

## **Identificar, Ordenar e Relacionar Eventos**

**Viviana Isabel Guerreiro Grave Cabrita**

Dissertação para obtenção do Grau de Mestre em  
**Engenharia Informática e de Computadores**

Orientadores: Prof. Doutor Nuno João Neves Mamede  
Prof. Doutor Jorge Manuel Evangelista Baptista

### **Júri**

Presidente: Prof. Doutor João Emílio Segurado Pavão Martins  
Orientador: Prof. Doutor Nuno João Neves Mamede  
Vogal: Prof. Doutor Bruno Emanuel da Graça Martins

**Novembro 2014**



# Agradecimentos

Gostaria de agradecer ao meu orientador, Professor Nuno João Neves Mamede, e ao meu co-orientador, Professor Jorge Manuel Evangelista Baptista, pelo apoio e disponibilidade dada ao longo do desenvolvimento da dissertação, assim como a experiência e a atenção ao detalhe, tornando possível o constante aperfeiçoamento deste trabalho.

Deixo também um agradecimento a Vera Cabarrão e a Cláudio Diniz pela ajuda e colaboração dada.

Um obrigado especial aos meus pais e ao meu irmão pelo suporte e compreensão durante o decorrer deste trabalho. E, por fim, um agradecimento grande aos meus amigos que me acompanharam durante este percurso, preenchendo o final desta etapa de muitos momentos de força e vivacidade.

Lisboa, 21 de Setembro de 2014  
Viviana Isabel Guerreiro Grave Cabrita



# Resumo

As operações de extração de informação e análise de texto destacam-se nas tarefas de Processamento de Língua Natural (PLN) e são essenciais para o desenvolvimento e melhoria de sistemas capazes de, entre outras aplicações, realizar sumarizações e traduções de texto sem a intervenção humana.

O trabalho descrito neste trabalho foca-se na identificação, relação e ordenação temporal de eventos, e contribuiu para o desenvolvimento da cadeia de processamento STRING (Statistical and Rule-Based Natural Language Processing), desenvolvido pelo Laboratório de Sistemas de Língua Falada (L<sup>2</sup>F) do Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento em Lisboa (INESC-ID).

Este documento contém um estudo de diversos sistemas de ordenação de eventos já existentes e dos verbos, nomes predicativos, conectores discursivos e advérbios de tempo no português, analisando a influência das conjunções, das preposições e dos advérbios na ordenação de eventos. Estendeu-se a STRING de modo a identificar mais eventos e foi desenvolvido e integrado um módulo responsável por detectar relações entre os eventos e ordená-los em relação com a linha do tempo. A tarefa de ordenação de eventos subdividiu-se em 4 fases: identificação de eventos; relações temporais; relações entre eventos; ordenação dos eventos. A identificação das relações é feita com base na informação já extraída pela STRING e no estudo efetuado em torno dos eventos e dos conectores discursivos no português. A partir das relações extraídas, dos eventos e das expressões temporais associadas aos eventos, o módulo extrai a ordem expressa entre os eventos.

Para a avaliação, anotou-se um corpus a partir de diversos excertos de conteúdo jornalístico, literário e enciclopédico, com 162 frases, contendo 101 eventos e 607 relações de ordem entre os eventos. Procedeu-se à avaliação das várias fases da tarefa de ordenação de eventos, em separado e como um todo, obtendo-se respetivamente, uma medida-f de 72.8%, 36.4%, 39%, 44.3% e 12.2%.



# Abstract

The operations of analysis, extraction and processing of text stands out within the Natural Language Processing (NLP) task and are essential for the development and improvement of systems capable of, among other applications, to perform summarizations and translations of text without human intervention.

The work described in this paper focused on the identification, matching and temporal ordering of events, and contributed to the development of the processing chain STRING (Statistical and Rule-Based Natural Language Processing), developed by the Laboratory for Spoken Language Systems (L<sup>2</sup>F) of Institute of Systems and Computer Engineering - Research and Development in Lisbon (INESC-ID).

This document contains also a study of verbs, predicate names, connectors and adverbs in Portuguese, studying the influence of conjunctions, prepositions and adverbs in events sorting.

The system STRING was extended in order to identify more events and it was developed a module for detecting relations between events and sort them in a timeline. The task of sorting events was subdivided into four phases: identification of events; temporal relations; relationships between events; event sorting. The identification of relationships is based on the information already extracted by STRING and the study made around events and discourse connectores in Portuguese. Using the extracted relationships and temporal expressions associated with events, the module sorts those events.

For the evaluation, a corpus was annotated from several excerpts from journalistic, literary and encyclopedic content, with 162 phrases, containing 101 events and 607 order relations between events. Proceeded to the evaluation of the various phases of the task of event sorting, separately and as a whole, it was obtained, respectively, a measure-f of 72.8%, 36.4%, 39%, 44.3% and 12.2%.





# Palavras-Chave

## Palavras-Chave

Extração de eventos

Ordenação de eventos

Expressões temporais

Processamento de Língua Natural (PLN)

Português

## Keywords

Event extraction

Event sorting

Temporal Expressions

Natural Language Processing (NLP)

Portuguese



# Índice

<b>Agradecimentos</b>	<b>i</b>
<b>Resumo</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>Palavras-Chave</b>	<b>vii</b>
<b>Lista de Figuras</b>	<b>xiii</b>
<b>Lista de Tabelas</b>	<b>xv</b>
<b>Lista de Abreviaturas</b>	<b>xvii</b>
<b>Lista de Termos</b>	<b>xix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Definição do Problema . . . . .	1
1.2 Objetivos . . . . .	2
1.3 Conceitos Básicos . . . . .	2
1.3.1 Definição de Evento . . . . .	2
1.3.2 Processamento de Texto . . . . .	4
1.3.3 Lógicas Temporais . . . . .	5
1.4 Contributos . . . . .	6
1.5 Estrutura do Documento . . . . .	6
<b>2 Estado da Arte</b>	<b>7</b>
2.1 Esquemas de Anotação . . . . .	7
2.1.1 STAG . . . . .	7
2.1.2 TIDES . . . . .	8
2.1.3 TimeML . . . . .	8
2.2 Sistemas de Ordenação Temporal de Eventos . . . . .	11
2.2.1 Terseo . . . . .	11
2.2.2 XTM . . . . .	12

2.2.3	Chambers et al.	13
2.2.4	TRIPS e TRIOS	14
2.2.5	NCSU-INDI e NCSU-JOINT	16
2.2.6	Maršić	16
2.3	Cadeia de processamento STRING	18
2.3.1	Arquitetura	18
2.3.2	Classificação de Entidades Temporais	19
2.3.3	Normalização de Entidades Temporais	20
2.3.4	Eventos que são atualmente normalizados	22
2.3.5	Dependências Auxiliares	22
2.4	Sinopse	26
<b>3</b>	<b>Identificação e Ordenação de Eventos Não Normalizados em Português</b>	<b>27</b>
3.1	Eventos	27
3.1.1	Nominalizações	27
3.1.2	Verbos	28
3.1.3	Nomes predicativos	32
3.2	Conectores Discursivos	32
3.2.1	Conjunções Coordenativas	33
3.2.2	Conjunções Subordinativas	34
3.2.3	Preposições	36
3.3	Advérbios e Locuções Adverbiais	38
3.4	Conclusão	39
<b>4</b>	<b>Identificação, Relação e Ordenação de Eventos</b>	<b>41</b>
4.1	Arquitetura	41
4.1.1	Estrutura do Módulo de Ordenação de Eventos	42
4.1.2	Ficheiros de Entrada e Saída	42
4.2	Identificação de Relações de Eventos	43
4.2.1	Dependências	43
4.2.2	Extração com base nas dependências existentes	45
4.3	Ordenação de Eventos	48
4.3.1	Dependências	48
4.3.2	Análise dos Eventos	49
4.3.3	Análise dos Conectores	50
4.3.4	Análise das Dependências	51
4.3.5	Representação gráfica da ordenação	51
4.4	Implementação	54

<b>5</b>	<b>Avaliação</b>	<b>57</b>
5.1	Anotação . . . . .	57
5.2	Métodos de Avaliação e Técnicas Utilizadas . . . . .	57
5.2.1	Avaliação Estrita e Relaxada . . . . .	58
5.2.2	Métricas Utilizadas . . . . .	59
5.2.3	Delimitação das tarefas a avaliar . . . . .	60
5.3	Resultados . . . . .	61
5.4	Discussão . . . . .	64
<b>6</b>	<b>Conclusão e Trabalhos Futuros</b>	<b>71</b>
	<b>Referências</b>	<b>73</b>
<b>A</b>	<b>Corpus</b>	<b>77</b>
<b>B</b>	<b>Diretivas de classificação e anotação</b>	<b>83</b>



