:s] das [Prep2:das:f:pl] modificações [N:f:pl] ( por [Prep1:por] exemplo [N:m:s] : área [N:f:s] agrícola [Adj:2gen:s] @ habitação [N:f:s] , área [N:f:s] florestal [Adj:2gen:s] @ área [N:f:s] industrial [Adj:2gen:s] , . . . ) .

2 . Material O trabalho [N:m:s] foi [V:cj] desenvolvido [Adj:m:s] [Pp:m:s] em [Prep1:em] estações [N:f:pl] de [Prep1:de] tratamento [N:m:s] de [Prep1:de] imagem [N:f:s] de [Prep1:de] tipo [N:m:s] SUN ( SUN 3 / 80 , SUN 3 / 110 e [Conj:e] SUN

SPARC ), em [Prep1:em] ambiente [N:m:s] UNIX . Para [Prep1:para] os [Art:def:m:pl] [Pron:poss] [Pron:dem] trabalhos [N:m:pl] de [Prep1:de] tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] , de [Prep1:de] correcção [N:f:s] geométrica [Adj:f:s] de [Prep1:de] imagem [N:f:s] e [Conj:e] de [Prep1:de] gestão [N:f:s] , tratamento [N:m:s] e [Conj:e] cartografia [N:f:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] geográficos [Adj:m:pl] foram [V:cj] utilizados [Pp:m:pl] [Adj:m:pl] os [Art:def:m:pl] [Pron:poss] [Pron:dem] programas [N:m:pl] PLANETES , RECAL e [Conj:e] SAVANE , respectivamente [Adv] . Estes [Pron:dem:lp:m:pl] programas [N:m:pl] foram [V:cj] desenvolvidos [Adj:m:pl] [Pp:m:pl] integralmente [Adv] no [Prep2:no:m:s] Laboratoire d ' Informatique Appliquée ( LIA ) do [Prep2:do:m:s] Institut Français de [Prep1:de] Recherche Scientifique pour le Développement en Coopération - ORSTOM , centro [N:m:s] de [Prep1:de] Bondy , laboratório [N:m:s] onde [Adv] realizei [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] parte [N:f:s] [V:cj] de [Prep1:de] tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] .

As [Art:def:f:pl] imagens [N:f:pl] seleccionadas [Pp:f:pl] [Adj:f:pl] cobrem [V:cj] uma [Art:indef:f:s] área [N:f:s] de [Prep1:de] 10 Km [N:m:s] x 10 Km [N:m:s] , área [N:f:s] que [Adv] [Pron:rel] se [Conj] [Pron:poss] situa [V:cj] na [Prep2:na:f:s] periferia [N:f:s] da [Prep2:da:f:s] cidade [N:f:s] nova [Adj:f:s] de [Prep1:de] Marne la Vallée , a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] NE da [Prep2:da:f:s] cidade [N:f:s] de [Prep1:de] Paris , e [Conj:e] apresentam [V:cj] os [Art:def:m:pl] [Pron:poss] [Pron:dem] parâmetros [N:m:pl] que [Adv] [Pron:rel] constam [V:cj] no [Prep2:no:m:s] Quadro [N:m:s] I . Os [Art:def:m:pl] [Pron:poss] [Pron:dem] dados [Pp:m:pl] [N:m:pl] foram [V:cj] registados [Pp:m:pl] [Adj:m:pl] com [Prep1:com] inclinação [N:f:s] quase [Adv] vertical [Adj:2gen:s] # # # para [Prep1:para] quaisquer [Pron:indef:pl] dos [Prep2:dos:m:pl] canais [N:m:pl] considerados [Adj:m:pl] [Pp:m:pl] .

Para [Prep1:para] além [Adv] das [Prep2:das:f:pl] imagens [N:f:pl] enunciadas [Adj:f:pl] [Pp:f:pl] foi [V:cj] utilizada [Pp:f:s] [Adj:f:s] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] Carta [N:f:s] dos [Prep2:dos:m:pl] Modos [N:m:pl] de [Prep1:de] Ocupação [N:f:s] do [Prep2:do:m:s] Solo [N:m:s] , à [Prep2:à:f:s] escala [N:f:s] de [Prep1:de] 1 : 25000 , folha [N:f:s] nº 18 , de [Prep1:de] 1982 , elaborada [Adj:f:s] [Pp:f:s] pelo [Prep2:pelo:m:s] Institut d ' Aménagement et d ' Urbanisme de [Prep1:de] la Région Ile-de-France , no [Prep2:no:m:s] âmbito [N:m:s] do [Prep2:do:m:s] Atlas [N:m:s] de [Prep1:de] l ' Occupation du Sol [N:m:s] / Modes d ' Occupation du Sol [N:m:s] ( MOS [Sigla] ) da [Prep2:da:f:s] região [N:f:s] [N:f:s] de [Prep1:de] Ile-de-France . Esta [Pron:dem:lp:f:s] carta [N:f:s] serviu [V:cj] de [Prep1:de] base [N:f:s] para [Prep1:para] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] determinação [N:f:s] da [Prep2:da:f:s] mudança [N:f:s] de [Prep1:de] afectação [N:f:s] das [Prep2:das:f:pl] zonas [N:f:pl] MOS [Sigla] . Por [Prep1:por] zona [N:f:s] MOS [Sigla] deve [V:cj] entender-se toda [Adj:f:s] [Adv] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] unidade [N:f:s] mínima [Adj:f:s] de [Prep1:de] referência [N:f:s] espacial [Adj:2gen:s] para [Prep1:para] efeitos [N:m:pl] de [Prep1:de] classificação [N:f:s] da [Prep2:da:f:s] ocupação [N:f:s] do [Prep2:do:m:s] solo [N:m:s] à [Prep2:à:f:s] qual [Pron:rel] também [Adv] estão [V:cj] associados [Adj:m:pl] [Pp:m:pl] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] demográfica [Adj:f:s] e [Conj:e] socio-económica que [Adv] [Pron:rel] , no [Prep2:no:m:s] conjunto [N:m:s] , constituem [V:cj] uma [Art:indef:f:s] base [N:f:s] de

## **Anexo 5**

.....

Sequências extraídas com ExtracTerm,  
a partir de texto etiquetado não desambiguado

Teledeteção [N:f:s] em [Prep1:em] áreas [N:f:pl] {N:f:s + Prep1:em + N:f:pl}  
 Teledeteção [N:f:s] em [Prep1:em] áreas [N:f:pl] periurbanas [Adj:f:pl] {N:f:s + Prep1:em + N:f:pl + Adj:f:pl}  
 áreas [N:f:pl] periurbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
 combinação [N:f:s] de [Prep1:de] índices [N:m:pl] {N:f:s + Prep1:de + N:m:pl}  
 combinação [N:f:s] de [Prep1:de] índices [N:m:pl] temáticos [Adj:m:pl] {N:f:s + Prep1:de + N:m:pl + Adj:m:pl}  
 índices [N:m:pl] temáticos [Adj:m:pl] {N:m:pl + Adj:m:pl}  
 mudança [N:f:s] de [Prep1:de] uso [N:m:s] [V:cj] {N:f:s + Prep1:de + N:m:s}  
 uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 imagens [N:f:pl] digitais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
 imagens [N:f:pl] digitais [Adj:2gen:pl] Spot [Sigla] {N:f:pl + Adj:2gen:pl + Sigla}  
 imagens [N:f:pl] digitais [Adj:2gen:pl] Spot [Sigla] HRV [Sigla] {N:f:pl + Adj:2gen:pl + Sigla + Sigla}  
 Spot [Sigla] HRV [Sigla] {Sigla + Sigla}

em áreas periurbanas : combinação de índices temáticos para localizar a mudança de uso do solo com recurso às imagens digitais Spot HRV .

manifestações [N:f:pl] territoriais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
 crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] {N:m:s + Prep2:das:f:pl + N:f:pl}  
 periferias [N:f:pl] urbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
 estatística [Adj:f:s] [N:f:s] descritiva [Adj:f:s] {N:f:s + Adj:f:s}  
 análise [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] {N:f:s + Prep1:de + N:m:pl}  
 análise [N:f:s] individual [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
 análise [N:f:s] individual [Adj:2gen:s] de [Prep1:de] variáveis [Adj:2gen:pl] [N:f:pl] {N:f:s + Adj:2gen:s + Prep1:de + N:f:pl}  
 análise [N:f:s] de [Prep1:de] variáveis [Adj:2gen:pl] [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
 análise [N:f:s] em [Prep1:em] componentes [N:f:pl] {N:f:s + Prep1:em + N:f:pl}  
 análise [N:f:s] em [Prep1:em] componentes [N:f:pl] principais [Adj:2gen:pl] {N:f:s + Prep1:em + N:f:pl + Adj:2gen:pl}  
 componentes [N:f:pl] principais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
 análise [N:f:s] factorial [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
 análise [N:f:s] factorial [Adj:2gen:s] de [Prep1:de] correspondências [N:f:pl] {N:f:s + Adj:2gen:s + Prep1:de + N:f:pl}  
 classificação [N:f:s] ascendente [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
 unidades [N:f:pl] territoriais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}

As manifestações territoriais do crescimento das periferias urbanas têm sido abundantemente tratadas nas ópticas económica e financeira , social e política , demográfica e habitacional , fundiária e funcional . As técnicas e os métodos para localizar , descrever e analisar esse crescimento baseiam-se , sobretudo , na estatística descritiva e na análise de dados ( análise individual de variáveis ,

análise de variáveis duas a duas , análise em componentes principais , análise factorial de correspondências , classificação ascendente hierárquica , etc . ) . A informação é tratada diacronicamente para deduzir , a partir de escalas muito variadas , as transformações que vão ocorrendo em unidades territoriais definidas administrativamente ( freguesia , concelho , commune , département , län , etc ) .

---

abordagem [N:f:s] estatística [Adj:f:s] [N:f:s] {N:f:s + Adj:f:s}  
abordagem [N:f:s] estatística [Adj:f:s] [N:f:s] {N:f:s + N:f:s}  
estatística [Adj:f:s] [N:f:s] do [Prep2:do:m:s] fenómeno [N:m:s] {N:f:s + Prep2:do:m:s + N:m:s}  
consideração [N:f:s] da [Prep2:da:f:s] unidade [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}  
unidade [N:f:s] administrativa [Adj:f:s] {N:f:s + Adj:f:s}  
unidade [N:f:s] mínima [Adj:f:s] {N:f:s + Adj:f:s}  
unidade [N:f:s] mínima [Adj:f:s] de [Prep1:de] referenciação [N:f:s] {N:f:s + Adj:f:s + Prep1:de + N:f:s}  
referenciação [N:f:s] espacial [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
diversidade [N:f:s] de [Prep1:de] situações [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
conjunto [N:m:s] completo [Adj:m:s] {N:m:s + Adj:m:s}  
conjunto [N:m:s] completo [Adj:m:s] de [Prep1:de] variáveis [Adj:2gen:pl] [N:f:pl] {N:m:s + Adj:m:s + Prep1:de + N:f:pl}  
ocupação [N:f:s] física [Adj:f:s] [N:f:s] {N:f:s + Adj:f:s}  
ocupação [N:f:s] física [Adj:f:s] [N:f:s] {N:f:s + N:f:s}  
física [Adj:f:s] [N:f:s] do [Prep2:do:m:s] território [N:m:s] {N:f:s + Prep2:do:m:s + N:m:s}  
definição [N:f:s] de [Prep1:de] coroas [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
zonas [N:f:pl] periurbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
caracterização [N:f:s] detalhada [Adj:f:s] [Pp:f:s] {N:f:s + Adj:f:s}  
detecção [N:f:s] de [Prep1:de] descontinuidades [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
fronteiras [N:f:pl] de [Prep1:de] forma [N:f:s] {N:f:pl + Prep1:de + N:f:s}  
fronteiras [N:f:pl] de [Prep1:de] forma [N:f:s] sistemática [Adj:f:s] {N:f:pl + Prep1:de + N:f:s + Adj:f:s}  
forma [N:f:s] sistemática [Adj:f:s] {N:f:s + Adj:f:s}  
fenómeno [N:m:s] de [Prep1:de] periurbanização [N:f:s] {N:m:s + Prep1:de + N:f:s}

A abordagem estatística do fenómeno é delicada . Em primeiro lugar porque a consideração da unidade administrativa como unidade mínima de referenciação espacial esconde a diversidade de situações no seu interior . Em segundo lugar , porque dificilmente se obtém um conjunto completo de variáveis ( demográficas , socio-económicas , de uso e ocupação física do território ) que responda em simultâneo à definição de coroas , sectores ou zonas periurbanas e à sua caracterização detalhada . Os dados , quando existem , são de difícil similaridade e raramente permitem a detecção de descontinuidades e de fronteiras de forma sistemática , como exige a reflexão sobre o fenómeno de periurbanização .

---

representação [N:f:s] cartográfica [Adj:f:s] {N:f:s + Adj:f:s}  
 uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 crescimento [N:m:s] urbano [Adj:m:s] {N:m:s + Adj:m:s}  
 mutações [N:f:pl] de [Prep1:de] uso [N:m:s] [V:cj] {N:f:pl + Prep1:de + N:m:s}  
 área [N:f:s] ocupada [Pp:f:s] [Adj:f:s] {N:f:s + Adj:f:s}  
 introdução [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] {N:f:s + Prep1:de + N:m:pl}  
 dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] {N:m:pl + Prep1:de + N:f:s}  
 dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] física [Adj:f:s] [N:f:s] {N:m:pl  
 + Prep1:de + N:f:s + Adj:f:s}  
 natureza [N:f:s] física [Adj:f:s] [N:f:s] {N:f:s + Adj:f:s}  
 natureza [N:f:s] física [Adj:f:s] [N:f:s] {N:f:s + N:f:s}  
 análise [N:f:s] quantitativa [Adj:f:s] {N:f:s + Adj:f:s}  
 problema [N:f:s] de [Prep1:de] crescimento [N:m:s] {N:f:s + Prep1:de + N:m:s}  
 crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] {N:m:s + Prep2:das:f:pl +  
 N:f:pl}  
 periferias [N:f:pl] urbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}

A classificação e a representação cartográfica do uso do solo e da sua evolução permitem abordar o crescimento urbano segundo uma óptica físico-funcional . O cálculo , a partir de cartas , das mutações de uso e da área ocupada por cada uso , possibilita a introdução de dados de natureza física na análise quantitativa do problema de crescimento das periferias urbanas .