# Lista de Figuras

1.1	Representação das relações presentes na frase <i>Chad esperava-me um tanto impaciente, quando cheguei a casa. Tinha trazido o jantar do restaurante chinês da esquina.</i> . . . . .	2
1.2	Ordenação dos eventos presentes na frase <i>Isto aconteceu em Março e aquilo em Junho desse ano.</i> . . . . .	2
1.3	Exemplos na representação de Reinhenbach, considerando o <i>instante do evento (E), de referência (R) e da frase (F).</i> . . . . .	5
1.4	Representação das relações de Allen. . . . .	6
2.1	Estrutura do sistema TERSEO . . . . .	11
2.2	Representação da arquitetura da cadeia STRING. . . . .	18
4.1	Representação da arquitetura da solução. . . . .	41
4.2	Representação da arquitetura interna do módulo de Ordenação de Eventos. . . . .	42
4.3	Representação da arquitetura interna do módulo de Ordenação de Eventos. . . . .	43
4.4	Condições necessárias para o evento A ser anterior ao B. . . . .	48
4.5	Condições necessárias para dois eventos serem simultâneos entre si. . . . .	48
4.6	Relações adicionais a considerar em relação às fronteiras de início e fim de eventos. . . . .	49
4.7	Representação de eventos através de uma linha cronológica. . . . .	52
4.8	Representação da ordem dos eventos através de uma linha cronológica. . . . .	52
4.9	Representação da ordem dos eventos através de um grafo. . . . .	52
4.10	Exemplo complexo da representação da ordem dos eventos através de um grafo. . . . .	53
4.11	Exemplo de representação da ordem de eventos e expressões temporais através de um grafo. . . . .	53
4.12	Exemplo da árvore sintática gerada pela STRING. . . . .	54
4.13	Exemplo da árvore de relações gerado pelo módulo de ordenação de eventos. . . . .	56
4.14	Exemplo do grafo de ordem gerado pelo módulo de ordenação de eventos. . . . .	56
5.1	Representação de vários grafos gerados sobre as dependências extraídas da solução. . . . .	64
5.2	Representação de grafos com muitas relações de ordem <i>simultâneo</i> . . . . .	65
5.3	Comparação entre a representação adotada na solução, à esquerda, com o resultado de alinhar os eventos <i>simultâneos</i> . . . . .	69
5.4	Comparação entre a representação gráfica da solução atual, à esquerda, com a solução ideal, à direita. . . . .	69





# Lista de Tabelas

2.1	Tipos de expressões temporais . . . . .	9
2.2	Relações aspetuais . . . . .	10
2.3	Resultados do sistema TERSEO . . . . .	12
2.4	Resultados da avaliação restrita na relação temporal de eventos . . . . .	13
2.5	Resultados da avaliação relaxada na relação temporal de eventos . . . . .	13
2.6	Resultados dos testes . . . . .	14
2.7	Resultados obtidos na extração de eventos . . . . .	15
2.8	Resultados obtidos na relação temporal de eventos . . . . .	15
2.9	Resultados obtidos na relação temporal de eventos . . . . .	16
2.10	Resultados da avaliação restrita na relação temporal de eventos . . . . .	17
2.11	Resultados da avaliação relaxada na relação temporal de eventos . . . . .	17
2.12	Exemplos de expressões temporais identificadas pela STRING . . . . .	20
2.13	Exemplos do uso do atributo VAL_NORM . . . . .	21
2.14	Conversão entre unidades temporais . . . . .	21
2.15	Tipos de eventos, palavras-chave e respectivos exemplos . . . . .	23
3.1	Classes morfossintáticas consideradas na ordenação de eventos . . . . .	32
3.2	Exemplos de orações subordinadas temporais . . . . .	36
5.1	Detalhes do corpus de teste utilizado para avaliação do sistema. . . . .	58
5.2	Exemplo das iterações realizadas pelo sistema de avaliação sobre resultados obtidos do processo de avaliação . . . . .	61
5.3	Resultados obtidos na avaliação estrita em termos de valores <i>corretos</i> , <i>espúrios</i> e <i>em falta</i> . . . . .	62
5.4	Resultados obtidos na avaliação estrita com base nas métricas <i>precisão</i> , <i>abrangência</i> e <i>medida-f</i> . . . . .	62
5.5	Resultados obtidos na avaliação relaxada em termos de valores <i>corretos</i> , <i>espúrios</i> e <i>em falta</i> . . . . .	63
5.6	Resultados obtidos na avaliação relaxada com base nas métricas <i>precisão</i> , <i>abrangência</i> e <i>medida-f</i> . . . . .	63
5.7	Matriz de confusão obtida a partir dos resultados . . . . .	63



# Lista de Abreviaturas

<b>ALINK</b>	Aspectual <b>LINK</b>	Ligação Aspectual
<b>DARPA</b>	<b>D</b> efense <b>A</b> dvanced <b>R</b> esearch <b>P</b> rojects <b>A</b> gency	Agência de Projetos de Pesquisa Avançada de Defesa
<b>HMM</b>	<b>H</b> idden <b>M</b> arkov <b>M</b> odel	Modelos Escondidos de Markov
<b>INESC-ID</b>		<b>I</b> nstituto <b>E</b> ngenharia e de <b>S</b> istemas e <b>C</b> omputadores - <b>I</b> nvestigação e <b>D</b> esenvolvimento
<b>LexMan</b>	<b>L</b> exical <b>M</b> orfological <b>A</b> nalyzer	Analizador Lexical Morfológico
<b>L<sup>2</sup>F</b>	<b>L</b> exical <b>2</b> <b>F</b> eature-based <b>L</b> anguage <b>S</b> ystems <b>L</b> aboratory	Laboratório de Sistemas de Língua Falada
<b>MARv</b>	<b>M</b> orphosyntactic <b>A</b> mbiguity <b>R</b> esolver	[módulo de] Desambiguação Morfossintática (estatístico)
<b>MaxEnt</b>	<b>M</b> axim <b>E</b> ntropy	<b>M</b> áxima <b>E</b> ntropia
<b>RuDriCo</b>	<b>R</b> ule- <b>D</b> riven <b>C</b> onverter	Conversor baseado em regras
<b>SLINK</b>	<b>S</b> ubordination <b>LINK</b>	Ligação de Subordinação
<b>STAG</b>	<b>S</b> heffield <b>T</b> emporal <b>A</b> notation <b>G</b> uidelines	Diretivas de Anotação Temporal
<b>STRING</b>	<b>S</b> tatistical and <b>R</b> ule-based natural lan- guage processing	
<b>TERSEO</b>	<b>T</b> emporal <b>E</b> xpression <b>R</b> esolution <b>S</b> ystem Applied to <b>E</b> vent <b>O</b> rdering	Sistema de Resolução de Expressões Temporais Aplicada à Ordenação de Eventos
<b>TIDES</b>	<b>T</b> ranslingual <b>I</b> nformation <b>D</b> etection, <b>E</b> xtraction and <b>S</b> ummarization	
<b>TLINK</b>	<b>T</b> emporal <b>LINK</b>	Ligação Temporal
<b>XIP</b>	<b>X</b> erox <b>I</b> cremental <b>P</b> arser	Analizador Sintático (da Xerox)
<b>XTM</b>	<b>X</b> IP <b>T</b> emporal <b>M</b> odule	Módulo temporal do analisador Sintático (da Xerox)
<b>XML</b>	<b>E</b> xtensible <b>M</b> arkup <b>L</b> anguage	



# Lista de Termos

<b>Termo</b>	<b>Significado</b>
<b>Abrangência</b>	Medida de avaliação que mede a proporção de respostas corretas obtidas pelo sistema, em relação às respostas corretas existentes/possíveis.
<b>Corpus</b>	Conjunto de textos (escritos ou orais) de uma dada língua (ou variedade de uma língua), coligido com base em critérios explícitos, com vista a um objetivo de investigação preciso e que obedece a critérios de amostragem de forma a poder ser considerado representativo da língua (ou variedade) de que constitui um modelo à escala.
<b>Medida-f</b>	Medida de avaliação que combina as medidas de <i>precisão</i> e <i>abrangência</i> , usando-se geralmente a <i>média harmónica</i> , isto é, atribuindo igual peso à precisão e à abrangência.
<b>Traço</b>	Especificação de um atributo e respectivo valor.
<b>Precisão</b>	Medida de avaliação que mede a proporção de respostas corretas em relação ao conjunto de respostas obtidos.



# Capítulo 1

## Introdução

Este trabalho enquadra-se na área do Processamento de Língua Natural, pretendendo contribuir para o desenvolvimento da cadeia de processamento STRING (Statistical and Rule-Based Natural Language Processing) [34], desenvolvida pelo Laboratório de Sistemas de Língua Falada (L<sup>2</sup>F) do Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID). O objectivo deste projeto consistiu em detetar e relacionar eventos entre si, ordenando-os num eixo temporal. Esta nova funcionalidade foi integrada como um novo módulo da cadeia.

### 1.1 Definição do Problema

A identificação, classificação e ordenação de eventos destacam-se, entre outros, na atividade de extração e análise automática de informação, de modo a permitir uma melhor compreensão, sumarização e processamento do conteúdo relevante dos textos. Considere-se o segmento:

*“Chad esperava-me um tanto impaciente, quando cheguei a casa. Tinha trazido o jantar do restaurante chinês da esquina.”*

Os eventos envolvidos nestas duas frases são *esperar(-me)*, *chegar* e *trazer*. Através da análise do tempo do verbo *esperar-me*, determina-se que Chad tinha começado a esperar antes de o sujeito da enunciação (eu) ter chegado a casa, e, inclusive, ainda estava à espera no momento de chegada. Com base no tempo composto do verbo *trazer* e na frase anterior, é possível determinar que o evento *trazer o jantar* ocorreu ainda antes do evento *chegar a casa*, ainda que não seja possível detectar se Chad já estava à espera. A Figura 1.1 representa as relações entre os três eventos presentes neste exemplo<sup>1</sup>.

Além dos tempos verbais, as expressões temporais desempenham igualmente uma função importante na ordenação de eventos. No exemplo usado, todos os eventos são anteriores ao momento de enunciação. Assim sendo, mesmo quando não é possível relacionar diretamente dois eventos entre si, como acontece entre *esperar(-me)* e *trazer*, é possível estabelecer uma relação temporal entre os eventos e o momento de enunciação, como no exemplo *Isto aconteceu em Março e aquilo em Junho desse ano*, representado na Figura 1.2.

---

<sup>1</sup>É possível considerar ainda o evento expresso pelo adjetivo *impaciente*, concomitante com o evento *esperar(-me)*, o que não foi aqui levado em consideração.

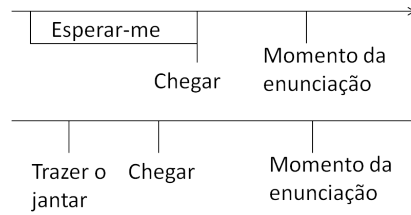


Figura 1.1: Representação das relações presentes na frase *Chad esperava-me um tanto impaciente, quando cheguei a casa. Tinha trazido o jantar do restaurante chinês da esquina.*

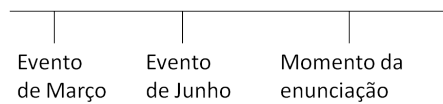


Figura 1.2: Ordenação dos eventos presentes na frase *Isto aconteceu em Março e aquilo em Junho desse ano.*

## 1.2 Objetivos

Pretendeu-se identificar os eventos que ainda não são identificados pela cadeia STRING [34] e estender as funcionalidades da STRING, criando um módulo com a capacidade de relacionar e ordenar os eventos, com os seguintes requisitos:

- 1- Estabelecer, no mínimo, as relações de *antes*, *simultâneo* ou *desconhecido* entre os eventos<sup>2</sup>;
- 2- Considerar os eventos como pontos no tempo, ignorando a duração dos mesmos para a respetiva ordenação;
- 3- Permitir uma visualização gráfica da ordem estabelecida entre os eventos.

O objetivo deste trabalho inclui ainda a anotação de um *corpus* de teste, a definição da estratégia a adotar para relacionar os eventos e a construção do novo módulo com as funcionalidades acima descritas e a sua integração na STRING.

## 1.3 Conceitos Básicos

Segue-se uma breve descrição do conceito de evento e das relações entre eventos considerados neste estudo. De seguida, descrevem-se alguns problemas a ter em conta no desenvolvimento dos sistemas de processamento de texto e as lógicas temporais consideradas para a ordenação dos eventos.

### 1.3.1 Definição de Evento

Como descrito em vários dicionários, de um modo genérico, os *eventos* [11] são acontecimentos localizáveis espacial e temporalmente, ainda que, por vezes, não seja possível obter essa informação. A definição de evento, sendo bastante genérica, pode ainda ser especificada ou modificada conforme o contexto em que é usada.<sup>3</sup>

Em sistemas de processamento de texto, considera-se que um evento é uma expressão linguística que representa um predicado semântico. Em suma, os verbos plenos (*entrevistar*), nomes predicativos (*entrevista*) e outras expressões ou construções com valor predicativo podem expressar um evento.

<sup>2</sup>A relação *depois* não é considerada explicitamente por ser complementar à relação *antes*.

<sup>3</sup>Num contexto computacional, por exemplo, consideram-se que os eventos são acontecimentos que provocam mudanças de estado no sistema em causa.



Assim, a definição de evento aqui empregue difere um pouco do seu sentido quotidiano, no sentido de abranger também predicados que não exprimem necessariamente a noção de *acontecimentos*. As expressões de medida (*pesar*), estados psicológicos (*irritar*) e predicados causativos (*causar*) são exemplos de eventos, nesta aceção precisa:

*Aquele livro pesa 2 quilos.*

*A Joana irritou-se com o João.*

*O sismo causou o tsunami.*

No entanto, mesmo entre sistemas de processamento de texto, a definição de evento pode variar de sistema para sistema, dependendo da perspectiva e objetivo do projecto. Como descrito ao longo deste trabalho, alguns dos sistemas estudados consideram que a inicialização ou finalização de um evento é igualmente um evento. Por conseguinte, na frase que se segue, é possível identificar-se um (*ler*) ou dois (*acabou e ler*) eventos, conforme a perspectiva adoptada.

*Ele já acabou de ler o livro.*

Para este trabalho, neste último caso, identifica-se apenas um evento (*ler*) e ignora-se o verbo auxiliar (*acabar*). Também não se considera os adjetivos predicativos (1)<sup>4</sup>, as construções preposicionais locativas (2) e os advérbios (3) como eventos:

1. *O Pedro ficou cansado.*
2. *O livro está sobre a mesa.*
3. *O Pedro fez isso depressa.*

Adicionalmente, um evento pode ainda pertencer a uma cadeia de eventos, podendo desencadear ou ser desencadeado por outros eventos. Como exemplo, na frase que se segue, o evento *adormecer* foi provocado pelo evento *sono*.

*Ele adormeceu de sono.*

A partir da interpretação da frase, sendo o evento *sono* o desencadeador do evento *adormecer*, determina-se que *sono* ocorreu em primeiro lugar.

Do mesmo modo, é possível, por vezes, identificar a ordem entre dois ou mais eventos a partir da interpretação da frase. Em alternativa, os eventos podem ser ordenados entre si com base no momento de ocorrência (1), iniciação (2) ou término dos mesmos (3).

1. *O João nasceu a 3 de Março de 1987.*
2. *A 25 de Abril de 1975, deu-se a revolução dos cravos.*

---

<sup>4</sup>Considera-se um adjectivo como predicativo se este exprime um predicado semântico e pode ocorrer auxiliado por um verbo copulativo, v.g. *ser* ou *estar* (e suas variantes).

### 3. *O Pedro obteve o doutoramento em 2004.*

A ordenação de eventos, seguindo uma ou ambas possibilidades, já foi realizada no passado em sistemas dedicados ao processamento de outras línguas. Como parte deste trabalho, precedeu-se à análise do português e à ordenação dos eventos baseadas, sobretudo, na interpretação das relações existentes entre os mesmos.

#### 1.3.2 Processamento de Texto

O processamento de língua natural [32] passa pela análise: da categoria das palavras (análise morfológica); da relação estrutural entre as palavras (análise sintática); do significado das mesmas (análise semântica); da relação do significado com os objectivos comunicativos (análise pragmática). Ao contrário das línguas artificiais, construídas por pessoas para serem mais facilmente manipuláveis por máquinas, as línguas naturais são conhecidas pela presença frequente de ambiguidades, sendo possível encontrá-las a nível sintático, semântico e morfológico.

Entre os tipos de modelos e algoritmos que se podem aplicar no processamento de texto para auxiliar a resolução deste tipo de problemas, destacam-se as máquinas de estados finitos, sistemas baseados em regras formais, lógica, modelos probabilísticos e técnicas de aprendizagem automática.

As *máquinas de estados finitos* são modelos formais constituídos por estados, transições entre estados e valores de entrada, correspondente, neste tipo de sistemas, ao texto a processar. Entre as variações possíveis, destacam-se o uso das cadeias de Markov (Markov Chains) e Modelos Escondidos de Markov (HMM, do ing. Hidden Markov Model) [44]. Tratam-se de extensões de autómatos finitos, definidos por estados e transições entre estados, cujos arcos de transição têm pesos probabilísticos associados.

Entre os *sistemas baseados em regras*, destacam-se o uso de gramáticas regulares, gramáticas livres de contexto e gramáticas aumentadas (*feature-augmented*).

As máquinas de estados finitos e os sistemas baseados em regras constituem as ferramentas principais para trabalhar sobre a análise morfológica e sintática. Para a análise sintática é também comum, como verificado em alguns dos sistemas analisados ao longo deste trabalho, a criação de estruturas, árvores sintáticas, para organização e decisão sobre o texto de entrada. Sobre essas estruturas, recorre-se frequentemente a algoritmos [62] como a Pesquisa em Profundidade Primeiro, Melhor Pesquisa e Procura A\* [20, 30, 39], de modo a tomar as decisões mais correctas sobre o texto.

Os *sistemas baseados em lógica* constituem um outro tipo de modelo, com um papel importante na captura do conhecimento da língua, e utilizam técnicas como a Lógica de Primeira Ordem e Cálculo Predicativo. Ainda que estes modelos possam ser usados atualmente, já não são um foco no processamento de língua natural. Inclusive, nenhum dos sistemas analisados no âmbito deste trabalho recorreu a este tipo de método.

A *aprendizagem automática* foca-se na automatização e aperfeiçoamento das técnicas existentes, podendo ser usada para a criação de autómatos, regras, heurísticas e classificadores, conforme o tipos de modelos aplicados para o processamento de texto.

Adicionalmente, qualquer um dos tipos de modelos referidos (máquinas de estados, sistemas de regras, lógica e aprendizagem automática) pode ainda ser aperfeiçoado com a aplicação de probabilidades, obtendo-se *modelos probabilísticos* como as Cadeias de Markov, o HMM, a Máxima Entropia (MaxEnt) [7] e as gramáticas aumentadas.

Para a avaliação dos sistemas de processamento de texto, utilizam-se geralmente as métricas *precisão*, *abrangência* e *medida-f* e um corpus linguístico [25], que deve representar uma amostra representativa da língua (ou da variedade da língua) para o qual o sistema desenvolvido se destina.

### 1.3.3 Lógicas Temporais

#### Reichenbach

Reichenbach [3, 27, 45] considera os eventos como pontos numa linha temporal, podendo determinado evento ocorrer *antes*, em *simultâneo* ou *depois* de outros eventos. Considera a noção de tempo presente, passado e futuro para a representação dos eventos, mas relativo a pontos de referência: momento de enunciação, instante do evento e instante de referência. O uso destas referências permitem relacionar pares de eventos que ocorrem ambos no passado ou no futuro. No exemplo que se segue, ambos eventos ocorreram no passado, ou seja, os instantes dos eventos precederam o instante de enunciação:

Eu *tinha feito* isto quando ele *chegou*.

No entanto, a frase seguinte tem implícita a ideia que algo se sucedeu ao evento *fazer*, pelo que o instante de referência sucede o instante do evento:

Eu *tinha feito* isto.

Segue-se, na Figura 1.3, diversos exemplos com as respetivas representações, segundo Reichenbach.