---

realização [N:f:s] de [Prep1:de] cartas [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
 cartas [N:f:pl] de [Prep1:de] uso [N:m:s] [V:cj] {N:f:pl + Prep1:de + N:m:s}  
 uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 evolução [N:f:s] de [Prep1:de] uso [N:m:s] [V:cj] {N:f:s + Prep1:de + N:m:s}  
 uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 interpretação [N:f:s] estereoscópica [Adj:f:s] {N:f:s + Adj:f:s}  
 interpretação [N:f:s] estereoscópica [Adj:f:s] de [Prep1:de] fotografia [N:f:s] {N:f:s +  
 Adj:f:s + Prep1:de + N:f:s}  
 fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Adj:f:s}  
 fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] {N:f:s + Adj:f:s + Adj:2gen:s}  
 cartas [N:f:pl] temáticas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
 uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 Carta [N:f:s] de [Prep1:de] Uso [N:m:s] [V:cj] {N:f:s + Prep1:de + N:m:s}  
 Uso [N:m:s] [V:cj] do [Prep2:do:m:s] Solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 Carta [N:f:s] de [Prep1:de] Evolução [N:f:s] {N:f:s + Prep1:de + N:f:s}  
 Evolução [N:f:s] de [Prep1:de] Uso [N:m:s] [V:cj] {N:f:s + Prep1:de + N:m:s}  
 Uso [N:m:s] [V:cj] do [Prep2:do:m:s] Solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 quantificação [N:f:s] das [Prep2:das:f:pl] cartas [N:f:pl] {N:f:s + Prep2:das:f:pl +  
 N:f:pl}  
 área [N:f:s] ocupada [Pp:f:s] [Adj:f:s] {N:f:s + Adj:f:s}  
 área [N:f:s] das [Prep2:das:f:pl] zonas [N:f:pl] {N:f:s + Prep2:das:f:pl + N:f:pl}  
 análise [N:f:s] quantitativa [Adj:f:s] {N:f:s + Adj:f:s}

transformações [N:f:pl] ocorridas [Pp:f:pl] [Adj:f:pl] {N:f:pl + Adj:f:pl}

Partindo destas duas hipóteses concluímos em 1989 a primeira etapa de um trabalho que teve por objectivo a definição duma metodologia para a realização de cartas de uso do solo e de evolução de uso do solo , por interpretação estereoscópica de fotografia aérea vertical . O resultado final é constituído por duas cartas temáticas do Concelho de Almada , onde está representado o uso do solo em 1986 ( Carta de Uso do Solo ) e a sua evolução entre 1967 e 1986 ( Carta de Evolução de Uso do Solo ) , segundo uma legenda em 12 classes . A quantificação das cartas permitiu apurar a área ocupada por cada uso em 1986 e a área das zonas que sofreram modificação entre 1967 e 1986 . Posteriormente procedeu-se à análise quantitativa das transformações ocorridas em 20 anos .

---

informação [N:f:s] obtida [Pp:f:s] [Adj:f:s] {N:f:s + Adj:f:s}  
uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
interpretação [N:f:s] de [Prep1:de] fotografia [N:f:s] {N:f:s + Prep1:de + N:f:s}  
interpretação [N:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Prep1:de + N:f:s + Adj:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Adj:f:s}  
unidade [N:f:s] de [Prep1:de] análise [N:f:s] {N:f:s + Prep1:de + N:f:s}  
zona [N:f:s] urbana [Adj:f:s] {N:f:s + Adj:f:s}  
zona [N:f:s] urbana [Adj:f:s] homogénea [Adj:f:s] {N:f:s + Adj:f:s + Adj:f:s}  
abordagem [N:f:s] do [Prep2:do:m:s] crescimento [N:m:s] {N:f:s + Prep2:do:m:s + N:m:s}  
crescimento [N:m:s] urbano [Adj:m:s] {N:m:s + Adj:m:s}  
fase [N:f:s] de [Prep1:de] fotointerpretação [N:f:s] {N:f:s + Prep1:de + N:f:s}  
fase [N:f:s] de [Prep1:de] fotointerpretação [N:f:s] combinada [Adj:f:s] [Pp:f:s] {N:f:s + Prep1:de + N:f:s + Adj:f:s}  
fotointerpretação [N:f:s] combinada [Adj:f:s] [Pp:f:s] {N:f:s + Adj:f:s}  
interpretação [N:f:s] comparativa [Adj:f:s] {N:f:s + Adj:f:s}  
interpretação [N:f:s] comparativa [Adj:f:s] de [Prep1:de] fotografia [N:f:s] {N:f:s + Adj:f:s + Prep1:de + N:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Adj:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] {N:f:s + Adj:f:s + Adj:2gen:s}  
datas [N:f:pl] diferentes [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
introdução [N:f:s] de [Prep1:de] subjectividade [N:f:s] {N:f:s + Prep1:de + N:f:s}  
classificação [N:f:s] das [Prep2:das:f:pl] mutações [N:f:pl] {N:f:s + Prep2:das:f:pl + N:f:pl}  
mutações [N:f:pl] de [Prep1:de] uso [N:m:s] [V:cj] {N:f:pl + Prep1:de + N:m:s}  
uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
utilização [N:f:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] {N:f:s + Prep2:dos:m:pl + N:m:pl}  
dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] {N:m:pl + Prep1:de + N:m:s}  
tratamento [N:m:s] numérico [Adj:m:s] {N:m:s + Adj:m:s}  
tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] {N:m:s + Adj:m:s}

+ Prep1:de + N:f:s}

noção [N:f:s] de [Prep1:de] análise [N:f:s] {N:f:s + Prep1:de + N:f:s}

A informação obtida sobre o uso do solo , por segmentação e interpretação de fotografia aérea , é extraordinariamente precisa e exaustiva e a unidade de análise , a zona urbana homogénea , responde satisfatoriamente à abordagem do crescimento urbano das periferias . Contudo , a fase de fotointerpretação combinada ( interpretação comparativa de fotografia aérea vertical de datas diferentes ) é muito morosa e não evita a introdução de subjectividade na classificação das mutações de uso do solo . A utilização dos dados de satélite e do tratamento numérico de imagem anula parcialmente os inconvenientes enunciados e permite respeitar a noção de análise por zona .

---

processo [N:m:s] [V:cj] de [Prep1:de] fotointerpretação [N:f:s] {N:m:s + Prep1:de + N:f:s}

processo [N:m:s] [V:cj] de [Prep1:de] fotointerpretação [N:f:s] analógica [Adj:f:s] {N:m:s + Prep1:de + N:f:s + Adj:f:s}

fotointerpretação [N:f:s] analógica [Adj:f:s] {N:f:s + Adj:f:s}

dados [Pp:m:pl] [N:m:pl] registados [Pp:m:pl] [Adj:m:pl] {N:m:pl + Adj:m:pl}

classificação [N:f:s] do [Prep2:do:m:s] uso [N:m:s] [V:cj] {N:f:s + Prep2:do:m:s + N:m:s}

uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}

estudo [N:m:s] diacrónico [Adj:m:s] {N:m:s + Adj:m:s}

modificações [N:f:pl] ocorridas [Pp:f:pl] [Adj:f:pl] {N:f:pl + Adj:f:pl}

1 . Problema O problema que se põe é o da construção de um método que permita " simular " o processo de fotointerpretação analógica utilizando a teledeteccção por satélite . Por outro lado , trata-se de saber se os dados registados por satélite possibilitam a classificação do uso do solo em meio densamente edificado e o estudo diacrónico das modificações ocorridas .

---

aplicação [N:f:s] dos [Prep2:dos:m:pl] métodos [N:m:pl] {N:f:s + Prep2:dos:m:pl + N:m:pl}

métodos [N:m:pl] de [Prep1:de] fotointerpretação [N:f:s] {N:m:pl + Prep1:de + N:f:s}

tratamento [N:m:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] {N:m:s + Prep2:dos:m:pl + N:m:pl}

dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] {N:m:pl + Prep1:de + N:m:s}

utilização [N:f:s] de [Prep1:de] suporte [N:m:s] [V:cj] {N:f:s + Prep1:de + N:m:s}

suporte [N:m:s] [V:cj] em [Prep1:em] papel [N:m:s] {N:m:s + Prep1:em + N:m:s}

monitor [N:m:s] da [Prep2:da:f:s] estação [N:f:s] {N:m:s + Prep2:da:f:s + N:f:s}

estação [N:f:s] de [Prep1:de] tratamento [N:m:s] {N:f:s + Prep1:de + N:m:s}

estação [N:f:s] de [Prep1:de] tratamento [N:m:s] numérico [Adj:m:s] {N:f:s + Prep1:de + N:m:s + Adj:m:s}

tratamento [N:m:s] numérico [Adj:m:s] {N:m:s + Adj:m:s}  
 tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] {N:m:s + Adj:m:s + Prep1:de + N:f:s}  
 modo [N:m:s] pancromático [Adj:m:s] {N:m:s + Adj:m:s}  
 composição [N:f:s] colorida [Adj:f:s] [Pp:f:s] {N:f:s + Adj:f:s}  
 metros [N:m:pl] das [Prep2:das:f:pl] imagens [N:f:pl] {N:m:pl + Prep2:das:f:pl + N:f:pl}  
 imagens [N:f:pl] SPOT [Sigla] {N:f:pl + Sigla}  
 modo [N:m:s] pancromático [Adj:m:s] {N:m:s + Adj:m:s}  
 imagem [N:f:s] próxima [Adj:f:s] {N:f:s + Adj:f:s}  
 fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Adj:f:s}  
 segmentação [N:f:s] do [Prep2:do:m:s] território [N:m:s] {N:f:s + Prep2:do:m:s + N:m:s}  
 território [N:m:s] em [Prep1:em] entidades [N:f:pl] {N:m:s + Prep1:em + N:f:pl}  
 território [N:m:s] em [Prep1:em] entidades [N:f:pl] diferentes [Adj:2gen:pl] {N:m:s + Prep1:em + N:f:pl + Adj:2gen:pl}  
 entidades [N:f:pl] diferentes [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
 unidades [N:f:pl] de [Prep1:de] paisagem [N:m:s] {N:f:pl + Prep1:de + N:m:s}

A aplicação dos métodos de fotointerpretação ao tratamento dos dados de satélite é possível e frequentemente eficaz através da utilização de suporte em papel, de filme ou directamente do monitor da estação de tratamento numérico de imagem, quer em modo pancromático quer em composição colorida. É importante referir que a resolução de 10 metros das imagens SPOT, em modo pancromático, permite obter uma imagem próxima duma fotografia aérea à escala aproximada de 1 : 50 . 000 e até de 1 : 30 . 000 . A fotointerpretação é desde logo possível ( segmentação do território em entidades diferentes, identificação das grandes unidades de paisagem, . . . ) mas a subjectividade é inevitável .

---

## **Anexo 6**

---

Texto etiquetado com o ExtracTerm,  
com a aplicação de regras de desambiguação

Teledetecção [N:f:s] em [Prep1:em] áreas [N:f:pl] periurbanas [Adj:f:pl] : combinação [N:f:s] de [Prep1:de] índices [N:m:pl] temáticos [Adj:m:pl] para [Prep1:para] localizar [V:inf] a [Art:def:f:s] mudança [N:f:s] de [Prep1:de] uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] com [Prep1:com] recurso [N:m:s] às [Prep2:às:s:pl] imagens [N:f:pl] digitais [Adj:2gen:pl] Spot [Sigla]HRV[Sigla].

Manifestações [N:f:pl] territoriais [Adj:2gen:pl] do [Prep2:do:m:s] crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] urbanas [Adj:f:pl] têm [V:cj] sido [Pp] abundantemente [Adv] tratadas [Pp:f:pl] [Adj:f:pl] nas [Prep2:nas:f:pl] ópticas [N:f:pl] económica [Adj:f:s] e [Conj:e] financeira [Adj:f:s] , social [Adj:2gen:s] e [Conj:e] política [N:f:s] , demográfica [Adj:f:s] e [Conj:e] habitacional [Adj:2gen:s] , fundiária [Adj:f:s] e [Conj:e] funcional [Adj:2gen:s] . As [Art:def:f:pl] técnicas [N:f:pl] e [Conj:e] os [Art:def:m:pl] [Pron:poss] [Pron:dem] métodos [N:m:pl] para [Prep1:para] localizar [V:inf] , descrever [V:inf] e [Conj:e] analisar [V:inf] esse [Pron:dem:3p:m:s] crescimento [N:m:s] baseiam-se , sobretudo [Adv] , na [Prep2:na:f:s] estatística [N:f:s] descritiva [Adj:f:s] e [Conj:e] na [Prep2:na:f:s] análise [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] ( análise [N:f:s] individual [Adj:2gen:s] de [Prep1:de] variáveis [N:f:pl] , análise [N:f:s] de [Prep1:de] variáveis [N:f:pl] duas [Num:card] a [Art:def:f:s] [Pron:poss:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] duas [Num:card] , análise [N:f:s] em [Prep1:em] componentes [N:f:pl] principais [Adj:2gen:pl] , análise [N:f:s] factorial [Adj:2gen:s] de [Prep1:de] correspondências [N:f:pl] , classificação [N:f:s] ascendente [Adj:2gen:s] hierárquica [Adj:f:s] , etc . ) . A [Art:def:f:s] informação [N:f:s] é [V:cj] tratada [Pp:f:s] diacronicamente [Adv] para [Prep1:para] deduzir [V:inf] , a [Art:def:f:s] partir [V:inf] de [Prep1:de] escalas [N:f:pl] muito [Adv] [Pron:indef:m:s] variadas [Adj:f:pl] , as [Art:def:f:pl] transformações [N:f:pl] que [Adv] [Pron:rel] vão [V:cj] ocorrendo [Ger] em [Prep1:em] unidades [N:f:pl] territoriais [Adj:2gen:pl] definidas [Pp:f:pl] administrativamente [Adv] ( freguesia [N:f:s] , concelho [N:m:s] , commune , département , län [N:] , etc ) .