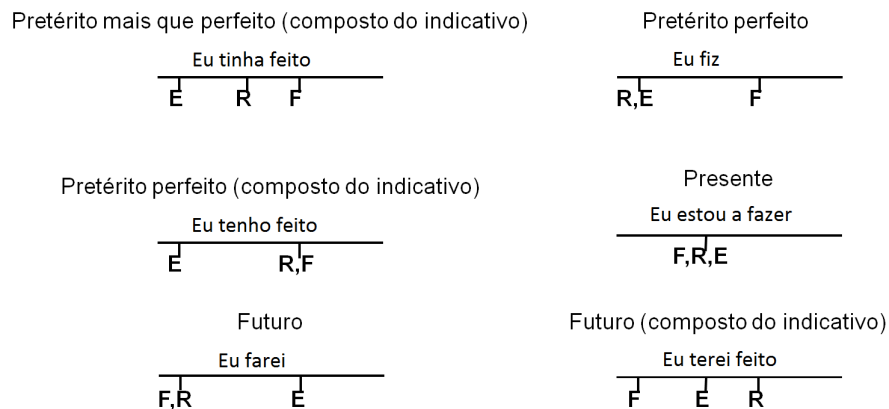


Figura 1.3: Exemplos na representação de Reichenbach, considerando o *instante do evento* (E), *de referência* (R) e *da frase* (F).

#### Interval-based Logic

Interval-based Logic [2] é um tipo de lógica temporal baseado em intervalos e considera as relações de inclusão, sobreposição e interseção entre cada par de eventos. Existem 13 relações diferentes segundo este tipo de lógica,

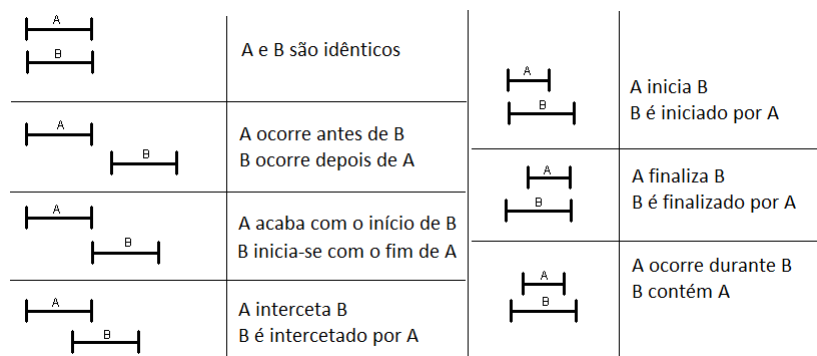


Figura 1.4: Representação das relações de Allen.

frequentemente chamadas de relações de Allen, considerando todas as combinações possíveis, como representado na Figura 1.4.

No entanto, a complexidade computacional para todas estas relações e entre todos os pares de eventos é elevada, pelo que se pode recorrer a simplificações como a adaptada pelo sistema XTM [28]: *antes*, *depois*, *igual*, *inclui*, *inclusão*, *sobreposição* e *sobreposto*.

## 1.4 Contributos

Este trabalho pretende contribuir para o desenvolvimento do sistema STRING, através da construção de um módulo de identificação, relacionamento e ordenação de eventos, permitindo representar essas relações e respectiva ordem num formalismo gráfico, mais adequado à interpretação por um humano. Para tal, definir-se-ão eventos e a respectiva ordenação, construindo-se um corpus anotado que permitirá a avaliação objectiva do sistema.

## 1.5 Estrutura do Documento

Este documento descreve o estado da arte no Capítulo 2, detalhando alguns sistemas que abordaram o problema de identificação, relação e ordenação de eventos. Inicia-se com a introdução de esquemas de anotação utilizados, seguindo da lógica e modelos aplicados por cada sistema e os resultados obtidos.

No Capítulo 3, analisa-se a estrutura gramatical do português e o modo como esta auxilia a identificação dos eventos e a relação e ordenação dos mesmos. Prossegue-se para a estruturação e implementação da solução no Capítulo 4.

No Capítulo 5, descreve-se o corpus criado para a avaliação do módulo e o modo como o mesmo foi avaliado. Apresentam-se igualmente os resultados obtidos e a análise efetuada, descrevendo os problemas encontrados. Finaliza-se o documento, no Capítulo 6, com a conclusão do trabalho.

## Capítulo 2

# Estado da Arte

Segue-se a introdução de alguns esquemas de anotação de eventos, expressões temporais e relações de ordem entre eventos. Estes esquemas foram usados, ou serviram de template, para os sistemas desenvolvidos em anos anteriores, descritos ao longo deste capítulo. Para cada sistema, descreve-se a representação adoptada, os métodos utilizados na resolução do problema de ordenação de eventos e os resultados obtidos nas avaliações realizadas.

Após a descrição e análise dos sistemas existentes, introduz-se a cadeia de processamento STRING [34] (Statistical e Rule Based Natural Language Processing), sistema que foi expandido com o módulo desenvolvido ao longo deste trabalho. A STRING é um sistema bastante modular, permitindo uma adição fácil de funcionalidades, como o módulo desenvolvido, e já extrai bastante informação do texto, tendo um esquema de representação próprio. Entre a informação extraída pela STRING, já se identifica e classifica alguns eventos e expressões temporais, assim como variados elementos gramaticais que auxiliam a resolução do problema abordado neste trabalho.

Conclui-se este capítulo com uma pequena análise da STRING e dos sistemas semelhantes, assim como algumas notas sobre os métodos adoptados pelos mesmos.

### 2.1 Esquemas de Anotação

#### 2.1.1 STAG

STAG (*Sheffield Temporal Annotation Guidelines*) [53] é um esquema de anotação que permite a anotação de expressões temporais, eventos e relações temporais entre eventos.

Utiliza a etiqueta TIMEX para a anotação de expressões temporais e a etiqueta EVENT para a identificação de eventos, relacionando-os através da adição de atributos nas etiquetas, para guardar o tipo de relação e os identificadores das expressões temporais ou eventos envolvidos.

### 2.1.2 TIDES

O esquema de anotação TIDES (*Translingual Information Detection, Extraction, and Summarization*) [22] foi criado com o apoio da DARPA (*Defense Advanced Research Projects Agency*) e com o objetivo de identificar e normalizar expressões temporais, para que estas pudessem futuramente ser usadas para sistemas de pergunta-resposta, tradução automática e de sumarização.

TIDES define um princípio para a anotação de expressões temporais: Se uma expressão é determinável pela interpretação humana, ela deve ser detetada e o valor deve ser determinado a partir de evidências presentes no documento analisado.

As expressões temporais são anotadas com a etiqueta TIMEX2 [22] e identificadas a partir de nomes, como *verão*, *mês* e *Natal*, padrões temporais especiais, como *12/12/2012* e *8:10*, adjetivos, como *recente* e *anual*, advérbios, como *hoje* e *ultimamente*, números, conjunções e preposições.

Segue-se um exemplo de normalização, demonstrando a extração de valores a partir de datas relativas e a relação estabelecida entre elas:

"I'm leaving on vacation *two weeks from next Tuesday*." (Eu vou de férias daqui a *duas semanas a partir da próxima terça-feira*):

I'm leaving on vacation <TIMEX2 VAL="1999-08-03"> two weeks from

<TIMEX2 VAL="1999-07-20">next Tuesday</TIMEX2></TIMEX2>

### 2.1.3 TimeML

O TimeML [40, 51, 52, 54] é uma linguagem de especificação, que permite representar expressões temporais, eventos e as relações entre eventos, combinando e estendendo os esquemas TIDES e STAG.

Os principais problemas abordados são:

- 1- *time stamping* de eventos (identificação dos eventos e ordenação temporal dos mesmos);
- 2- ordenação de eventos entre si;
- 3- raciocínio sobre as expressões temporais relativas (tais como as expressões *último mês* e *duas horas antes*);
- 4- raciocínio sobre a persistência dos eventos (determinação da duração dos eventos e seus resultados).

O esquema de anotação TimeML utiliza as etiquetas EVENT, SIGNAL, TIMEX3, TLINK, SLINK e ALINK para a etiquetagem dos eventos, dos tempos e das relações identificadas.

A etiqueta EVENT é usada para assinalar e classificar os eventos. Estes podem ser expressos a partir de verbos, nominalizações, adjetivos e orações predicativas ou preposicionais <sup>1</sup>, podem ser vistos como pontos ou intervalos no tempo e descrever estados ou circunstâncias.

<sup>1</sup>Os conceitos de orações predicativas e preposicionais não são atualmente considerados pela STRING. As orações predicativas e preposicionais são subclasses de orações subordinativas que, por sua vez, serão consideradas para a ordenação de eventos.

Tipo	Exemplos
Absoluto	Janeiro, 4 de Fevereiro, Natal
Relativo	último mês, próxima semana, 3 dias atrás
Duração	4 horas, 5 meses

Tabela 2.1: Tipos de expressões temporais

As orações predicativas são orações subordinadas que exercem a função de predicativo do sujeito da oração principal, como no exemplo que se segue, utilizando a conjunção *que*: *O certo é que o João vai receber o prêmio.*

As orações preposicionais são orações que complementam o sentido da oração subordinativa ao desempenhar uma função de complemento preposicional, como na frase: *O João tem a certeza de que vai receber o prêmio.*

Os eventos são classificados quanto ao tempo, à classe ou ao aspeto. Relativo ao tempo, o evento ocorre no *passado, presente, futuro* ou *desconhecido*, quanto à classe, o evento pode representar uma ação, um estado ou uma percepção, entre outros, e, quanto ao aspeto, pode ser *perfeito, progressivo* (que se prolonga no tempo), *ambos* ou *nenhum*.

A etiqueta *SIGNAL* é usada para anotar secções do texto que indicam como os elementos temporais se relacionam, tais como indicadores de relações e conectores temporais. Alguns exemplos das expressões etiquetadas com *SIGNAL* são *durante, enquanto* e *se*. Segue-se um exemplo de uma frase com um indicador temporal assinalado pela etiqueta *SIGNAL*:

“Two days *before* the attack.” (dois dias *antes* do ataque.):

Two days <SIGNAL sid="1">before</SIGNAL> the attack.

A etiquetagem das entidades temporais é feita com a etiqueta *TIMEX3*, extensão do *TIMEX2*, detetando três tipos de expressões temporais, como exemplificado na Tabela 2.1: expressões absolutas, relativas e durações.

Segue-se um exemplo de uma frase com a expressão temporal identificada com a etiqueta *TIMEX3*:

“In *November 22, 2004*.” (a 22 de Novembro de 2012):

<TIMEX3 tid="1" type="DATE" value="2012-11-22">November 22, 2004

</TIMEX3>.

As relações temporais são expressas pelas etiquetas *LINK*: *TLINK* (Temporal *LINK*), *SLINK* (Subordination *LINK*) e *ALINK* (Aspectual *LINK*).

A etiqueta *TLINK* é usada para assinalar relações temporais entre eventos ou entre eventos e o tempo, com base nas 13 relações de Allen (ver Secção 1.3.3 deste documento), *SLINK* é usado para assinalar relações entre eventos e entre eventos e sinais. Por fim, a etiqueta *ALINK* é usada para representar relações aspetuais entre os eventos. As relações aspetuais consideradas são do tipo *inicialização, culminação, terminação* ou *continuação*, como exemplificado na Tabela 2.2.

Estas etiquetas (*TLINK, SLINK* e *ALINK*) contêm alguns atributos em comum (*eventInstanceID* e *signalID*) que representam os identificadores respetivos do evento e sinal encontrado.

Tipo	Exemplo
Inicialização	"John started to read." (O João começou a ler.)
Culminação	"John finished assembling the table." (O João acabou de montar a mesa.)
Terminação	"John stopped talking." (O João parou de falar.)
Continuação	"John kept talking." (O João continuou a falar.)

Tabela 2.2: Relações aspetuais

A etiqueta TLINK tem, como atributos, os tempos ou eventos que se relacionam, o id do sinal usado para identificar a relação, o tipo de relação e a expressão temporal referente à magnitude, se existir. Considere-se a seguinte frase: *John left two days before the attack.* (O João saiu dois dias antes do ataque.).

Identificam-se no exemplo dois eventos, *left* (saiu) e *attack* (ataque), o sinal *before* (antes de) e a expressão temporal *two days* (dois dias). Atribuíem-se os identificadores *ei1* e *ei2* aos eventos, *s1* ao sinal e *t1* à expressão temporal encontrada. Deste modo, a anotação resultante é a que se segue:

```
<TLINK eventInstanceID="ei1" signalID="s1" relatedToEvent="ei2"
relType="BEFORE" magnitude="t1"/>
```

A etiqueta SLINK tem os atributos referentes à instância do evento principal, ao sinal, ao evento subordinado e ao tipo de relação estabelecida. Considere-se a frase: *Bill wants to teach on Monday.* (O Bill quer ensinar à segunda-feira.)

Nesta frase, identificam-se dois eventos, o evento *querer* e *ensinar*, e o sinal *to* (que tem a função, na língua inglesa, de ligar dois verbos). A atribuição dos identificadores é feita à semelhança da etiqueta TLINK. Segue-se a anotação resultante:

```
<SLINK eventInstanceID="ei1" signalID="s1" subordinatedEvent="e2"
relType="IS_INCLUDED"/>
```

A etiqueta ALINK tem, como atributos, os eventos que se relacionam, o sinal e o tipo de relação. Considere-se a frase: "The boat begun to sink." (O barco começou a afundar-se.).

No exemplo, foram identificados dois eventos, identificados através dos verbos *begun* (começou) e *sink* (afundar)<sup>2</sup>, um sinal *to* (a) e uma relação de inicialização entre os eventos presentes.

"The boat begun to sink." (O barco começou a afundar-se.):

```
<ALINK eventInstanceID="ei1" signalID="s1"
relatedToEvent="e2" relType="INITIATES"/>
```

<sup>2</sup>Na STRING considera-se a existência de apenas um evento: o evento *afundar*. O verbo auxiliar *começou* é considerado um verbo aspetual devido às suas propriedades aspetuais.



## 2.2 Sistemas de Ordenação Temporal de Eventos

### 2.2.1 Terseo

O sistema TERSEO (*Temporal Expression Resolution System Applied to Event Ordering*) (2005) [49, 50] é uma aplicação web que permite a ordenação de eventos, baseada em regras e utilizando o esquema de anotação TIMEX2. A arquitetura do sistema de ordenação encontra-se representado na Figura 2.1.

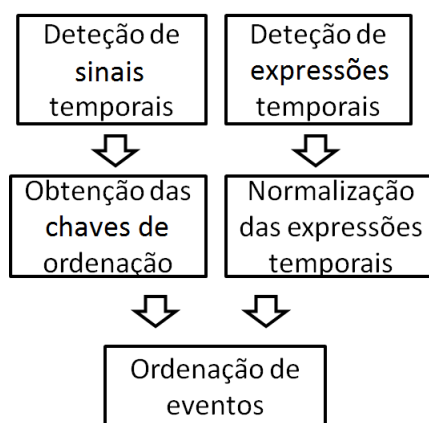


Figura 2.1: Estrutura do sistema TERSEO

Este sistema permite a ordenação dos eventos em três fases. Na primeira fase, são detetados os sinais (ex: *antes; durante; enquanto*), que ajudam a estabelecer a ordem entre eventos, e as expressões temporais. As expressões temporais podem ser explícitas, expressando datas absolutas (ex: *3 de Março*) e variáveis (ex: *Páscoa*), ou referências relativas (ex: *ontem*).

Na segunda fase, as expressões temporais detetadas passam por um processo de resolução anafórica e são etiquetadas. Em paralelo, os sinais temporais são usados para obter os tipos de ordenação a serem aplicados entre os eventos, na próxima fase.

Na terceira e última fase, o sistema usa as ordenações e as expressões temporais, após a normalização e etiquetação, para a ordenação dos eventos. Apenas são identificados os eventos presentes em frases que contenham expressões temporais, ou que se relacionam com outras frases através de um ou mais sinais temporais. Após a ordenação, é atribuído um número a cada evento, pela ordem temporal de ocorrência, podendo existir vários eventos com o mesmo número, se estes forem temporalmente concomitantes.

O sistema foi testado para as línguas espanhola e inglesa. Para o espanhol, foi utilizado um corpus para treino, anotado por dois anotadores diferentes, e um outro corpus para a avaliação. Para o teste em inglês utilizou-se apenas um corpus de teste, com as regras obtidas por métodos de aprendizagem automática. Nestas condições, obtiveram-se os resultados reportados na Tabela 2.3.

	corpus de treino (espanhol)	corpus de teste (espanhol)	corpus de teste (inglês)
Precisão	84%	88%	77%
Abrangência	71%	69%	62%
Medida-F	77%	77%	68%

Tabela 2.3: Resultados do sistema TERSEO

### 2.2.2 XTM

O processador temporal XTM (*XIP Temporal Module*) (2007) [28] é uma extensão do XIP (*XEROX Incremental Parsing*) [1]. XIP é um sistema que se baseia em regras e extrai relações gramaticais e papéis temáticos como ligações de dependência, dividindo-se em três fases:

- (i) fase de tokenização, análise morfológica e etiquetagem;
- (ii) segmentação e reconhecimento de entidades mencionadas;
- (iii) análise sintática profunda do texto.

A análise temporal, realizada pelo XTM, é feita em três fases: reconhecimento e interpretação de expressões temporais; associação de expressões temporais aos eventos e ordenação dos eventos presentes na mesma frase; ordenação de eventos no documento.

O processamento temporal a nível da frase é realizado com as outras tarefas de análise linguística, pelo que a associação entre expressões temporais e eventos é considerada um caso particular da atribuição de papéis temáticos a predicados. O sistema consegue detetar se os predicados presentes na frase estão temporalmente relacionados e que tipo de relações existe entre eles. Permite ainda calcular a distância temporal entre eventos, quando explícita, e inferir as relações dependentes da data de criação do documento.

As relações temporais são baseadas nas relações de Allen [2], considerando-se as relações *depois*, *antes*, *durante*, *inclui*, *sobrepo*, *é sobreposto* e *é igual*, e os eventos são identificados a partir de nomes e verbos, considerando-se a informação morfológica e aspetual. Se não for possível determinar especificamente o tipo de relação temporal estabelecida, utiliza-se uma representação própria para o efeito. Como exemplo, considere-se a perífrase verbal: *has been taken* (tem sido levado). Segundo os autores, o evento presente na frase pode ser considerado presente ou passado, pelo que é exprimido do seguinte modo: PRES-OR-PAST(taken).

De modo a demonstrar o tipo de relações estabelecidas entre eventos e expressões temporais, considere-se a frase: *People began gathering in Abuja Tuesday for the two day rally*. (As pessoas começaram a reunir-se em Abuja, na terça, para dois dias de rali.)

*People began gathering in Abuja Tuesday for the two day rally*. (As pessoas começaram a reunir-se em Abuja, na terça, para o rali de dois dias.)