abordagem [N:f:s] estatística [Adj:f:s] do [Prep2:do:m:s] fenómeno [N:m:s] é [V:cj] delicada [Adj:f:s] . Em [Prep1:em] primeiro [Adj:m:s] [Adv] [Num:ord] lugar [N:m:s] porque [Adv] [Conj] a [Art:def:f:s] consideração [N:f:s] da [Prep2:da:f:s] unidade [N:f:s] administrativa [Adj:f:s] como [Adv] [Conj] unidade [N:f:s] mínima [Adj:f:s] de [Prep1:de] referência [N:f:s] espacial [Adj:2gen:s] esconde [V:cj] a [Art:def:f:s] diversidade [N:f:s] de [Prep1:de] situações [N:f:pl] no [Prep2:no:m:s] seu [Pron:poss:3p:m:s] interior [Adj:2gen:s] [N:m:s] . Em [Prep1:em] segundo [Adv] [N:m:s] [Num:ord] lugar [N:m:s] , porque [Adv] [Conj] dificilmente [Adv] se [Conj] [Pron:poss] obtém [V:cj] um [Art:indef] conjunto [N:m:s] completo [Adj:m:s] de [Prep1:de] variáveis [N:f:pl] ( demográficas [Adj:f:pl] , socio-económicas , de [Prep1:de] uso [N:m:s] e [Conj:e] ocupação [N:f:s] física [Adj:f:s] do [Prep2:do:m:s] território [N:m:s] ) que [Adv] [Pron:rel] responda [V:cj] em [Prep1:em] simultâneo [Adj:m:s] à [Prep2:à:f:s] definição [N:f:s] de [Prep1:de] coroa [N:f:pl] , sectores [N:m:pl] ou [Conj:ou] zonas [N:f:pl] periurbanas [Adj:f:pl] e [Conj:e] à [Prep2:à:f:s] sua [Pron:poss:3p:f:s] caracterização [N:f:s] detalhada [Adj:f:s] [Pp:f:s] . Os [Art:def:m:pl] [Pron:poss] [Pron:dem] dados [Pp:m:pl] [N:m:pl] , quando [Adv] [Conj] existem [V:cj] , são [V:cj] de [Prep1:de] difícil [Adj:2gen:s] similaridade [N:f:s] e [Conj:e] raramente [Adv] permitem [V:cj] a [Art:def:f:s] detecção [N:f:s] de [Prep1:de] descontinuidades [N:f:pl] e [Conj:e] de [Prep1:de] fronteiras [N:f:pl] de [Prep1:de] forma [N:f:s] sistemática [Adj:f:s] , como [Adv] [Conj] exige [V:cj] a [Art:def:f:s]

reflexão [N:f:s] sobre [Prep1:sobre] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] fenómeno [N:m:s] de [Prep1:de] periurbanização [N:f:s] .  
classificação [N:f:s] e [Conj:e] a [Art:def:f:s] representação [N:f:s] cartográfica [Adj:f:s] do [Prep2:do:m:s] uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] e [Conj:e] da [Prep2:da:f:s] sua [Pron:pess:3p:f:s] evolução [N:f:s] permitem [V:cj] abordar [V:inf] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] crescimento [N:m:s] urbano [Adj:m:s] segundo [Adv] [N:m:s] [Num:ord] uma [Art:indef:f:s] óptica [N:f:s] fisico-funcional . O [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] cálculo [N:m:s] [N:m:s] , a [Art:def:f:s] partir [V:inf] de [Prep1:de] cartas [N:f:pl] , das [Prep2:das:f:pl] mutações [N:f:pl] de [Prep1:de] uso [N:m:s] e [Conj:e] da [Prep2:da:f:s] área [N:f:s] ocupada [Pp:f:s] por [Prep1:por] cada [Pron:indef] uso [N:m:s] [V:cj] , possibilita [V:cj] a [Art:def:f:s] introdução [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] física [Adj:f:s] na [Prep2:na:f:s] análise [N:f:s] quantitativa [Adj:f:s] do [Prep2:do:m:s] problema [N:f:s] de [Prep1:de] crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] urbanas [Adj:f:pl] .  
destas [Prep2:destas:f:pl] duas [Num:card] hipóteses [N:f:pl] concluímos [V:cj] em [Prep1:em] 1989 a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] primeira [Adj:f:s] [Adv] etapa [N:f:s] de [Prep1:de] um [Art:indef] trabalho [N:m:s] que [Adv] [Pron:rel] teve [V:cj] por [Prep1:por] objectivo [Adj:m:s] a [Art:def:f:s] definição [N:f:s] duma [Prep2:duma:f:s] metodologia [N:f:s] para [Prep1:para] a [Art:def:f:s] realização [N:f:s] de [Prep1:de] cartas [N:f:pl] de [Prep1:de] uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] e [Conj:e] de [Prep1:de] evolução [N:f:s] de [Prep1:de] uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] , por [Prep1:por] interpretação [N:f:s] estereoscópica [Adj:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] . O [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] resultado [N:m:s] final [Adj] é [V:cj] constituído [Pp:m:s] por [Prep1:por] duas [Num:card] cartas [N:f:pl] temáticas [Adj:f:pl] do [Prep2:do:m:s] Concelho [N:m:s] de [Prep1:de] Almada , onde [Adv] está [V:cj] representado [Pp:m:s] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] em [Prep1:em] 1986 ( Carta [N:f:s] de [Prep1:de] Uso [N:m:s] do [Prep2:do:m:s] Solo [N:m:s] ) e [Conj:e] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] sua [Pron:pess:3p:f:s] evolução [N:f:s] entre [Prep1:entre] 1967 e [Conj:e] 1986 ( Carta [N:f:s] de [Prep1:de] Evolução [N:f:s] de [Prep1:de] Uso [N:m:s] do [Prep2:do:m:s] Solo [N:m:s] ) , segundo [Adv] [N:m:s] [Num:ord] uma [Art:indef:f:s] legenda [N:f:s] em [Prep1:em] 12 classes [N:f:pl] . A [Art:def:f:s] quantificação [N:f:s] das [Prep2:das:f:pl] cartas [N:f:pl] permitiu [V:cj] apurar [N:inf] a [Art:def:f:s] área [N:f:s] ocupada [Pp:f:s] por [Prep1:por] cada [Pron:indef] uso [N:m:s] [V:cj] em [Prep1:em] 1986 e [Conj:e] a [Art:def:f:s] área [N:f:s] das [Prep2:das:f:pl] zonas [N:f:pl] que [Adv] [Pron:rel] sofreram [V:cj] modificação [N:f:s] entre [Prep1:entre] 1967 e [Conj:e] 1986 . Posteriormente [Adv] procedeu-se à [Prep2:à:f:s] análise [N:f:s] quantitativa [Adj:f:s] das [Prep2:das:f:pl] transformações [N:f:pl] ocorridas [Pp:f:pl] [Adj:f:pl] em [Prep1:em] 20 anos [N:m:pl] .  
informação [N:f:s] obtida [Pp:f:s] sobre [Prep1:sobre] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] , por [Prep1:por] segmentação [N:f:s] e [Conj:e] interpretação [N:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] , é [V:cj] extraordinariamente [Adv] precisa [V:cj] e [Conj:e] exaustiva [Adj:f:s] e [Conj:e] a [Art:def:f:s] unidade [N:f:s] de [Prep1:de] análise [N:f:s] , a [Art:def:f:s] zona [N:f:s] urbana [Adj:f:s] homogénea [Adj:f:s] , responde [V:cj]

satisfatoriamente [Adv] à [Prep2:à:f:s] abordagem [N:f:s] do [Prep2:do:m:s] crescimento [N:m:s] urbano [Adj:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] . Contudo [Conj] , a [Art:def:f:s] fase [N:f:s] de [Prep1:de] fotointerpretação [N:f:s] combinada [Adj:f:s] [Pp:f:s] ( interpretação [N:f:s] comparativa [Adj:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] de [Prep1:de] datas [N:f:pl] diferentes [Adj:2gen:pl] ) é [V:cj] muito [Adv] [Pron:indef:m:s] morosa [Adj:f:s] e [Conj:e] não [Adv] evita [V:cj] a [Art:def:f:s] introdução [N:f:s] de [Prep1:de] subjectividade [N:f:s] na [Prep2:na:f:s] classificação [N:f:s] das [Prep2:das:f:pl] mutações [N:f:pl] de [Prep1:de] uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] . A [Art:def:f:s] utilização [N:f:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] e [Conj:e] do [Prep2:do:m:s] tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] anula [V:cj] parcialmente [Adv] os [Art:def:m:pl] [Pron:poss] [Pron:dem] inconvenientes [Adj:2gen:pl] enunciados [Pp:m:pl] e [Conj:e] permite [V:cj] respeitar [V:inf] a [Art:def:f:s] noção [N:f:s] de [Prep1:de] análise [N:f:s] por [Prep1:por] zona [N:f:s] .

. Problema [N:f:s]

problema [N:f:s] que [Adv] [Pron:rel] se [Conj] [Pron:poss] põe [V:cj] é [V:cj] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] da [Prep2:da:f:s] construção [N:f:s] de [Prep1:de] um [Art:indef] método [N:m:s] que [Adv] [Pron:rel] permita [V:cj] " simular [V:inf] " o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] processo [N:m:s] [V:cj] de [Prep1:de] fotointerpretação [N:f:s] analógica [Adj:f:s] utilizando [Ger] a [Art:def:f:s] teledetecção [N:f:s] por [Prep1:por] satélite [N:m:s] . Por [Prep1:por] outro [Pron:dem:m:s] [Pron:indef:m:s] lado [N:m:s] , trata-se de [Prep1:de] saber [V:inf] se [Conj] [Pron:poss] os [Art:def:m:pl] [Pron:poss] [Pron:dem] dados [Pp:m:pl] [N:m:pl] registados [Pp:m:pl] por [Prep1:por] satélite [N:m:s] possibilitam [V:cj] a [Art:def:f:s] classificação [N:f:s] do [Prep2:do:m:s] uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] em [Prep1:em] meio [Adj:m:s] [Adv] densamente [Adv] edificado [Adj:m:s] [Pp:m:s] e [Conj:e] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] estudo [N:m:s] diacrónico [Adj:m:s] das [Prep2:das:f:pl] modificações [N:f:pl] ocorridas [Pp:f:pl] [Adj:f:pl] . aplicação [N:f:s] dos [Prep2:dos:m:pl] métodos [N:m:pl] de [Prep1:de] fotointerpretação [N:f:s] ao [Prep2:ao:m:s] tratamento [N:m:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] é [V:cj] possível [Adj:2gen:s] e [Conj:e] frequentemente [Adv] eficaz [Adj:2gen:s] através da [Loc] utilização [N:f:s] de [Prep1:de] suporte [N:m:s] em [Prep1:em] papel [N:m:s] , de [Prep1:de] filme [N:m:s] ou [Conj:ou] directamente [Adv] do [Prep2:do:m:s] monitor [N:m:s] da [Prep2:da:f:s] estação [N:f:s] de [Prep1:de] tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] , quer [V:cj] em [Prep1:em] modo [N:m:s] pancromático [Adj:m:s] quer [V:cj] em [Prep1:em] composição [N:f:s] colorida [Adj:f:s] [Pp:f:s] . É importante [Adj:2gen:s] referir [V:inf] que [Adv] [Pron:rel] a [Art:def:f:s] resolução [N:f:s] de [Prep1:de] 10 metros [N:m:pl] das [Prep2:das:f:pl] imagens [N:f:pl] SPOT [Sigla] , em [Prep1:em] modo [N:m:s] pancromático [Adj:m:s] , permite [V:cj] obter [V:inf] uma [Art:indef:f:s] imagem [N:f:s] próxima [Adj:f:s] duma [Prep2:duma:f:s] fotografia [N:f:s] aérea [Adj:f:s] à [Prep2:à:f:s] escala [N:f:s] aproximada [Pp:f:s] [Pp:f:s] de [Prep1:de] 1 : 50 . 000 e [Conj:e] até [Adv] [Prep1:até] de [Prep1:de] 1 : 30 . 000 . A [Art:def:f:s] fotointerpretação [N:f:s] é [V:cj] desde logo [Loc] possível [Adj:2gen:s] ( segmentação [N:f:s] do [Prep2:do:m:s] território [N:m:s] em [Prep1:em] entidades [N:f:pl] diferentes [Adj:2gen:pl] , identificação [N:f:s] das [Prep2:das:f:pl]

grandes [Adj:2gen:pl] unidades [N:f:pl] de [Prepl:de] paisagem [N:m:s] , . . . ) mas [Adv] [Conj] a [Art:def:f:s] subjectividade [N:f:s] é [V:cj] inevitável [Adj:2gen:s] . que [Adv] [Pron:rel] vamos [V:cj] expor [V:inf] é [V:cj] um [Art:indef] procedimento [N:m:s] , ainda [Adv] experimental [Adj:2gen:s] , que [Adv] [Pron:rel] permite [V:cj] reconhecer [V:inf] automaticamente [Adv] as [Art:def:f:pl] zonas [N:f:pl] que [Adv] [Pron:rel] mudaram [V:cj] de [Prepl:de] afectação [N:f:s] utilizando [Ger] a [Art:def:f:s] combinação [N:f:s] automática [Adj:f:s] de [Prepl:de] índices [N:m:pl] temáticos [Adj:m:pl] calculados [Pp:m:pl] a [Art:def:f:s] partir [V:inf] de [Prepl:de] dados [Pp:m:pl] [N:m:pl] de [Prepl:de] satélite [N:m:s] de [Prepl:de] alta [Adj:f:s] resolução\_espacial [N:f:s] . Consideramos [V:cj] , pois [Adv] [Conj] , a [Art:def:f:s] combinação [N:f:s] de [Prepl:de] índices [N:m:pl] temáticos [Adj:m:pl] como [Adv] [Conj] um [Art:indef] ensaio [N:m:s] preliminar [Adj:2gen:s] de [Prepl:de] " simulação [N:f:s] " do [Prep2:do:m:s] processo [N:m:s] de [Prepl:de] fotointerpretação [N:f:s] . Exclui-se desta [Prep2:desta:f:s] fase [N:f:s] a [Art:def:f:s] determinação [N:f:s] do [Prep2:do:m:s] sentido [N:m:s] [Pp:m:s] [Adj:m:s] das [Prep2:das:f:pl] modificações [N:f:pl] ( por [Prepl:por] exemplo [N:m:s] : área [N:f:s] agrícola [Adj:2gen:s] @ habitação [N:f:s] , área [N:f:s] florestal [Adj:2gen:s] @ área [N:f:s] industrial [Adj:2gen:s] , . . . ) .

. Material [N:m:s]

trabalho [N:m:s] foi [V:cj] desenvolvido [Pp:m:s] em [Prepl:em] estações [N:f:pl] de [Prepl:de] tratamento [N:m:s] de [Prepl:de] imagem [N:f:s] de [Prepl:de] tipo [N:m:s] SUN ( SUN 3 / 80 , SUN 3 / 110 e [Conj:e] SUN SPARC ) , em [Prepl:em] ambiente [N:m:s] UNIX . Para [Prepl:para] os [Art:def:m:pl] [Pron:poss] [Pron:dem] trabalhos [N:m:pl] de [Prepl:de] tratamento [N:m:s] numérico [Adj:m:s] de [Prepl:de] imagem [N:f:s] , de [Prepl:de] correcção [N:f:s] geométrica [Adj:f:s] de [Prepl:de] imagem [N:f:s] e [Conj:e] de [Prepl:de] gestão [N:f:s] , tratamento [N:m:s] e [Conj:e] cartografia [N:f:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] geográficos [Adj:m:pl] foram [V:cj] utilizados [Pp:m:pl] os [Art:def:m:pl] [Pron:poss] [Pron:dem] programas [N:m:pl] PLANETES , RECAL e [Conj:e] SAVANE , respectivamente [Adv] . Estes [Pron:dem:lp:m:pl] programas [N:m:pl] foram [V:cj] desenvolvidos [Pp:m:pl] integralmente [Adv] no [Prep2:no:m:s] Laboratoire d ' Informatique Appliquée ( LIA ) do [Prep2:do:m:s] Institut Français de [Prepl:de] Recherche Scientifique pour le Développement en Coopération - ORSTOM , centro [N:m:s] de [Prepl:de] Bondy , laboratório [N:m:s] onde [Adv] realizei [V:cj] a [Art:def:f:s] parte [N:f:s] [V:cj] de [Prepl:de] tratamento [N:m:s] numérico [Adj:m:s] de [Prepl:de] imagem [N:f:s] .

imagens [N:f:pl] seleccionadas [Pp:f:pl] cobrem [V:cj] uma [Art:indef:f:s] área [N:f:s] de [Prepl:de] 10 Km [N:m:s] x 10 Km [N:m:s] , área [N:f:s] que [Adv] [Pron:rel] se [Conj] [Pron:poss] situa [V:cj] na [Prep2:na:f:s] periferia [N:f:s] da [Prep2:da:f:s] cidade [N:f:s] nova [Adj:f:s] de [Prepl:de] Marne la Vallée , a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prepl:a] NE da [Prep2:da:f:s] cidade [N:f:s] de [Prepl:de] Paris , e [Conj:e] apresentam [V:cj] os [Art:def:m:pl] [Pron:poss] [Pron:dem] parâmetros [N:m:pl] que [Adv] [Pron:rel] constam [V:cj] no [Prep2:no:m:s] Quadro [N:m:s] I . Os [Art:def:m:pl] [Pron:poss] [Pron:dem] dados [Pp:m:pl] [N:m:pl] foram [V:cj] registados [Pp:m:pl] com [Prepl:com] inclinação [N:f:s] quase [Adv] vertical [Adj:2gen:s] # # # para [Prepl:para] quaisquer [Pron:indef:pl] dos [Prep2:dos:m:pl] canais [N:m:pl] considerados [Adj:m:pl] [Pp:m:pl] .

além das [Loc] imagens [N:f:pl] enunciadas [Pp:f:pl] foi [V:cj] utilizada [Pp:f:s] a [Art:def:f:s] Carta [N:f:s] dos [Prep2:dos:m:pl] Modos [N:m:pl] de [Prep1:de] Ocupação [N:f:s] do [Prep2:do:m:s] Solo [N:m:s] , à [Prep2:à:f:s] escala [N:f:s] de [Prep1:de] 1 : 25000 , folha [N:f:s] nº 18 , de [Prep1:de] 1982 , elaborada [Adj:f:s] [Pp:f:s] pelo [Prep2:pelo:m:s] Institut d ' Aménagement et d ' Urbanisme de [Prep1:de] la Région Ile-de-France , no [Prep2:no:m:s] âmbito [N:m:s] do [Prep2:do:m:s] Atlas [N:m:s] de [Prep1:de] l ' Occupation du Sol [N:m:s] / Modes d ' Occupation du Sol [N:m:s] ( MOS [Sigla] ) da [Prep2:da:f:s] região [N:f:s] [N:f:s] de [Prep1:de] Ile-de-France . Esta [Pron:dem:1p:f:s] carta [N:f:s] serviu [V:cj] de [Prep1:de] base [N:f:s] para [Prep1:para] a [Art:def:f:s] determinação [N:f:s] da [Prep2:da:f:s] mudança [N:f:s] de [Prep1:de] afectação [N:f:s] das [Prep2:das:f:pl] zonas [N:f:pl] MOS [Sigla] . Por [Prep1:por] zona [N:f:s] MOS [Sigla] deve [V:cj] entender-se toda [Adj:f:s] [Adv] a [Art:def:f:s] unidade [N:f:s] mínima [Adj:f:s] de [Prep1:de] referenciação [N:f:s] espacial [Adj:2gen:s] para [Prep1:para] efeitos [N:m:pl] de [Prep1:de] classificação [N:f:s] da [Prep2:da:f:s] ocupação [N:f:s] do [Prep2:do:m:s] solo [N:m:s] à [Prep2:à:f:s] qual [Pron:rel] também [Adv] estão [V:cj] associados [Pp:m:pl] dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] demográfica [Adj:f:s] e [Conj:e] socio-económica que [Adv] [Pron:rel] , no [Prep2:no:m:s] conjunto [N:m:s] , constituem [V:cj] uma [Art:indef:f:s] base [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] relacionáveis [Adj:2gen:pl] .

I - características [Adj:f:pl] [N:m:pl] das [Prep2:das:f:pl] imagens [N:f:pl] e [Conj:e] das [Prep2:das:f:pl] bandas [N:f:pl] espectrais [Adj:2gen:pl] [ Quadro [N:m:s] ] Método [N:m:s] .

. 1 . Procedimento [N:m:s] geral [Adj:2gen:s] .

Quadro [N:m:s] II descrevem-se as [Art:def:f:pl] etapas [N:f:pl] fundamentais [Adj:2gen:pl] do [Prep2:do:m:s] procedimento [N:m:s] adoptado [Adj:m:s] [Pp:m:s] . Esse [Pron:dem:3p:m:s] procedimento [N:m:s] apoiou-se em [Prep1:em] dois [Num:card] princípios [N:m:pl] . O [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] primeiro [Adj:m:s] [Adv] [Num:ord] , considera [V:cj] que [Adv] [Pron:rel] a [Art:def:f:s] detecção [N:f:s] das [Prep2:das:f:pl] modificações [N:f:pl] na [Prep2:na:f:s] ocupação [N:f:s] do [Prep2:do:m:s] solo [N:m:s] se [Conj] [Pron:poss] faz [V:cj] tomando [Ger] como [Adv] [Conj] unidade [N:f:s] mínima [Adj:f:s] de [Prep1:de] análise [N:f:s] a [Art:def:f:s] zona [N:f:s] MOS [Sigla] , o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] segundo [Adv] [N:m:s] [Num:ord] supõe [V:cj] que [Adv] [Pron:rel] as [Art:def:f:pl] mutações [N:f:pl] procuradas [Pp:f:pl] obrigam [V:cj] à [Prep2:à:f:s] generalização [N:f:s] da [Prep2:da:f:s] legenda [N:f:s] e [Conj:e] da [Prep2:da:f:s] Carta [N:f:s] dos [Prep2:dos:m:pl] Modos [N:m:pl] de [Prep1:de] Ocupação [N:f:s] do [Prep2:do:m:s] Solo [N:m:s] .

generalização [N:f:s] da [Prep2:da:f:s] legenda [N:f:s] - passagem [N:f:s] de [Prep1:de] 47 classes [N:f:pl] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] 12 classes [N:f:pl] de [Prep1:de] uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] - baseou-se no [Prep2:no:m:s] agrupamento [N:m:s] , numa [Prep2:numa:f:s] mesma [Adv] [N:f:s] [Pron:dem:f:s] classe [N:f:s] , dos [Prep2:dos:m:pl] usos [N:m:pl] do [Prep2:do:m:s] solo [N:m:s] pertencentes [Adj:2gen:pl] à [Prep2:à:f:s] mesma [Adv] [N:f:s] [Pron:dem:f:s] família [N:f:s] de [Prep1:de] funcionalidades [N:f:pl] .

## **Anexo 7**

Sequências extraídas com ExtracTerm,  
a partir de texto etiquetado desambiguado

Teledeteção [N:f:s] em [Prep1:em] áreas [N:f:pl] {N:f:s + Prep1:em + N:f:pl}  
 Teledeteção [N:f:s] em [Prep1:em] áreas [N:f:pl] periurbanas [Adj:f:pl] {N:f:s +  
 Prep1:em + N:f:pl + Adj:f:pl}  
 áreas [N:f:pl] periurbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
 combinação [N:f:s] de [Prep1:de] índices [N:m:pl] {N:f:s + Prep1:de + N:m:pl}  
 combinação [N:f:s] de [Prep1:de] índices [N:m:pl] temáticos [Adj:m:pl] {N:f:s +  
 Prep1:de + N:m:pl + Adj:m:pl}  
 índices [N:m:pl] temáticos [Adj:m:pl] {N:m:pl + Adj:m:pl}  
 mudança [N:f:s] de [Prep1:de] uso [N:m:s] {N:f:s + Prep1:de + N:m:s}  
 uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
 imagens [N:f:pl] digitais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
 imagens [N:f:pl] digitais [Adj:2gen:pl] Spot [Sigla] {N:f:pl + Adj:2gen:pl + Sigla}  
 imagens [N:f:pl] digitais [Adj:2gen:pl] Spot [Sigla] HRV [Sigla] {N:f:pl + Adj:2gen:pl  
 + Sigla + Sigla}  
 Spot [Sigla] HRV [Sigla] {Sigla + Sigla}

Teledeteção em áreas periurbanas : combinação de índices temáticos para  
 localizar a mudança de uso do solo com recurso às imagens digitais Spot HRV .

---

manifestações [N:f:pl] territoriais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
 crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] {N:m:s + Prep2:das:f:pl +  
 N:f:pl}  
 periferias [N:f:pl] urbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
 estatística [N:f:s] descritiva [Adj:f:s] {N:f:s + Adj:f:s}  
 análise [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] {N:f:s + Prep1:de + N:m:pl}  
 análise [N:f:s] individual [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
 análise [N:f:s] individual [Adj:2gen:s] de [Prep1:de] variáveis [N:f:pl] {N:f:s +  
 Adj:2gen:s + Prep1:de + N:f:pl}  
 análise [N:f:s] de [Prep1:de] variáveis [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
 análise [N:f:s] em [Prep1:em] componentes [N:f:pl] {N:f:s + Prep1:em + N:f:pl}  
 análise [N:f:s] em [Prep1:em] componentes [N:f:pl] principais [Adj:2gen:pl] {N:f:s +  
 Prep1:em + N:f:pl + Adj:2gen:pl}  
 componentes [N:f:pl] principais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
 análise [N:f:s] factorial [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
 análise [N:f:s] factorial [Adj:2gen:s] de [Prep1:de] correspondências [N:f:pl] {N:f:s +  
 Adj:2gen:s + Prep1:de + N:f:pl}  
 classificação [N:f:s] ascendente [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
 unidades [N:f:pl] territoriais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}

manifestações territoriais do crescimento das periferias urbanas têm sido  
 abundantemente tratadas nas ópticas económica e financeira , social e política ,  
 demográfica e habitacional , fundiária e funcional . As técnicas e os métodos  
 para localizar , descrever e analisar esse crescimento baseiam-se , sobretudo , na  
 estatística descritiva e na análise de dados ( análise individual de variáveis ,  
 análise de variáveis duas a duas , análise em componentes principais , análise

factorial de correspondências , classificação ascendente hierárquica , etc . ) . A informação é tratada diacronicamente para deduzir , a partir de escalas muito variadas , as transformações que vão ocorrendo em unidades territoriais definidas administrativamente ( freguesia , concelho , commune , département , län , etc ) .

---

abordagem [N:f:s] estatística [Adj:f:s] {N:f:s + Adj:f:s}  
consideração [N:f:s] da [Prep2:da:f:s] unidade [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}  
unidade [N:f:s] administrativa [Adj:f:s] {N:f:s + Adj:f:s}  
unidade [N:f:s] mínima [Adj:f:s]  
unidade [N:f:s] mínima [Adj:f:s] de [Prep1:de] referência [N:f:s] {N:f:s + Adj:f:s + Prep1:de + N:f:s}  
referência [N:f:s] espacial [Adj:2gens] {N:f:s + Adj:2gens}  
diversidade [N:f:s] de [Prep1:de] situações [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
conjunto [N:m:s] completo [Adj:m:s] {N:m:s + Adj:m:s}  
conjunto [N:m:s] completo [Adj:m:s] de [Prep1:de] variáveis [N:f:pl] {N:m:s + Adj:m:s + Prep1:de + N:f:pl}  
ocupação [N:f:s] física [Adj:f:s] {N:f:s + Adj:f:s}  
definição [N:f:s] de [Prep1:de] coroa [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
zonas [N:f:pl] periurbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
caracterização [N:f:s] detalhada [Adj:f:s] [Pp:f:s] {N:f:s + Adj:f:s}  
detecção [N:f:s] de [Prep1:de] descontinuidades [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
fronteiras [N:f:pl] de [Prep1:de] forma [N:f:s] {N:f:pl + Prep1:de + N:f:s}  
fronteiras [N:f:pl] de [Prep1:de] forma [N:f:s] sistemática [Adj:f:s] {N:f:pl + Prep1:de + N:f:s + Adj:f:s}  
forma [N:f:s] sistemática [Adj:f:s] {N:f:s + Adj:f:s}  
fenómeno [N:m:s] de [Prep1:de] periurbanização [N:f:s] {N:m:s + Prep1:de + N:f:s}

abordagem estatística do fenómeno é delicada . Em primeiro lugar porque a consideração da unidade administrativa como unidade mínima de referência espacial esconde a diversidade de situações no seu interior . Em segundo lugar , porque dificilmente se obtém um conjunto completo de variáveis ( demográficas , socio-económicas , de uso e ocupação física do território ) que responda em simultâneo à definição de coroa , sectores ou zonas periurbanas e à sua caracterização detalhada . Os dados , quando existem , são de difícil similaridade e raramente permitem a detecção de descontinuidades e de fronteiras de forma sistemática , como exige a reflexão sobre o fenómeno de periurbanização .