Identificam-se três eventos, *began*(começou), *gathering*(reunir) e *rally*(rali), a referência temporal *Tuesday* (terça-feira) e a duração *two days* (dois dias), obtendo-se as seguintes associações temporais:

TEMP(began, Tuesday)

	Relação	XTM	melhor resultado
Tarefa 1	Precisão	53%	62%
	Abrangência	25%	62%
	Medida-F	34%	62%
Tarefa 2	Precisão	78%	80%
	Abrangência	57%	80%
	Medida-F	66%	80%
Tarefa 3	Precisão	42%	55%
	Abrangência	42%	55%
	Medida-F	42%	55%

Tabela 2.4: Resultados da avaliação restrita na relação temporal de eventos

	Relação	XTM	melhor resultado
Tarefa 1	Precisão	63%	64%
	Abrangência	30%	64%
	Medida-F	41%	64%
Tarefa 2	Precisão	84%	84%
	Abrangência	62%	81%
	Medida-F	71%	81%
Tarefa 3	Precisão	58%	64%
	Abrangência	58%	64%
	Medida-F	58%	64%

Tabela 2.5: Resultados da avaliação relaxada na relação temporal de eventos

#### TEMP(rally, two day)

O sistema foi avaliado no TempEval-1 [58], considerando uma avaliação restrita (Tabela 2.4) e relaxada (Tabela 2.5) para cada uma das três tarefas de relações temporais da campanha [57]:

1. Determinar as relações temporais entre eventos e expressões temporais na mesma frase;
2. Determinar a relação de cada evento em relação à data de criação do documento;
3. Determinar as relações temporais entre eventos de frases distintas.

#### 2.2.3 Chambers et al.

Chambers et al. (2008) [12] realizaram um trabalho de relacionamento e de ordenação temporal de eventos com uma abordagem de aprendizagem automática sobre o TimeBank Corpus [8, 41], considerando uma análise local e global dos eventos identificados

Relação	Classificação local			Classificação local, global e temporal		
	antes	depois	desconhecido	antes	depois	desconhecido
Precisão	61,4%	57,6%	53,0%	63,7%	60,3%	52,0%
Abrangência	55,4%	53,1%	62,8%	57,1%	54,3%	62,9%
Medida-F	58,2%	55,3%	57,5%	60,2%	57,1%	56,9%

Tabela 2.6: Resultados dos testes

O Timebank Corpus é um corpus constituído por 186 artigos com os eventos e expressões temporais etiquetadas, com base no modelo de anotação do TimeML [40, 51, 52, 54]. Cada evento está etiquetado com a informação temporal existente sobre ele, como tempo, modalidade e aspeto. As relações usadas entre eventos são baseadas nas relações de Allen [2].

A classificação local é baseada num outro trabalho anterior de Chambers et al. [13], que permite a ordenação de pares de eventos a partir de métodos de aprendizagem supervisionados, como *support vector machines* [10, 23, 61].

A abordagem global foi introduzida para evitar a ocorrência de paradoxos e contradições, causadas pelas decisões locais. Por exemplo, o evento 1 pode ser classificado como anterior do evento 2o qual, por sua vez, que precede o evento 3, apesar de este ser anterior ao evento 1. Considerou-se inicialmente, para este modelo, uma abordagem gananciosa (*greedy*) [62] mas, após testes, obtiveram-se melhores resultados com o uso da Programação Linear Inteira (*Integer Linear Programming*) [48].

Em conjunto com a abordagem local e global, considerou-se os dados temporais associados aos eventos para criar uma maior conexão entre os mesmos. Os testes foram realizados considerando apenas 3 relações (*antes*, *depois* e *desconhecido*), com 13% de relações desconhecidas entre os eventos anotados. Os resultados são reportados na Tabela 2.6, podendo-se verificar que, de um modo geral, se obteve resultados melhores ao se utilizar a abordagem local e global e a informação temporal obtida.

#### 2.2.4 TRIPS e TRIOS

O analisador TRIPS (2010) [56] utiliza uma gramática livre de contexto e produz formas lógicas a partir de texto, usando uma gramática da língua inglesa e um dicionário semântico. Para palavras que não são conhecidas do dicionário do TRIPS, o sistema utiliza a WordNet [19, 29, 38]. É usado uma representação em grafos do tipo *packed-forest* [31], construindo os nós num sentido de baixo para cima, usando uma estratégia de procura semelhante à  $A^*$  [30], baseada em estatísticas do pré-processamento e em regras e pesos lexicais. Após a obtenção do grafo, usa-se um algoritmo de programação dinâmica para procurar a sequência de formas lógicas com menor custo.

O TRIOS (*TRIPS Temporal Reasoning System*) é um sistema que utiliza o analisador TRIPS para extrair as expressões temporais, os eventos, os respetivos traços e relações presentes no texto. Após obter-se as formas lógicas com o TRIPS, aplicam-se padrões de extração previa e manualmente codificados. Enquanto é realizada a extração, os eventos são filtrados e as relações temporais são identificadas usando classificadores *Markov Logic Networks* [47] e *Conditional Random Field* [60].

	TRIPS	TRIOS	melhor resultado
Precisão	55%	80%	81%
Abrangência	88%	74%	88%
Medida-F	67%	84%	84%

Tabela 2.7: Resultados obtidos na extração de eventos

	Relação	TRIPS	TRIOS	melhor resultado
Tarefa 1	Precisão	63%	65%	65%
	Abrangência	52%	52%	52%
	Medida-F	57%	58%	
Tarefa 2	Precisão	76%	79%	82%
	Abrangência	69%	67%	69%
	Medida-F	72%	73%	
Tarefa 3	Precisão	58%	56%	58%
	Abrangência	50%	42%	50%
	Medida-F	54%	48%	
Tarefa 4	Precisão	59%	60%	66%
	Abrangência	54%	46%	54%
	Medida-F	56%	52%	

Tabela 2.8: Resultados obtidos na relação temporal de eventos

O esquema de anotação usado é uma extensão personalizada do TimeML [40, 51, 52, 54], tendo sido proposto uma extensão do corpus TIMEBANK [8, 41], TRIOS-TIMEBANK [55], com a adição de atributos e de uma nova etiqueta, RLINK, que permite relacionar outros tipos de elementos com os eventos.

Segundo os resultados obtidos no TempEval-2 [42, 43], na avaliação da extração de eventos, Tabela 2.7, o sistema TRIOS obteve a maior abrangência (88%), enquanto o sistema TRIPS conseguiu obter a segunda maior precisão (80% em 81%).

Nas tarefas de relações temporais, apesar deste sistema ter utilizado a extração automática em vez de corpus anotados manualmente, conseguiu os resultados da Tabela 2.8, ultrapassando a abrangência e precisão de alguns competidores. A avaliação da relação entre eventos subdividiu-se em quatro tarefas:

1. Determinar as relações temporais entre eventos e expressões temporais na mesma frase;
2. Determinar a relação de cada evento em relação à data de criação do documento;
3. Determinar as relações temporais entre eventos de frases distintas;
4. Determinar as relações pares entre eventos, onde o primeiro é considerado sintaticamente subordinante em relação ao segundo.

	NCSU-INDI	NCSU-JOINT	melhor resultado
Tarefa 1	63%	62%	65%
Tarefa 2	68%	21%	82%
Tarefa 3	48%	51%	58%
Tarefa 4	66%	25%	66%

Tabela 2.9: Resultados obtidos na relação temporal de eventos

### 2.2.5 NCSU-INDI e NCSU-JOINT

Ha et al. (2010) [26] abordaram o problema de relações temporais como a junção de quatro tarefas:

- (i) previsão de relações temporais entre eventos e expressões temporais presentes na mesma frase;
- (ii) relação dos eventos com o tempo de criação do documento;
- (iii) relação entre os eventos principais de frases vizinhas;
- (iv) relações entre eventos principais e sintaticamente dependentes.

Os sistemas NCSU-INDI e NCSU-JOINT [26] usam uma abordagem de aprendizagem automática com *Markov Logic Networks* [47]. O sistema NCSU-INDI consiste, no entanto, em quatro classificadores distintos, um para cada tarefa, ao contrário do NCSU-JOINT, que considera a composição das tarefas.

Para a aprendizagem automática, obteve-se traços a partir do corpus, dos analisadores sintáticos e das relações semânticas lexicais extraídas de duas bases de dados externas: a WordNet [19, 29, 38]; VerbOcean [14], grafo de relações semânticas de verbos onde se estabelece associações e respectivos pesos.

Ambos sistemas foram avaliados no Temp-Eval 2010 [42, 43] em termos de precisão, obtendo-se os valores da Tabela 2.9

### 2.2.6 Maršić

O sistema de Maršić (2011) [35], WVALI, permite a identificação, normalização e ordenação de eventos e de expressões temporais, utilizando a base de dados lexical WordNet 2.0 [19, 29, 38] e o esquema de anotação TIMEX2 [22], adaptado de modo a ficar em conformidade com o TimeML [40, 51, 52, 54].

A tarefa de identificação das expressões temporais divide-se em três módulos: um módulo de identificação baseado em regras, um módulo para verificar a correção sintática e um módulo para desambiguação.

A classificação de eventos é feita com o auxílio do WordNet, atribuindo a cada verbo a sua classe mais relevante. A normalização dos eventos segue a abordagem do TimeML, usando a etiqueta EVENT.

Para a identificação das relações entre eventos, o sistema considera a informação existente nos advérbios de tempo, nas orações subordinadas e o tempo e aspeto dos verbos. O problema de ordenação de eventos é subdividido em dois: a identificação das relações dentro da mesma frase e a ordenação temporal entre frases diferentes.