---

representação [N:f:s] cartográfica [Adj:f:s] {N:f:s + Adj:f:s}  
uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
crescimento [N:m:s] urbano [Adj:m:s] {N:m:s + Adj:m:s}  
mutações [N:f:pl] de [Prep1:de] uso [N:m:s] {N:f:pl + Prep1:de + N:m:s}  
introdução [N:f:s] de [Prep1:de] dados [Pp:m:pl] [N:m:pl] {N:f:s + Prep1:de + N:m:pl}  
dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] {N:m:pl + Prep1:de + N:f:s}

dados [Pp:m:pl] [N:m:pl] de [Prep1:de] natureza [N:f:s] física [Adj:f:s] {N:m:pl + Prep1:de + N:f:s + Adj:f:s}  
natureza [N:f:s] física [Adj:f:s] {N:f:s + Adj:f:s}  
análise [N:f:s] quantitativa [Adj:f:s] {N:f:s + Adj:f:s}  
problema [N:f:s] de [Prep1:de] crescimento [N:m:s] {N:f:s + Prep1:de + N:m:s}  
crescimento [N:m:s] das [Prep2:das:f:pl] periferias [N:f:pl] {N:m:s + Prep2:das:f:pl + N:f:pl}  
periferias [N:f:pl] urbanas [Adj:f:pl] {N:f:pl + Adj:f:pl}

classificação e a representação cartográfica do uso do solo e da sua evolução permitem abordar o crescimento urbano segundo uma óptica físico-funcional . O cálculo , a partir de cartas , das mutações de uso e da área ocupada por cada uso , possibilita a introdução de dados de natureza física na análise quantitativa do problema de crescimento das periferias urbanas .

---

realização [N:f:s] de [Prep1:de] cartas [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
cartas [N:f:pl] de [Prep1:de] uso [N:m:s] {N:f:pl + Prep1:de + N:m:s}  
uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
evolução [N:f:s] de [Prep1:de] uso [N:m:s] {N:f:s + Prep1:de + N:m:s}  
uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
interpretação [N:f:s] estereoscópica [Adj:f:s] {N:f:s + Adj:f:s}  
interpretação [N:f:s] estereoscópica [Adj:f:s] de [Prep1:de] fotografia [N:f:s] {N:f:s + Adj:f:s + Prep1:de + N:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Adj:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] {N:f:s + Adj:f:s + Adj:2gen:s}  
cartas [N:f:pl] temáticas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
uso [N:m:s] [V:cj] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
Carta [N:f:s] de [Prep1:de] Uso [N:m:s] {N:f:s + Prep1:de + N:m:s}  
Uso [N:m:s] do [Prep2:do:m:s] Solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
Carta [N:f:s] de [Prep1:de] Evolução [N:f:s] {N:f:s + Prep1:de + N:f:s}  
Evolução [N:f:s] de [Prep1:de] Uso [N:m:s] {N:f:s + Prep1:de + N:m:s}  
Uso [N:m:s] do [Prep2:do:m:s] Solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
quantificação [N:f:s] das [Prep2:das:f:pl] cartas [N:f:pl] {N:f:s + Prep2:das:f:pl + N:f:pl}  
área [N:f:s] das [Prep2:das:f:pl] zonas [N:f:pl] {N:f:s + Prep2:das:f:pl + N:f:pl}  
análise [N:f:s] quantitativa [Adj:f:s] {N:f:s + Adj:f:s}

destas duas hipóteses concluímos em 1989 a primeira etapa de um trabalho que teve por objectivo a definição duma metodologia para a realização de cartas de uso do solo e de evolução de uso do solo , por interpretação estereoscópica de fotografia aérea vertical . O resultado final é constituído por duas cartas temáticas do Concelho de Almada , onde está representado o uso do solo em 1986 ( Carta de Uso do Solo ) e a sua evolução entre 1967 e 1986 ( Carta de Evolução de Uso do Solo ) , segundo uma legenda em 12 classes . A quantificação das cartas permitiu apurar a área ocupada por cada uso em 1986 e

a área das zonas que sofreram modificação entre 1967 e 1986 . Posteriormente procedeu-se à análise quantitativa das transformações ocorridas em 20 anos .

---

uso [N:m:s] [V:e] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
interpretação [N:f:s] de [Prep1:de] fotografia [N:f:s] {N:f:s + Prep1:de + N:f:s}  
interpretação [N:f:s] de [Prep1:de] fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Prep1:de + N:f:s + Adj:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Adj:f:s}  
unidade [N:f:s] de [Prep1:de] análise [N:f:s] {N:f:s + Prep1:de + N:f:s}  
zona [N:f:s] urbana [Adj:f:s] {N:f:s + Adj:f:s}  
zona [N:f:s] urbana [Adj:f:s] homogénea [Adj:f:s] {N:f:s + Adj:f:s + Adj:f:s}  
abordagem [N:f:s] do [Prep2:do:m:s] crescimento [N:m:s] {N:f:s + Prep2:do:m:s + N:m:s}  
crescimento [N:m:s] urbano [Adj:m:s] {N:m:s + Adj:m:s}  
fase [N:f:s] de [Prep1:de] fotointerpretação [N:f:s] {N:f:s + Prep1:de + N:f:s}  
fase [N:f:s] de [Prep1:de] fotointerpretação [N:f:s] combinada [Adj:f:s] [Pp:f:s] {N:f:s + Prep1:de + N:f:s + Adj:f:s}  
fotointerpretação [N:f:s] combinada [Adj:f:s] [Pp:f:s] {N:f:s + Adj:f:s}  
interpretação [N:f:s] comparativa [Adj:f:s] {N:f:s + Adj:f:s}  
interpretação [N:f:s] comparativa [Adj:f:s] de [Prep1:de] fotografia [N:f:s] {N:f:s + Adj:f:s + Prep1:de + N:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] {N:f:s + Adj:f:s}  
fotografia [N:f:s] aérea [Adj:f:s] vertical [Adj:2gen:s] {N:f:s + Adj:f:s + Adj:2gen:s}  
datas [N:f:pl] diferentes [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
introdução [N:f:s] de [Prep1:de] subjectividade [N:f:s] {N:f:s + Prep1:de + N:f:s}  
classificação [N:f:s] das [Prep2:das:f:pl] mutações [N:f:pl] {N:f:s + Prep2:das:f:pl + N:f:pl}  
mutações [N:f:pl] de [Prep1:de] uso [N:m:s] {N:f:pl + Prep1:de + N:m:s}  
uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}  
utilização [N:f:s] dos [Prep2:dos:m:pl] dados [Pp:m:pl] [N:m:pl] {N:f:s + Prep2:dos:m:pl + N:m:pl}  
dados [Pp:m:pl] [N:m:pl] de [Prep1:de] satélite [N:m:s] {N:m:pl + Prep1:de + N:m:s}  
tratamento [N:m:s] numérico [Adj:m:s] {N:m:s + Adj:m:s}  
tratamento [N:m:s] numérico [Adj:m:s] de [Prep1:de] imagem [N:f:s] {N:m:s + Adj:m:s + Prep1:de + N:f:s}  
noção [N:f:s] de [Prep1:de] análise [N:f:s] {N:f:s + Prep1:de + N:f:s}

informação obtida sobre o uso do solo , por segmentação e interpretação de fotografia aérea , é extraordinariamente precisa e exaustiva e a unidade de análise , a zona urbana homogénea , responde satisfatoriamente à abordagem do crescimento urbano das periferias . Contudo , a fase de fotointerpretação combinada ( interpretação comparativa de fotografia aérea vertical de datas diferentes ) é muito morosa e não evita a introdução de subjectividade na classificação das mutações de uso do solo . A utilização dos dados de satélite e do tratamento numérico de imagem anula parcialmente os inconvenientes

enunciados e permite respeitar a noção de análise por zona .

---

## . Problema

---

processo [N:m:s] [V:cj] de [Prep1:de] fotointerpretação [N:f:s] {N:m:s + Prep1:de + N:f:s}

processo [N:m:s] [V:cj] de [Prep1:de] fotointerpretação [N:f:s] analógica [Adj:f:s]  
{N:m:s + Prep1:de + N:f:s + Adj:f:s}

fotointerpretação [N:f:s] analógica [Adj:f:s] {N:f:s + Adj:f:s}

classificação [N:f:s] do [Prep2:do:m:s] uso [N:m:s] {N:f:s + Prep2:do:m:s + N:m:s}

uso [N:m:s] do [Prep2:do:m:s] solo [N:m:s] {N:m:s + Prep2:do:m:s + N:m:s}

estudo [N:m:s] diacrónico [Adj:m:s] {N:m:s + Adj:m:s}

modificações [N:f:pl] ocorridas [Pp:f:pl] [Adj:f:pl] {N:f:pl + Adj:f:pl}

problema que se põe é o da construção de um método que permita " simular " o processo de fotointerpretação analógica utilizando a teledetecção por satélite .  
Por outro lado , trata-se de saber se os dados registados por satélite possibilitam a classificação do uso do solo em meio densamente edificado e o estudo diacrónico das modificações ocorridas .

---

## **Anexo 8**

---

Texto etiquetado com ExracTerm,  
a partir de novo *corpus*, com aplicação de regras de desambiguação

Capítulo [N:m:s] 1 - A [Art:def:f:s] radiação [N:f:s] electromagnética Introdução . aquisição [N:f:s] de [Prep1:de] imagens [N:f:pl] por [Prep1:por] Detecção [N:f:s] Remota [Adj:f:s] ( DR ) baseia-se no [Prep2:no:m:s] facto [N:m:s] de [Prep1:de] os [Art:def:m:pl] [Pron:poss] [Pron:dem] objectos [N:m:pl] da [Prep2:da:f:s] superfície [N:f:s] terrestre [Adj:2gen:s] reflectirem radiação\_ electromagnética [N:f:s] ( REM [Sigla] ) proveniente [Adj:2gen:s] do [Prep2:do:m:s] Sol [N:m:s] e [Conj:e] emitirem eles [Pron:pess:3p:m:pl] próprios [Adj:m:pl] REM [Sigla] . A [Art:def:f:s] intensidade [N:f:s] e [Conj:e] composição [N:f:s] espectral [Adj:2gen:s] da [Prep2:da:f:s] REM [Sigla] emitida [Adj:f:s] [Pp:f:s] assim\_ como [Loc] a [Art:def:f:s] percentagem [N:f:s] de [Prep1:de] REM [Sigla] reflectida [Pp:f:s] [Adj:f:s] dependem [V:cj] das [Prep2:das:f:pl] características [N:m:pl] físicas [Adj:f:pl] e [Conj:e] químicas [Adj:f:pl] [N:f:pl] do [Prep2:do:m:s] objecto [N:m:s] . Assim [Adv] , é [V:cj] possível [Adj:2gen:s] inferir [V:inf] das [Prep2:das:f:pl] características [N:m:pl] físicas [Adj:f:pl] e [Conj:e] químicas [Adj:f:pl] [N:f:pl] de [Prep1:de] objectos [N:m:pl] à [Prep2:à:f:s] superfície [N:f:s] da [Prep2:da:f:s] Terra [N:f:s] através [Adv] da [Prep2:da:f:s] REM [Sigla] proveniente [Adj:2gen:s] desses [Prep2:desses:m:pl] objectos [N:m:pl] e [Conj:e] registada [Pp:f:s] [Adj:f:s] por [Prep1:por] sensores [N:m:pl] a [Art:def:f:s] bordo [N:m:s] de [Prep1:de] satélites [N:m:pl] artificiais [Adj:2gen:pl] da [Prep2:da:f:s] Terra [N:f:s] , designados [Adj:m:pl] [Adj:m:pl] satélites [N:m:pl] de [Prep1:de] Detecção [N:f:s] Remota [Adj:f:s] .

Propagação [N:f:s] da [Prep2:da:f:s] Radiação [N:f:s] Electromagnética A REM [Sigla] propaga-se através\_ do [Loc] espaço [N:m:s] sem [Prep1:sem] suporte [N:m:s] material [N:m:s] e [Conj:e] só [Adj] é [V:cj] observável [Adj:2gen:s] através\_ dos [Loc] efeitos [N:m:pl] que [Adv] [Pron:rel] provoca [V:cj] . Apresenta [V:cj] propriedades [N:f:pl] ondulatórias [Adj:f:pl] quando [Adv] [Conj] se [Conj] [Pron:poss] propaga [V:cj] no [Prep2:no:m:s] espaço [N:m:s] e [Conj:e] quando [Adv] [Conj] interage com [Prep1:com] a [Art:def:f:s] matéria [N:f:s] a [Prep1:a] uma [Art:indef:f:s] escala [N:f:s] macroscópica [Adj:f:s] [Adj:f:s] e [Conj:e] apresenta [V:cj] propriedades [N:f:pl] corpusculares quando [Adv] [Conj] interage com [Prep1:com] a [Art:def:f:s] matéria [N:f:s] a [Prep1:a] uma [Art:indef:f:s] escala [N:f:s] atómica [Adj:f:s] .

fenómenos [N:m:pl] de [Prep1:de] interferência [N:f:s] e [Conj:e] difracção [N:f:s] explicam-se através\_ da [Loc] teoria [N:f:s] ondulatória [Adj:f:s] , proposta [N:f:s] [Pp:f:s] pelo [Prep2:pelo:m:s] físico [Adj:m:s] britânico James C . Maxwell ; este [Pron:dem:1p:m:s] elaborou uma [Art:indef:f:s] formulação matemática [N:f:s] para [Prep1:para] os [Art:def:m:pl] [Pron:poss] [Pron:dem] fenómenos [N:m:pl] eléctricos [Adj:m:pl] e [Conj:e] magnéticos [Adj:m:pl] que [Adv] [Pron:rel] postula que [Adv] [Pron:rel] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] é [V:cj] a [Art:def:f:s] manifestação [N:f:s] da [Prep2:da:f:s] existência [N:f:s] de [Prep1:de] dois [Num:card] campos [N:m:pl] de [Prep1:de] forças [N:f:pl] ( eléctrico [Adj:m:s] e [Conj:e] magnético [Adj:m:s] ) , conjugados [Adj:m:pl] [Pp:m:pl] ; a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] é [V:cj] gerada [Pp:f:s] quando [Adv] [Conj] se [Conj] [Pron:poss] produzem [V:cj] variações [N:f:pl] na [Prep2:na:f:s] intensidade [N:f:s] ou [Conj:ou] direcção [N:f:s] de [Prep1:de] qualquer [Pron:indef] dos [Prep2:dos:m:pl] campos [N:m:pl] de [Prep1:de] forças [N:f:pl] presentes [Adj:2gen:pl] e [Conj:e] propaga-se na [Prep2:na:f:s] forma [N:f:s] de [Prep1:de]

duas [Num:card] ondas [N:f:pl] sinusoidais [Adj:2gen:pl] , em [Prep1:em] fase [N:f:s] , situadas [Pp:f:pl] [Adj:f:pl] em [Prep1:em] dois [Num:card] planos [Adj:m:pl] [N:m:pl] ortogonais [Adj:2gen:pl] , como [Adv] [Conj] se [Conj] [Pron:poss] apresenta [V:cj] na [Prep2:na:f:s] figura [N:f:s] 1 . 1 .  
 outro [Pron:dem:m:s] [Pron:indef:m:s] lado [N:m:s] , outros [Pron:dem:m:pl] [Pron:indef:m:pl] fenómenos [N:m:pl] , tais [Adj:pl] [Pron:dem:pl] como [Adv] [Conj] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] efeito [N:m:s] fotoeléctrico , explicam-se através da [Loc] teoria [N:f:s] corpuscular , proposta [N:f:s] [Pp:f:s] por [Prep1:por] Max Planck [Np] , que [Adv] [Pron:rel] postula que [Adv] [Pron:rel] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] é [V:cj] constituída [Pp:f:s] por [Prep1:por] corpúsculos [N:m:pl] designados [Adj:m:pl] [Adj:m:pl] fótons [N:f:pl] cuja [Pron:rel:f:s] quantidade [N:f:s] de [Prep1:de] energia [N:f:s] ( Q ) é [V:cj] directamente [Adv] proporcional [Adj:2gen:s] à [Prep2:à:f:s] frequência [N:f:s] ( f ) da [Prep2:da:f:s] onda [N:f:s] de [Prep1:de] propagação [N:f:s] :

[ Fórmula [N:f:s] ]

$h = 6.626 \times 10^{-34}$  Joules é [V:cj] designada [Pp:f:s] por [Prep1:por] constante [Adj] [N:f:s] de [Prep1:de] Planck [Np] . Em [Prep1:em] Detecção [N:f:s] Remota [Adj:f:s] , analisam-se os [Art:def:m:pl] [Pron:poss] [Pron:dem] fenómenos [N:m:pl] de [Prep1:de] propagação [N:f:s] da [Prep2:da:f:s] REM [Sigla] à [Prep2:à:f:s] luz [N:f:s] da [Prep2:da:f:s] teoria [N:f:s] ondulatória [Adj:f:s] e [Conj:e] os [Art:def:m:pl] [Pron:poss] [Pron:dem] fenómenos [N:m:pl] de [Prep1:de] interacção [N:f:s] da [Prep2:da:f:s] radiação [N:f:s] com [Prep1:com] a [Art:def:f:s] matéria [N:f:s] à [Prep2:à:f:s] luz [N:f:s] da [Prep2:da:f:s] teoria [N:f:s] corpuscular . As [Art:def:f:pl] alterações [N:f:pl] que [Adv] [Pron:rel] a [Art:def:f:s] energia [N:f:s] electromagnética [Adj:f:s] provoca [V:cj] quando [Adv] [Conj] interage com [Prep1:com] um [Art:indef] corpo [N:m:s] explicam-se pela [Prep2:pela:f:s] transferência [N:f:s] da [Prep2:da:f:s] energia [N:f:s] transportada [Pp:f:s] pelos [Prep2:pelos:m:pl] fótons [N:f:pl] para [Prep1:para] os [Art:def:m:pl] [Pron:poss] [Pron:dem] átomos [N:m:pl] desse [Prep2:desse:m:s] corpo [N:m:s] .

principais [Adj:2gen:pl] sensores [N:m:pl] em [Prep1:em] operação [N:f:s] que [Adv] [Pron:rel] registam [V:cj] informação [N:f:s] para [Prep1:para] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] desenvolvimento [N:m:s] de [Prep1:de] aplicações [N:f:pl] no [Prep2:no:m:s] âmbito [N:m:s] da [Prep2:da:f:s] avaliação [N:f:s] dos [Prep2:dos:m:pl] recursos [N:m:pl] naturais [Adj:2gen:pl] são [V:cj] sensores [N:m:pl] ópticos [Adj:m:pl] que [Adv] [Pron:rel] registam [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] reflectida [Pp:f:s] [Adj:f:s] ou [Conj:ou] emitida [Adj:f:s] [Pp:f:s] pela [Prep2:pela:f:s] superfície [N:f:s] terrestre [Adj:2gen:s] ou [Conj:ou] dispositivos [N:m:pl] de [Prep1:de] microondas [N:m:pl] designados [Adj:m:pl] [Adj:m:pl] RADAR [N:m:s] ( RAdio Detection And Ranging ) . Este [Pron:dem:1p:m:s] capítulo [N:m:s] vai [V:cj] descrever [V:inf] os [Art:def:m:pl] [Pron:poss] [Pron:dem] processos [N:m:pl] subjacentes à [Prep2:à:f:s] aquisição [N:f:s] de [Prep1:de] imagens [N:f:pl] por [Prep1:por] sensores [N:m:pl] ópticos [Adj:m:pl] , sensíveis [Adj:2gen:pl] às [Prep2:às:s:pl] bandas [N:f:pl] em [Prep1:em] que [Adv] [Pron:rel] se [Conj] [Pron:poss] propaga [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] Solar [Adj:2gen:s] que [Adv] [Pron:rel]

se [Conj] [Pron:poss] reflecte [V:cj] na [Prep2:na:f:s] superfície [N:f:s] terrestre [Adj:2gen:s] .

Espectro [N:m:s] Electromagnético [Adj:m:s]

ondas\_electromagnéticas [N:f:pl] propagam-se , no [Prep2:no:m:s] vácuo [Adj:m:s] , à [Prep2:à:f:s] velocidade [N:f:s] da [Prep2:da:f:s] luz [N:f:s] (  $c = 299792458$  m / s ) e [Conj:e] com [Prep1:com] uma [Art:indef:f:s] frequência [N:f:s] (  $f$  ) e [Conj:e] um [Art:indef] comprimento [N:m:s] de [Prep1:de] onda [N:f:s] ( ) tais [Adj:pl] [Pron:dem:pl] que [Adv] [Pron:rel] :

[ Fórmula [N:f:s] ]

frequência [N:f:s] e [Conj:e] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] comprimento [N:m:s] de [Prep1:de] onda [N:f:s] estabelecem [V:cj] a [Art:def:f:s] forma [N:f:s] da [Prep2:da:f:s] onda [N:f:s] electromagnética [Adj:f:s] e [Conj:e] esta [Pron:dem:1p:f:s] está [V:cj] relacionada [Pp:f:s] com [Prep1:com] a [Art:def:f:s] quantidade [N:f:s] de [Prep1:de] energia [N:f:s] transportada [Pp:f:s] [Adj:f:s] através\_da [Loc] expressão [N:f:s] [N:f:s] que [Adv] [Pron:rel] se [Conj] [Pron:poss] obtém [V:cj] resolvendo [Ger] ( 1 . 2 ) em [Prep1:em] ordem [N:f:s] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a]  $f$  e [Conj:e] substituindo [Ger] em [Prep1:em] ( 1 . 1 ) :

[ Fórmula [N:f:s] ]

assim\_que [Loc] a [Art:def:f:s] quantidade [N:f:s] de [Prep1:de] energia [N:f:s] transportada [Pp:f:s] pela [Prep2:pela:f:s] onda [N:f:s] electromagnética [Adj:f:s] é [V:cj] inversamente [Adv] proporcional [Adj:2gen:s] ao [Prep2:ao:m:s] comprimento [N:m:s] de [Prep1:de] onda [N:f:s] . Este [Pron:dem:1p:m:s] efeito [N:m:s] é [V:cj] responsável [Adj:2gen:s] pelo [Prep2:pelo:m:s] facto [N:m:s] de [Prep1:de] sensores [N:m:pl] que [Adv] [Pron:rel] funcionam [V:cj] em [Prep1:em] bandas [N:f:pl] de [Prep1:de] comprimento [N:m:s] de [Prep1:de] onda [N:f:s] maiores [Adj:2gen:pl] , como [Adv] [Conj] por [Prep1:por] exemplo [N:m:s] as [Art:def:f:pl] bandas [N:f:pl] térmicas [Adj:f:pl] do [Prep2:do:m:s] satélite [N:m:s] Landsat [Sigla] , terem [V:cj] de [Prep1:de] registar [V:inf] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] de [Prep1:de] um [Art:indef] elemento [N:m:s] de [Prep1:de] terreno [N:m:s] de [Prep1:de] maiores [Adj:2gen:pl] dimensões [N:f:pl] , para [Prep1:para] que [Adv] [Pron:rel] a [Art:def:f:s] quantidade [N:f:s] de [Prep1:de] energia [N:f:s] que [Adv] [Pron:rel] chega [V:cj] ao [Prep2:ao:m:s] sensor [N:m:s] seja [V:cj] detectável [Adj:2gen:s] . Por [Prep1:por] outro [Pron:dem:m:s] [Pron:indef:m:s] lado [N:m:s] explica [V:cj] também [Adv] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] facto [N:m:s] de [Prep1:de] os [Art:def:m:pl] [Pron:poss] [Pron:dem] sensores [N:m:pl] de [Prep1:de] alta [Adj:f:s] resolução\_espacial [N:f:s] ( que [Adv] [Pron:rel] detectam [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] que [Adv] [Pron:rel] parte [N:f:s] [V:cj] de [Prep1:de] elementos [N:m:pl] de [Prep1:de] terreno [N:m:s] de [Prep1:de] pequenas [Adj:f:pl] dimensões [N:f:pl] ) ficarem limitados [Pp:m:pl] [Adj:m:pl] a [Art:def:f:s] comprimentos [N:m:pl] de [Prep1:de] onda [N:f:s] mais [Adj] [Adv] [Pron:indef] baixos [Adj:m:pl] como [Adv] [Conj] as [Art:def:f:pl] bandas [N:f:pl] do [Prep2:do:m:s] visível [Adj:2gen:s] e [Conj:e] infravermelho [Adj:m:s] [N:m:s] próximo [Adj:m:s] .

REM [Sigla] é [V:cj] classificada [Pp:f:s] em [Prep1:em] função [N:f:s] da [Prep2:da:f:s] sua [Pron:poss:3p:f:s] frequência [N:f:s] ( da [Prep2:da:f:s] sua

[Pron:poss:3p:f:s] quantidade [N:f:s] de [Prep1:de] energia [N:f:s] ) por [Prep1:por] intermédio [Adj:m:s] do [Prep2:do:m:s] espectro [N:m:s] electromagnético [Adj:m:s] que [Adv] [Pron:rel] consiste [V:cj] numa [Prep2:numa:f:s] escala [N:f:s] de [Prep1:de] frequências [N:f:pl] , cujos [Pron:rel:m:pl] intervalos [N:m:pl] são [V:cj] designados [Adj:m:pl] [Adj:m:pl] bandas [N:f:pl] espectrais [Adj:2gen:pl] Fig . 1 . 2 - Espectro [N:m:s] Electromagnético [Adj:m:s]

sensores [N:m:pl] multispectrais [Adj:2gen:pl] de [Prep1:de] REM [Sigla] colocados [Adj:m:pl] [Pp:m:pl] em [Prep1:em] plataformas [N:f:pl] espaciais [Adj:2gen:pl] , são [V:cj] concebidos [Pp:m:pl] para [Prep1:para] adquirir [V:inf] informação [N:f:s] em [Prep1:em] bandas [N:f:pl] muito [Adv] [Pron:indef:m:s] estreitas [Adj:f:pl] do [Prep2:do:m:s] espectro [N:m:s] electromagnético [Adj:m:s] . A [Art:def:f:s] localização [N:f:s] destas [Prep2:destas:f:pl] bandas [N:f:pl] está [V:cj] condicionada [Pp:f:s] às [Prep2:às:s:pl] zonas [N:f:pl] do [Prep2:do:m:s] espectro [N:m:s] , designadas [Adj:f:pl] [Pp:f:pl] por [Prep1:por] janelas [N:f:pl] espectrais [Adj:2gen:pl] , em [Prep1:em] que [Adv] [Pron:rel] a [Art:def:f:s] atmosfera [N:f:s] deixa [V:cj] passar [V:inf] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] proveniente [Adj:2gen:s] do [Prep2:do:m:s] Sol [N:m:s] e [Conj:e] àquelas [Prep2:àquelas:f:pl] zonas [N:f:pl] em [Prep1:em] que [Adv] [Pron:rel] a [Art:def:f:s] Terra [N:f:s] emite [V:cj] REM [Sigla] .