	Relação	WVALI	melhor resultado
Tarefa 1	Precisão	62%	62%
	Abrangência	62%	62%
	Medida-F	62%	62%
Tarefa 2	Precisão	80%	80%
	Abrangência	80%	80%
	Medida-F	80%	80%
Tarefa 3	Precisão	54%	55%
	Abrangência	54%	55%
	Medida-F	54%	55%

Tabela 2.10: Resultados da avaliação restrita na relação temporal de eventos

	Relação	WVALI	melhor resultado
Tarefa 1	Precisão	64%	64%
	Abrangência	64%	64%
	Medida-F	64%	64%
Tarefa 2	Precisão	81%	84%
	Abrangência	81%	81%
	Medida-F	81%	81%
Tarefa 3	Precisão	64%	64%
	Abrangência	64%	64%
	Medida-F	64%	64%

Tabela 2.11: Resultados da avaliação relaxada na relação temporal de eventos

As relações temporais existentes dentro das frases são identificadas usando árvores sintáticas ao nível da frase e propagação de baixo para cima das relações entre os constituintes. Entre frases diferentes são usados apenas os eventos principais de ambas para estabelecer as relações. Se não for possível haver consenso na ordenação a atribuir, o sistema usa os dados estatísticos do TimeBank [8, 41]. Quando é detetado um evento condicionado pelo tempo de criação do documento em análise, ou seja, uma referência do tipo enunciação, as relações temporais relacionadas com esse evento são analisadas e propagadas no sentido de cima para baixo, pela árvore sintática, se necessário.

Para os testes, Maršić utilizou os corpus TERN 2004 [21], anotado com o TIMEX2, e TimeBank, com o modelo de anotação TimeML.

O sistema também participou na TempEval-1 [58], obtendo os melhores resultados em quase todas as tarefas de relações temporais da campanha [57], tanto na avaliação restrita, Tabela 2.10, como relaxada, Tabela 2.11. Na tarefa 1, relacionamento de eventos dentro da mesma frase, obteve a maior precisão, abrangência e medida-f. Nas outras duas tarefas, relação com a data de criação do documento e relação de eventos entre frases distintas, o sistema conseguiu superar ou manter-se próximo dos resultados máximos obtidos.

## 2.3 Cadeia de processamento STRING

A STRING (*Statistical and Rule-Based Natural Language Processing*) [34] é uma cadeia de processamento de língua natural, desenvolvido para o idioma português pelo L<sup>2</sup>F, baseada em regras e em aprendizagem automática. A STRING tem uma estrutura modular que inclui os módulos LexMan (*Lexical Morphological analyzer*) [59], RuDriCo2 (*Rule Driven Converter*) [17, 18], MARv4 (*Morphossyntactic Ambiguity Resolver*) [16, 46] e XIP (*XEROX Incremental Parsing*) [1]. A arquitetura da STRING [34] está representada na Figura 2.2.

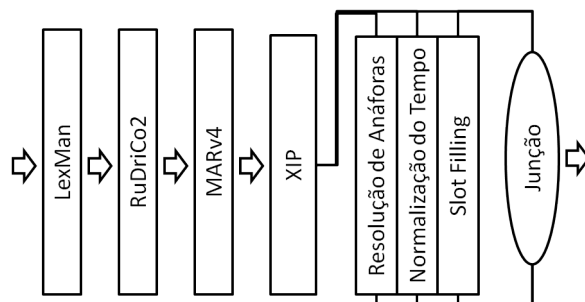


Figura 2.2: Representação da arquitetura da cadeia STRING.

Uma vez que é sobre a estrutura da STRING que o nosso trabalho se desenvolverá, apresentamo-la com mais pormenor nas secções seguintes.

### 2.3.1 Arquitetura

O primeiro módulo, o LexMan [59], é responsável pela segmentação e a etiquetagem morfossintática. O texto é segmentado e é realizada a identificação prévia de certos símbolos, tais como “<”, “>”, “#”, “\$”, “\*”, “+” e “%”, sinais de pontuação, endereços de e-mail, endereços de internet, números cardinais, ordinais ou romanos, abreviaturas, entre outros. Após este processo, é feita a análise e a etiquetagem morfossintática, identificando as classes gramaticais (nome, verbo, adjetivo, pronome, artigo, advérbio, preposição, conjunção, numeral, interjeição, pontuação e símbolo) e preenchendo os campos dos respetivos valores gramaticais (categoria, subcategoria, modo, tempo, pessoa, número, género, grau, caso, formação, sintático e semântico). Por fim, o LexMan faz a segmentação do texto em frases, considerando os sinais de pontuação “.”, “!” e “?” como fronteiras de frase.

À saída do LexMan, o texto é processado pelo módulo RuDriCo2 [17, 18], responsável por desfazer as contrações ou tratar como um único *token* as palavras compostas. Procede ainda à desambiguação morfológica baseada em regras, alterando a segmentação de texto quando necessário.

Após este processo de desambiguação baseada em regras, o módulo MARv4 [16, 46] efetua uma desambiguação estatística, recorrendo ao modelo Hidden Markov Model [44] e usando o algoritmo de Viterbi [44] para seleccionar as melhores sequências de etiquetas morfossintáticas. O modelo de língua é baseado em trigramas (modelos de segunda ordem), que codificam a informação referente a categorias, e unigramas, que codificam a informação lexical.

Depois do MARv4, o XIP [1] encarrega-se de executar uma análise sintática ao texto processado anteriormente,



adicionando informação lexical, sintática e semântica. Aplica regras de desambiguação morfossintática e gramáticas locais, calcula os *chunks* (constituintes elementares) e as respectivas dependências (sujeito, complemento direto, modificador, etc.). Os eventos atualmente identificados pela STRING são detetados e classificados neste módulo.

Por fim, é feita a resolução de anáforas, a normalização de expressões temporais e a identificação e associação de informação adicional às entidades mencionadas, nos respetivos módulos.

### 2.3.2 Classificação de Entidades Temporais

As expressões temporais classificadas pela STRING abrangem as expressões de referência *absoluta* (completa ou incompleta), de referência *relativa*, e as expressões de *duração* e de *frequência* [6, 37]. As expressões *genéricas* como a expressão *Verão*, quando ocorre em frases como *Eu gosto do Verão*, são igualmente consideradas, dado que podem ser usadas como referência para outras expressões temporais, como exemplificado: *Eu gosto do Verão. É a altura do ano em que eu consigo trabalhar melhor.*

Nesta segunda frase, a expressão anafórica *a altura do ano* refere-se à ocorrência temporal genérica *Verão*.

Uma referência *absoluta* não depende de outros elementos para que se consiga obter o seu valor. As referências *relativas*, por sua vez, podem ser de natureza *textual*, se o seu valor depender de uma outra expressão temporal presente no texto, ou de *enunciação*, se dependerem, para a sua interpretação, do momento de enunciação (em última instância, da data de criação do documento).

Exemplo de uma referência absoluta:

"Estarei de regresso *no segundo dia de Março de 2012*."

Exemplo de uma referência à enunciação:

"Portugal é *atualmente* um país desenvolvido, com um Índice de Desenvolvimento Humano (IDH) considerado como muito elevado."

Exemplo de uma referência de textual:

"A *25 de Novembro de 1975* diversos sectores da esquerda radical (essencialmente pára-quedaistas e polícia militar na Região Militar de Lisboa), provocados pelas notícias, levam a cabo uma tentativa de golpe de estado, que no entanto não tem nenhuma liderança clara. O Grupo dos Nove reage pondo em prática um plano militar de resposta, liderado por António Ramalho Eanes. Este triunfa e *no ano seguinte* consolida-se a democracia. "

A mesma referência pode ainda ser expressa por diferentes expressões temporais, como exemplificado nas frases que se seguem:

*Estarei de regresso no segundo dia de Março de 2012.*

*Estarei de regresso no dia 2 do terceiro mês, em 2012.*

*Estarei de regresso na primeira sexta-feira de Março de 2012.*

Tipo	Expressão Temporal	Exemplo
Referência Absoluta	<i>a 20 de Janeiro</i> <i>a 20/01/1990</i> <i>entre o dia 2 e 7 de Janeiro</i>	A Rita nasceu <i>a 20 de Janeiro</i> . A Rita nasceu <i>a 20/01/1990</i> . Acampamos <i>entre o dia 2 e 7 de Janeiro</i> .
Referência Relativa	<i>ontem</i>	Joguei futebol <i>ontem</i> .
Duração	<i>3 horas</i> <i>entre uma a duas horas</i>	O exame durou <i>3 horas</i> . O filme dura <i>entre uma a duas horas</i> .
Frequência	<i>todos os meses</i>	O Diogo faz desporto <i>todos os meses</i> .
Referência Genérica	<i>o Verão</i>	Adoro <i>o Verão</i> .

Tabela 2.12: Exemplos de expressões temporais identificadas pela STRING

*Estarei de regresso a 02-03-12.*

Seguem-se, na Tabela 2.12, exemplos dos diferentes tipos de expressões temporais consideradas.

Considerando ainda o exemplo de referência textual, a expressão temporal *no ano seguinte* não é equivalente a *25 de Novembro de 1976*, pois apresenta um grau de imprecisão considerável. Esta situação é igualmente considerado pela cadeia STRING, sendo assinaladas tanto a existência de um certo grau de vagueza como a granularidade do intervalo de tempo considerado.

Existe ainda um outro tipo de imprecisão identificado, assinalado através de elementos lexicais como o advérbio *por volta de* na frase que se segue:

*Estarei de regresso por volta do dia 15.*

Outros elementos presentes nas frases, por sua vez, alteram o sentido das expressões temporais, como as preposições *antes* e *depois*, exemplificados nas frases que se seguem:

*Estarei de regresso antes do dia 15.*

*Estarei de regresso depois do dia 15.*

### 2.3.3 Normalização de Entidades Temporais

Para normalizar as expressões temporais absolutas, que referenciam datas de uma forma precisa, usa-se o atributo VAL\_NORM, que segue a seguinte estrutura:

VAL\_NORM="«era»M«milénio»«século»«década»D«ano» «mês»«dia»T«hora»«minuto»«segundos»  
«milisegundo»E«estação»LM«limite-aberto»"

Seguem-se, na Tabela 2.13, alguns exemplos do uso do atributo VAL\_NORM.