banda [N:f:s] de [Prep1:de] rádio [N:f:s] [N:m:s] cobre [V:cj] a [Art:def:f:s] região [N:f:s] [N:f:s] dos [Prep2:dos:m:pl] comprimentos [N:m:pl] de [Prep1:de] onda [N:f:s] superiores [Adj:2gen:pl] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] 10 cm ; esta [Pron:dem:1p:f:s] banda [N:f:s] é [V:cj] utilizada [Pp:f:s] , principalmente [Adv] , por [Prep1:por] sensores [N:m:pl] de [Prep1:de] rádio [N:f:s] [N:m:s] activos [Adj:m:pl] como [Adv] [Conj] os [Art:def:m:pl] [Pron:poss] [Pron:dem] dispositivos [N:m:pl] de [Prep1:de] RADAR [N:m:s] de [Prep1:de] imagem [N:f:s] , os [Art:def:m:pl] [Pron:poss] [Pron:dem] altímetros e [Conj:e] os [Art:def:m:pl] [Pron:poss] [Pron:dem] sonares . A [Art:def:f:s] banda [N:f:s] das [Prep2:das:f:pl] microondas [N:m:pl] cobre [V:cj] a [Art:def:f:s] região [N:f:s] [N:f:s] adjacente [Adj:2gen:s] àquela [Prep2:àquela:f:s] , até\_ao [Loc] comprimento [N:m:s] de [Prep1:de] onda [N:f:s] de [Prep1:de] 1mm ; é [V:cj] utilizada [Pp:f:s] por [Prep1:por] radiómetros [N:m:pl] de [Prep1:de] microondas [N:m:pl] , espectómetros [N:m:pl] e [Conj:e] sistemas [N:m:pl] de [Prep1:de] RADAR [N:m:s] . A [Art:def:f:s] banda [N:f:s] dos [Prep2:dos:m:pl] infravermelhos [N:m:pl] cobre [V:cj] a [Art:def:f:s] região [N:f:s] [N:f:s] espectral [Adj:2gen:s] desde [Prep1:desde] o [Art:def:f:s] [Pron:pess:3p:s] [Pron:dem:3p:s] comprimento [N:m:s] de [Prep1:de] onda [N:f:s] de [Prep1:de] 1mm até [Adv] [Prep1:até] 0 . 7 ( m ; esta [Pron:dem:1p:f:s] região [N:f:s] [N:f:s] é [V:cj] subdividida [Pp:f:s] em [Prep1:em] sub-regiões designadas [Adj:f:pl] [Pp:f:pl] infra- vermelho [Adj:m:s] longínquo [Adj:m:s] , infravermelho [Adj:m:s] [N:m:s] térmico [Adj:m:s] e [Conj:e] infravermelho [Adj:m:s] [N:m:s] próximo [Adj:m:s] . Nestas [Prep2:nestas:f:pl] bandas [N:f:pl] utilizam-se muitos [Adv] [Pron:indef:m:pl] sistemas [N:m:pl] de [Prep1:de] imagem [N:f:s] : espectro- radiómetros , radiómetros [N:m:pl] , polarímetros e [Conj:e] lasers . Estes [Pron:dem:1p:m:pl] sistemas [N:m:pl] também [Adv] são [V:cj] utilizados [Pp:m:pl] na [Prep2:na:f:s] banda [N:f:s] do [Prep2:do:m:s] visível [Adj:2gen:s] ,

de [Prep1:de] 0 . 4 ( m - 0 . 7 ( m . As [Art:def:f:pl] bandas [N:f:pl] dos [Prep2:dos:m:pl] ultravioletas [Adj:2gen:pl] , dos [Prep2:dos:m:pl] raios [N:m:pl] x e [Conj:e] dos [Prep2:dos:m:pl] raios [N:m:pl] gama [N:m:s] não [Adv] são [V:cj] muito [Adv] [Pron:indef:m:s] utilizadas [Pp:f:pl] [Adj:f:pl] em [Prep1:em] DR por\_causa\_da [Loc] opacidade dos [Prep2:dos:m:pl] gases [N:m:pl] atmosféricos [Adj:m:pl] em [Prep1:em] relação [N:f:s] à [Prep2:à:f:s] REM [Sigla] proveniente [Adj:2gen:s] do [Prep2:do:m:s] Sol [N:m:s] , nestes [Prep2:nestes:m:pl] comprimentos [N:m:pl] de [Prep1:de] onda [N:f:s] . e [Conj:e] Definição [N:f:s] das [Prep2:das:f:pl] Grandezas [N:f:pl] Radiométricas [Adj:f:pl] .

neste [Prep2:neste:m:s] parágrafo [N:m:s] algumas [Pron:indef:f:pl] grandezas [N:f:pl] radiométricas [Adj:f:pl] utilizadas [Pp:f:pl] para [Prep1:para] caracterizar [V:inf] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] e [Conj:e] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] sua [Pron:poss:3p:f:s] interacção [N:f:s] com [Prep1:com] a [Art:def:f:s] matéria [N:f:s] , cuja [Pron:rel:f:s] síntese [N:f:s] se [Conj] [Pron:poss] apresenta [V:cj] no [Prep2:no:m:s] quadro [N:m:s] 1 . 1 .

Radiante [Adj:2gen:s] : a [Art:def:f:s] energia [N:f:s] radiante [Adj:2gen:s] é [V:cj] a [Art:def:f:s] quantidade [N:f:s] de [Prep1:de] energia [N:f:s] transportada [Pp:f:s] pela [Prep2:pela:f:s] REM [Sigla] . É uma [Art:indef:f:s] medida [N:f:s] [Pp:f:s] [Adj:f:s] da [Prep2:da:f:s] capacidade [N:f:s] que [Adv] [Pron:rel] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] REM [Sigla] tem [V:cj] de [Prep1:de] produzir [V:inf] alterações [N:f:pl] nos [Prep2:nos:m:pl] objectos [N:m:pl] em [Prep1:em] que [Adv] [Pron:rel] incide [V:cj] , alterando [Ger] a [Art:def:f:s] sua [Pron:poss:3p:f:s] temperatura [N:f:s] ou [Conj:ou] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] seu [Pron:poss:3p:m:s] estado [N:m:s] físico [Adj:m:s] . A [Art:def:f:s] quantidade [N:f:s] de [Prep1:de] energia [N:f:s] radiante [Adj:2gen:s] por [Prep1:por] unidade [N:f:s] de [Prep1:de] volume [N:m:s] é [V:cj] designada [Pp:f:s] densidade [N:f:s] de [Prep1:de] energia [N:f:s] radiante [Adj:2gen:s] ( W )

Radiante [Adj:2gen:s] : é [V:cj] a [Art:def:f:s] velocidade [N:f:s] com [Prep1:com] que [Adv] [Pron:rel] a [Art:def:f:s] energia [N:f:s] radiante [Adj:2gen:s] chega [V:cj] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] um [Art:indef] determinado [Adj:m:s] [Pp:m:s] ponto [N:m:s] .

de [Prep1:de] Fluxo [N:m:s] Radiante [Adj:2gen:s] : é [V:cj] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] fluxo [N:m:s] radiante [Adj:2gen:s] que [Adv] [Pron:rel] chega [V:cj] ( irradiância [N:f:s] - E [Conj:e] ) a [Prep1:a] uma [Art:indef:f:s] unidade [N:f:s] de [Prep1:de] área [N:f:s] de [Prep1:de] uma [Art:indef:f:s] superfície [N:f:s] plana [Adj:f:s] , ou [Conj:ou] que [Adv] [Pron:rel] parte [N:f:s] [V:cj] ( exitância - M ) , por [Prep1:por] reflexão [N:f:s] ou [Conj:ou] emissão [N:f:s] , da [Prep2:da:f:s] unidade [N:f:s] de [Prep1:de] área [N:f:s] de [Prep1:de] uma [Art:indef:f:s] superfície [N:f:s] plana [Adj:f:s] .

[ Quadro [N:m:s] ] - Grandezas [N:f:pl] Radiométricas [Adj:f:pl]

Radiante [Adj:2gen:s] : é [V:cj] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] fluxo [N:m:s] radiante [Adj:2gen:s] por [Prep1:por] unidade [N:f:s] de [Prep1:de] ângulo [N:m:s] sólido [Adj:m:s] , que [Adv] [Pron:rel] deixa [V:cj] uma [Art:indef:f:s] fonte [N:f:s] pontual [Adj:2gen:s] de [Prep1:de] REM [Sigla] , numa [Prep2:numa:f:s] determinada [Adj:f:s] [Pp:f:s] direcção [N:f:s] .

: é [V:cj] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] fluxo [N:m:s] radiante [Adj:2gen:s] , por [Prep1:por] unidade [N:f:s] de [Prep1:de] ângulo [N:m:s] sólido [Adj:m:s] , que [Adv] [Pron:rel] deixa [V:cj] uma [Art:indef:f:s] fonte [N:f:s] de [Prep1:de] radiação [N:f:s] ( não [Adv] pontual [Adj:2gen:s] ) , numa [Prep2:numa:f:s] dada [Pp:f:s] direcção [N:f:s] , por [Prep1:por] unidade [N:f:s] de [Prep1:de] área [N:f:s] projectada [Pp:f:s] [Adj:f:s] da [Prep2:da:f:s] fonte [N:f:s] , nessa [Prep2:nessa:f:s] direcção [N:f:s] .

. 1 . 3 - Noção [N:f:s] de [Prep1:de] Radiância Os sensores [N:m:pl] ópticos [Adj:m:pl] instalados [Pp:m:pl] em [Prep1:em] plataformas [N:f:pl] espaciais [Adj:2gen:pl] medem [V:cj] a [Art:def:f:s] radiância [N:f:s] dos [Prep2:dos:m:pl] objectos [N:m:pl] à [Prep2:à:f:s] superfície [N:f:s] da [Prep2:da:f:s] Terra [N:f:s] . Para [Prep1:para] cada [Pron:indef] elemento [N:m:s] de [Prep1:de] área [N:f:s] do [Prep2:do:m:s] terreno [N:m:s] , a [Art:def:f:s] informação [N:f:s] registada [Pp:f:s] [Adj:f:s] , por [Prep1:por] banda [N:f:s] espectral [Adj:2gen:s] , é [V:cj] um [Art:indef] número [N:m:s] inteiro [Adj:m:s] proporcional [Adj:2gen:s] à [Prep2:à:f:s] radiância [N:f:s] desse [Prep2:desse:m:s] elemento [N:m:s] de [Prep1:de] área [N:f:s] .

imagem [N:f:s] numérica [Adj:f:s] que [Adv] [Pron:rel] o [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] utilizador [N:m:s] recebe [V:cj] para [Prep1:para] processamento [N:m:s] consiste [V:cj] num [Prep2:num:m:s] conjunto [N:m:s] de [Prep1:de] matrizes [N:f:pl] Inteiras [Adj:f:pl] , tantas [Pron:indef:f:pl] quantas [Pron:interr] as [Art:def:f:pl] bandas [N:f:pl] espectrais [Adj:2gen:pl] em [Prep1:em] que [Adv] [Pron:rel] a [Art:def:f:s] imagem [N:f:s] foi [V:cj] registada [Pp:f:s] . Cada [Pron:indef] elemento [N:m:s] da [Prep2:da:f:s] matriz [N:f:s] é [V:cj] designado [Pp:m:s] [Pp:m:s] elemento [N:m:s] de [Prep1:de] imagem [N:f:s] ou [Conj:ou] píxel [N:m:s] ( " picture element " ) . O [Art:def:s] [Pron:pess:3p:s] [Pron:dem:3p:s] Inteiro [Adj:m:s] atribuído [Pp:m:s] a [Art:def:f:s] [Pron:pess:3p:f:s] [Pron:dem:3p:f:s] [Prep1:a] cada [Pron:indef] elemento [N:m:s] de [Prep1:de] imagem [N:f:s] é [V:cj] designado [Pp:m:s] [Pp:m:s] por [Prep1:por] nível [N:m:s] de [Prep1:de] intensidade [N:f:s] radiométrica [Adj:f:s] ou [Conj:ou] nível [N:m:s] radiométrico [Adj:m:s] ( NR ) .

## **Anexo 9**

---

Sequências extraídas com ExtracTerm,  
a partir de texto etiquetado desambiguado

---

aquisição [N:f:s] de [Prep1:de] imagens [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
Detecção [N:f:s] Remota [Adj:f:s] {N:f:s + Adj:f:s}  
objectos [N:m:pl] da [Prep2:da:f:s] superfície [N:f:s] {N:m:pl + Prep2:da:f:s + N:f:s}  
superfície [N:f:s] terrestre [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
composição [N:f:s] espectral [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
químicas [Adj:f:pl] [N:f:pl] do [Prep2:do:m:s] objecto [N:m:s] {N:f:pl + Prep2:do:m:s + N:m:s}  
químicas [Adj:f:pl] [N:f:pl] de [Prep1:de] objectos [N:m:pl] {N:f:pl + Prep1:de + N:m:pl}  
superfície [N:f:s] da [Prep2:da:f:s] Terra [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}  
bordo [N:m:s] de [Prep1:de] satélites [N:m:pl] {N:m:s + Prep1:de + N:m:pl}  
satélites [N:m:pl] artificiais [Adj:2gen:pl] {N:m:pl + Adj:2gen:pl}  
satélites [N:m:pl] de [Prep1:de] Detecção [N:f:s] {N:m:pl + Prep1:de + N:f:s}  
satélites [N:m:pl] de [Prep1:de] Detecção [N:f:s] Remota [Adj:f:s] {N:m:pl + Prep1:de + N:f:s + Adj:f:s}  
Detecção [N:f:s] Remota [Adj:f:s] {N:f:s + Adj:f:s}

aquisição de imagens por Detecção Remota ( DR ) baseia-se no facto de os objectos da superfície terrestre reflectirem radiação electromagnética ( REM ) proveniente do Sol e emitirem eles próprios REM . A intensidade e composição espectral da REM emitida assim como a percentagem de REM reflectida dependem das características físicas e químicas do objecto . Assim , é possível inferir das características físicas e químicas de objectos à superfície da Terra através da REM proveniente desses objectos e registada por sensores a bordo de satélites artificiais da Terra , designados satélites de Detecção Remota .

---

Propagação [N:f:s] da [Prep2:da:f:s] Radiação [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}  
propriedades [N:f:pl] ondulatórias [Adj:f:pl] {N:f:pl + Adj:f:pl}  
escala [N:f:s] macroscópica [Adj:f:s] [Adj:f:s] {N:f:s + Adj:f:s}  
escala [N:f:s] atómica [Adj:f:s] {N:f:s + Adj:f:s}

Propagação da Radiação Electromagnética A REM propaga-se através do espaço sem suporte material e só é observável através dos efeitos que provoca . Apresenta propriedades ondulatórias quando se propaga no espaço e quando interage com a matéria a uma escala macroscópica e apresenta propriedades corpusculares quando interage com a matéria a uma escala atómica .

---

fenómenos [N:m:pl] de [Prep1:de] interferência [N:f:s] {N:m:pl + Prep1:de + N:f:s}  
teoria [N:f:s] ondulatória [Adj:f:s] {N:f:s + Adj:f:s}  
fenómenos [N:m:pl] eléctricos [Adj:m:pl] {N:m:pl + Adj:m:pl}  
fenómenos [N:m:pl] eléctricos [Adj:m:pl] e [Conj:e] magnéticos [Adj:m:pl]

{N:m:pl + Adj:m:pl + Conje + Adj:m:pl}

manifestação [N:f:s] da [Prep2:da:f:s] existência [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}

campos [N:m:pl] de [Prep1:de] forças [N:f:pl] {N:m:pl + Prep1:de + N:f:pl}

campos [N:m:pl] de [Prep1:de] forças [N:f:pl] {N:m:pl + Prep1:de + N:f:pl}

forças [N:f:pl] presentes [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}

ondas [N:f:pl] sinusoidais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}

planos [Adj:m:pl] [N:m:pl] ortogonais [Adj:2gen:pl] {N:m:pl + Adj:2gen:pl}

fenómenos de interferência e difracção explicam-se através da teoria ondulatória , proposta pelo fisico britânico James C . Maxwell ; este elaborou uma formulação matemática para os fenómenos eléctricos e magnéticos que postula que a REM é a manifestação da existência de dois campos de forças ( eléctrico e magnético ) , conjugados ; a REM é gerada quando se produzem variações na intensidade ou direcção de qualquer dos campos de forças presentes e propaga-se na forma de duas ondas sinusoidais , em fase , situadas em dois planos ortogonais , como se apresenta na figura 1 . 1 .

---

corpúsculos [N:m:pl] designados [Adj:m:pl] {N:m:pl + Adj:m:pl}

quantidade [N:f:s] de [Prep1:de] energia [N:f:s] {N:f:s + Prep1:de + N:f:s}

onda [N:f:s] de [Prep1:de] propagação [N:f:s] {N:f:s + Prep1:de + N:f:s}

outro lado , outros fenómenos , tais como o efeito fotoeléctrico , explicam-se através da teoria corpuscular , proposta por Max Planck , que postula que a REM é constituída por corpúsculos designados fotões cuja quantidade de energia (  $Q$  ) é directamente proporcional à frequência (  $f$  ) da onda de propagação :

---

[ Fórmula ]

---

constante [Adj] [N:f:s] de [Prep1:de] Planck [Np] {N:f:s + Prep1:de + Np}

Detecção [N:f:s] Remota [Adj:f:s] {N:f:s + Adj:f:s}

fenómenos [N:m:pl] de [Prep1:de] propagação [N:f:s] {N:m:pl + Prep1:de + N:f:s}

luz [N:f:s] da [Prep2:da:f:s] teoria [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}

teoria [N:f:s] ondulatória [Adj:f:s] {N:f:s + Adj:f:s}

fenómenos [N:m:pl] de [Prep1:de] interacção [N:f:s] {N:m:pl + Prep1:de + N:f:s}

interacção [N:f:s] da [Prep2:da:f:s] radiação [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}

luz [N:f:s] da [Prep2:da:f:s] teoria [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}

energia [N:f:s] electromagnética [Adj:f:s] {N:f:s + Adj:f:s}

transferência [N:f:s] da [Prep2:da:f:s] energia [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}

$h = 6.626 \times 10^{-34}$  Joules é designada por constante de Planck . Em Detecção Remota , analisam-se os fenómenos de propagação da REM à luz da teoria

ondulatória e os fenómenos de interacção da radiação com a matéria à luz da teoria corpuscular . As alterações que a energia electromagnética provoca quando interage com um corpo explicam-se pela transferência da energia transportada pelos fotões para os átomos desse corpo .

---

sensores [N:m:pl] em [Prep1:em] operação [N:f:s] {N:m:pl + Prep1:em + N:f:s}  
desenvolvimento [N:m:s] de [Prep1:de] aplicações [N:f:pl] {N:m:s + Prep1:de + N:f:pl}  
âmbito [N:m:s] da [Prep2:da:f:s] avaliação [N:f:s] {N:m:s + Prep2:da:f:s + N:f:s}  
avaliação [N:f:s] dos [Prep2:dos:m:pl] recursos [N:m:pl] {N:f:s + Prep2:dos:m:pl + N:m:pl}  
recursos [N:m:pl] naturais [Adj:2gen:pl] {N:m:pl + Adj:2gen:pl}  
sensores [N:m:pl] ópticos [Adj:m:pl] {N:m:pl + Adj:m:pl}  
superfície [N:f:s] terrestre [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
dispositivos [N:m:pl] de [Prep1:de] microondas [N:m:pl] {N:m:pl + Prep1:de + N:m:pl}  
dispositivos [N:m:pl] de [Prep1:de] microondas [N:m:pl] designados [Pp:m:pl][Adj:m:pl] {N:m:pl + Prep1:de + N:m:pl + Adj:m:pl}  
microondas [N:m:pl] designados [Adj:m:pl] {N:m:pl + Adj:m:pl}  
aquisição [N:f:s] de [Prep1:de] imagens [N:f:pl] {N:f:s + Prep1:de + N:f:pl}  
sensores [N:m:pl] ópticos [Adj:m:pl] {N:m:pl + Adj:m:pl}  
superfície [N:f:s] terrestre [Adj:2gen:s] {N:f:s + Adj:2gen:s}

principais sensores em operação que registam informação para o desenvolvimento de aplicações no âmbito da avaliação dos recursos naturais são sensores ópticos que registam a REM reflectida ou emitida pela superfície terrestre ou dispositivos de microondas designados RADAR ( RADio Detection And Ranging ) . Este capítulo vai descrever os processos subjacentes à aquisição de imagens por sensores ópticos , sensíveis às bandas em que se propaga a REM Solar que se reflecte na superfície terrestre .

---

Espectro [N:m:s] Electromagnético [Adj:m:s] {N:m:s + Adj:m:s}

Espectro Electromagnético

---

velocidade [N:f:s] da [Prep2:da:f:s] luz [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}  
comprimento [N:m:s] de [Prep1:de] onda [N:f:s] {N:m:s + Prep1:de + N:f:s}

ondas\_ electromagnéticas propagam-se , no vácuo , à velocidade da luz (  $c = 299792458 \text{ m / s}$  ) e com uma frequência (  $f$  ) e um comprimento de onda (  $\lambda$  ) tais que :

---

[ Fórmula ]

comprimento [N:m:s] de [Prep1:de] onda [N:f:s] {N:m:s + Prep1:de + N:f:s}  
forma [N:f:s] da [Prep2:da:f:s] onda [N:f:s] {N:f:s + Prep2:da:f:s + N:f:s}  
onda [N:f:s] electromagnética [Adj:f:s] {N:f:s + Adj:f:s}  
quantidade [N:f:s] de [Prep1:de] energia [N:f:s] {N:f:s + Prep1:de + N:f:s}  
quantidade [N:f:s] de [Prep1:de] energia [N:f:s] transportada [Pp:f:s] [Adj:f:s]  
{N:f:s + Prep1:de + N:f:s + Adj:f:s}  
energia [N:f:s] transportada [Pp:f:s] [Adj:f:s] {N:f:s + Adj:f:s}

frequência e o comprimento de onda estabelecem a forma da onda  
electromagnética e esta está relacionada com a quantidade de energia  
transportada através da expressão que se obtém resolvendo ( 1 . 2 ) em ordem a  
f e substituindo em ( 1 . 1 ) :

---

[ Fórmula ]

---

quantidade [N:f:s] de [Prep1:de] energia [N:f:s] {N:f:s + Prep1:de + N:f:s}  
onda [N:f:s] electromagnética [Adj:f:s] {N:f:s + Adj:f:s}  
comprimento [N:m:s] de [Prep1:de] onda [N:f:s] {N:m:s + Prep1:de + N:f:s}  
facto [N:m:s] de [Prep1:de] sensores [N:m:pl] {N:m:s + Prep1:de + N:m:pl}  
bandas [N:f:pl] de [Prep1:de] comprimento [N:m:s] {N:f:pl + Prep1:de + N:m:s}  
comprimento [N:m:s] de [Prep1:de] onda [N:f:s] {N:m:s + Prep1:de + N:f:s}  
bandas [N:f:pl] térmicas [Adj:f:pl] {N:f:pl + Adj:f:pl}  
satélite [N:m:s] Landsat [Sigla] {N:m:s + Sigla}  
elemento [N:m:s] de [Prep1:de] terreno [N:m:s] {N:m:s + Prep1:de + N:m:s}  
quantidade [N:f:s] de [Prep1:de] energia [N:f:s] {N:f:s + Prep1:de + N:f:s}  
parte [N:f:s] [V:cj] de [Prep1:de] elementos [N:m:pl] {N:f:s + Prep1:de + N:m:pl}  
elementos [N:m:pl] de [Prep1:de] terreno [N:m:s] {N:m:pl + Prep1:de + N:m:s}  
comprimentos [N:m:pl] de [Prep1:de] onda [N:f:s] {N:m:pl + Prep1:de + N:f:s}  
infravermelho [Adj:m:s] [N:m:s] próximo [Adj:m:s] {N:m:s + Adj:m:s}

assim\_ que a quantidade de energia transportada pela onda electromagnética é  
inversamente proporcional ao comprimento de onda . Este efeito é responsável  
pelo facto de sensores que funcionam em bandas de comprimento de onda  
maiores , como por exemplo as bandas térmicas do satélite Landsat , terem de  
registar a REM de um elemento de terreno de maiores dimensões , para que a  
quantidade de energia que chega ao sensor seja detectável . Por outro lado  
explica também o facto de os sensores de alta resolução\_espacial ( que detectam  
a REM que parte de elementos de terreno de pequenas dimensões ) ficarem  
limitados a comprimentos de onda mais baixos como as bandas do visível e  
infravermelho próximo .

---

quantidade [N:f:s] de [Prep1:de] energia [N:f:s] {N:f:s + Prep1:de + N:f:s}  
espectro [N:m:s] electromagnético [Adj:m:s] {N:m:s + Adj:m:s}  
escala [N:f:s] de [Prep1:de] frequências [N:f:pl] {N:f:s + Prep1:de + N:f:pl}

bandas [N:f:pl] espectrais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
Espectro [N:m:s] Electromagnético [Adj:m:s] {N:m:s + Adj:m:s}

REM é classificada em função da sua frequência ( da sua quantidade de energia )  
por intermédio do espectro electromagnético que consiste numa escala de  
frequências , cujos intervalos são designados bandas espectrais Fig . 1 . 2 -  
Espectro Electromagnético

---

sensores [N:m:pl] multiespectrais [Adj:2gen:pl] {N:m:pl + Adj:2gen:pl}  
plataformas [N:f:pl] espaciais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}  
informação [N:f:s] em [Prep1:em] bandas [N:f:pl] {N:f:s + Prep1:em + N:f:pl}  
espectro [N:m:s] electromagnético [Adj:m:s] {N:m:s + Adj:m:s}  
zonas [N:f:pl] do [Prep2:do:m:s] espectro [N:m:s] {N:f:pl + Prep2:do:m:s +  
N:m:s}  
janelas [N:f:pl] espectrais [Adj:2gen:pl] {N:f:pl + Adj:2gen:pl}

sensores multiespectrais de REM colocados em plataformas espaciais , são  
concebidos para adquirir informação em bandas muito estreitas do espectro  
electromagnético . A localização destas bandas está condicionada às zonas do  
espectro , designadas por janelas espectrais , em que a atmosfera deixa passar a  
REM proveniente do Sol e àquelas zonas em que a Terra emite REM .

---

banda [N:f:s] de [Prep1:de] rádio [N:f:s] {N:f:s + Prep1:de + N:m:s}  
banda [N:f:s] de [Prep1:de] rádio [N:f:s] {N:f:s + Prep1:de + N:f:s}  
região [N:f:s] [N:f:s] dos [Prep2:dos:m:pl] comprimentos [N:m:pl] {N:f:s +  
Prep2:dos:m:pl + N:m:pl}  
comprimentos [N:m:pl] de [Prep1:de] onda [N:f:s] {N:m:pl + Prep1:de + N:f:s}  
sensores [N:m:pl] de [Prep1:de] rádio [N:f:s] {N:m:pl + Prep1:de + N:m:s}  
sensores [N:m:pl] de [Prep1:de] rádio [N:f:s] {N:m:pl + Prep1:de + N:f:s}  
dispositivos [N:m:pl] de [Prep1:de] RADAR [Sigla] {N:m:pl + Prep1:de + Sigla}  
RADAR [Sigla] de [Prep1:de] imagem [N:f:s] {Sigla+ Prep1:de + N:f:s}  
região [N:f:s] [N:f:s] adjacente [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
comprimento [N:m:s] de [Prep1:de] onda [N:f:s] {N:m:s + Prep1:de + N:f:s}  
radiómetros [N:m:pl] de [Prep1:de] microondas [N:m:pl] {N:m:pl + Prep1:de +  
N:m:pl}  
sistemas [N:m:pl] de [Prep1:de] RADAR [Sigla] {N:m:pl + Prep1:de + Sigla}  
banda [N:f:s] dos [Prep2:dos:m:pl] infravermelhos [N:m:pl] N:f:s +  
Prep2:dos:m:pl + N:m:pl}  
região [N:f:s] [N:f:s] espectral [Adj:2gen:s] {N:f:s + Adj:2gen:s}  
comprimento [N:m:s] de [Prep1:de] onda [N:f:s] {N:m:s + Prep1:de + N:f:s}  
infravermelho [Adj:m:s] [N:m:s] térmico [Adj:m:s] {N:m:s + Adj:m:s}  
infravermelho [Adj:m:s] [N:m:s] térmico [Adj:m:s] e [Conj:e] infravermelho  
[Adj:m:s] [N:m:s] {N:m:s + Adj:m:s + Conj:e + Adj:m:s}  
infravermelho [Adj:m:s] [N:m:s] próximo [Adj:m:s] {N:m:s + Adj:m:s}  
sistemas [N:m:pl] de [Prep1:de] imagem [N:f:s] {N:m:pl + Prep1:de + N:f:s}  
gases [N:m:pl] atmosféricos [Adj:m:pl]  
gases [N:m:pl] atmosféricos [Adj:m:pl] em [Prep1:em] relação [N:f:s] {N:m:pl +

Adj:m:pl + Prep1:em + N:f:s}

comprimentos [N:m:pl] de [Prep1:de] onda [N:f:s] {N:m:pl + Prep1:de + N:f:s}

banda de rádio cobre a região dos comprimentos de onda superiores a 10 cm ; esta banda é utilizada , principalmente , por sensores de rádioactivos como os dispositivos de RADAR de imagem , os altímetros e os sonares . A banda das microondas cobre a região adjacente àquela , até ao comprimento de onda de 1mm ; é utilizada por radiómetros de microondas , espectómetros e sistemas de RADAR . A banda dos infravermelhos cobre a região espectral desde o comprimento de onda de 1mm até 0 . 7 ( m ; esta região é subdividida em sub-regiões designadas infra- vermelho longínquo , infravermelho térmico e infravermelho próximo . Nestas bandas utilizam-se muitos sistemas de imagem : espectro-radiómetros , radiómetros , polarímetros e lasers . Estes sistemas também são utilizados na banda do visível , de 0 . 4 ( m - 0 . 7 ( m . As bandas dos ultravioletas , dos raios x e dos raios gama não são muito utilizadas em DR por causa da opacidade dos gases atmosféricos em relação à REM proveniente do Sol , nestes comprimentos de onda .

---