O atributo TEMP\_REF é usado para sinalizar o tipo de referência da expressão temporal em análise, podendo tomar os valores "ENUNCIACAO", "TEXTUAL" ou "ABSOLUTO", conforme, respetivamente, se a expressão faz referência ao momento de enunciação (ou data de documento), se se refere a outra expressão temporal presente

Expressão Temporal	Normalização
a 2 de Março de 2012	VAL_NORM="+CM----D20120302T-----E-LM- "
na primavera	VAL_NORM="+CM----D-----T-----EPRLM- "
antes da primavera	VAL_NORM="+CM----D-----T-----EPRLMA "
depois do 1º milénio	VAL_NORM="+CM01---D-----T-----E-LMP "
na 1ª década	VAL_NORM="+CM---01D-----T-----E-LM- "
dia 3 às 16:15	VAL_NORM="+CM----D----03T1615---E-LM- "
em 13 a.C.	VAL_NORM="-CM----D0013---T-----E-LM- "

Tabela 2.13: Exemplos do uso do atributo VAL\_NORM

Expressão Temporal	Conversão	Representação
1 milénio	1000 anos	VAL_DELTA="A1000D0H0M0S0M0", UMED="milénio"
1 década	10 anos	VAL_DELTA="A10D0H0M0S0M0", UMED="década"
meio ano	183 dias	VAL_DELTA="A0D183H0M0S0M0", UMED="ano"
1 mês	30 dias	VAL_DELTA="A0D30H0M0S0M0", UMED="mês"
1 semana	7 dias	VAL_DELTA="A0D7H0M0S0M0", UMED="semana"
meio dia	12 horas	VAL_DELTA="A0D0H12M0S0M0", UMED="dia"

Tabela 2.14: Conversão entre unidades temporais

no texto, ou se se trata de uma referência temporal absoluta, que pode ser normalizada diretamente a partir da expressão.

O atributo VAL\_DELTA é usado na normalização de datas relativas, expressões de duração e de frequência, usando a seguinte notação:

VAL\_DELTA="A«ano»D«dia»H«hora»M«minuto»S«segundo»M«milissegundo»"

Os meses, as décadas e os séculos são omitidos porque o atributo é flexível na representação e permite que se faça conversões entre unidades de tempo com diferentes níveis de granularidade. Por exemplo, *uma semana* é representado como *sete dias*, como se pode verificar nos exemplos da Tabela 2.14. Usa-se um segundo atributo, UMED, para guardar a granularidade da expressão original, como também é exemplificado na tabela, de modo a não se perder informação.

O atributo SENTIDO permite especificar a orientação do tempo do evento relativamente ao momento referido pela expressão temporal encontrada, ou seja, se o evento aconteceu *antes*, *depois*, *durante*, *a terminar* ou *a começar* no momento especificado pela expressão temporal.

Quando uma expressão temporal representa um momento impreciso, tal como *por volta das 3 horas* ou *perto de Setembro*, este facto é assinalado com o atributo FUZZY.

### 2.3.4 Eventos que são atualmente normalizados

Os eventos atualmente detetados e classificados pela STRING [4, 33] referem-se a acontecimentos e relações relevantes para a caracterização de entidades mencionadas, em particular as que envolvem pessoas, organizações, lugares e expressões temporais. Estes eventos são organizados em diferentes categorias.

Assim, na categoria LIFETIME, capturam-se os eventos que assinalam os momentos considerados mais marcantes ou relevantes na vida de uma pessoa, tais como as datas de *nascimento e morte*, as relações de *parentesco*, o local de *residência* ou a *formação académica*. Na categoria BUSINESS representam-se as relações profissionais entre entidades mencionadas, geralmente entre pessoas e organizações. Sob a classe LOCATION, classificam-se os eventos que identificam deslocações ou permanência temporária em determinados locais ou edifícios. A categoria PUBLIC-EVENT procura capturar eventos que envolvem a realização de eventos sociais, de natureza pública, bem como a comemoração de efemérides. Por fim, a categoria CRIME captura os eventos relacionados com crime ou administração da justiça.

A classificação de um evento envolve a identificação dos intervenientes, das condições em que o mesmo ocorre, das datas de início e fim, da localização e da duração, quando presentes ou relevantes para o tipo de evento considerado. Por outras palavras, identifica-se o predicado semântico e atribuem-se os papéis semânticos aos seus argumentos. Considere-se a frase:

*A Maria visitou a Alemanha.*

O evento identificado no exemplo, através do verbo *visitar*, representa uma deslocação, sendo classificado como LOCATION. De modo a especificar que o evento representa uma deslocação breve, atribui-se ainda uma palavra-chave (*keyword*): *local\_visit*. Na representação do exemplo que se segue, EVENT\_LEX associa a palavra-chave ao evento, EVENT\_LOCATION identifica o local visitado, EVENT\_PARTICIPANT indica o interveniente do evento e a relação EVENT\_PLACE associa o edifício ou local visitado.

```
EVENT_LEX(visitou,local_visit)
EVENT_LOCATION(visitou)
EVENT_PARTICIPANT(visitou,Maria)
EVENT_PLACE(visitou,Alemanha)
```

A Tabela 2.15 descreve os tipos de eventos já considerados para normalização. Estes eventos são normalizados deste modo devido à sua importância no sistema de extração e recuperação de informação que a STRING suporta (projeto OObian<sup>3</sup>).

### 2.3.5 Dependências Auxiliares

A STRING extrai ainda algumas dependências com base na análise sintática das frases e dos seus constituintes.

Relativo à análise do predicado da frase, extraem-se as dependências do predicado com: (i) o sujeito; (ii) o complemento indirecto; (iii) o complemento directo; (iv) o predicativo do sujeito e (v) outros modificadores de discurso, expressas do seguinte modo:

---

<sup>3</sup>O projeto OObian é uma plataforma de pesquisa e navegação de conteúdo em textos não estruturados. É um projeto desenvolvido no âmbito do programa QREN e feito em colaboração com a empresa MAISIS. Para mais informações, consultar o website oficial: <http://www.oobian.com/home/overview>

Tipo de evento	Palavras-chave	Exemplos
LIFETIME	nascimento morte idade parentesco residência educação	A Joana nasceu em Fevereiro. O André morreu em Julho. O Rui tem 20 anos. O Rui é irmão da Júlia. A Maria reside na Alemanha. O José estudou no Instituto Superior Técnico.
BUSINESS	trabalho fundação propriedade cliente afiliação	A Joana é rececionista de hotel. Bill Gates fundou a Microsoft. O Rui é acionista da Makro. O Rui faz sempre compras no Pingo Doce. A Maria é simpatizante da Worten.
LOCATION	local_stat local_visita	A Maria já não está no Porto. A Maria visitou a Alemanha.
PUBLIC-EVENT	público efémeride	A Joana foi ao batizado do Marcos. Celebrou-se ontem o 100º aniversário da Implementação da República.
CRIME	homicídio roubo	Dois homens foram mortos num tiroteio. Um homem de 25 anos roubou dois fios de ouro.

Tabela 2.15: Tipos de eventos, palavras-chave e respectivos exemplos

- (i) CDIR («predicado», «complemento direto»)
- (ii) CINDIR («predicado», «complemento indireto»)
- (iii) SUBJ («predicado», «sujeito»)
- (iv) PREDSUBJ («predicado», «predicativo do sujeito»)
- (v) MOD («predicado», «modificador»)

*O João ofereceu hoje um livro à Ana.*

- CDIR (ofereceu, livro)
- CINDIR (ofereceu, Ana)
- SUBJ (ofereceu, João)
- MOD (ofereceu, hoje)

*O Eduardo ficou encantado com essa notícia.*

- PREDSUBJ (ficou, notícia)

Para as orações subordinadas, extraem-se as dependências que representam as orações preposicionais, associando a preposição ao respectivo advérbio ou nome, e as dependências entre as conjunções subordinativas e os verbos da oração (vi) subordinada e (vii) subordinante.

- (vi) INTROD («conjunção», «verbo subordinado»)
- (vii) CONNECTOR («verbo subordinante», «conjunção»)

*O João esteve a estudar enquanto comia.*

- INTROD (enquanto, comia)
- CONNECTOR (estudar, enquanto)

É igualmente extraída uma dependência que associa (viii) o início e o final da oração subordinada.

- (viii) QBOUNDARY («conector», «primeiro elemento da oração», «último elemento da oração»)

*O Rui deitou-se no sofá assim que chegou a casa.*

- QBOUNDARY (assim que, assim que, chegou)

Para as relações coordenadas, extraem-se (ix) as relações de coordenação estabelecidas entre as conjunções e os elementos que são coordenados. A coordenação também é extraída com grupos nominais e, em certas circunstâncias, também com grupos preposicionais (x).

- (ix) COORD («conjunção», «elemento coordenado»)
- (x) PREPD («preposição», «nome ou advérbio»)

*O André chegou a casa, poisou a mochila e estendeu-se na cama.*

- COORD (e, chegou)
- COORD (e, poisou)
- COORD (e, estendeu-se)

*A Rita saiu de casa depois do jantar.*

- PREPD (jantar, depois de)

Relativamente aos eventos normalizados, é possível obter-se na STRING as dependências de ocorrência, início e conclusão do evento normalizado, associando-o a expressões temporais.

EVENT\_TIME-CALENDAR («evento», «expressão temporal»)

EVENT\_TIME-DURATION («evento», «expressão temporal»)

EVENT\_TIME-FREQUENCY («evento», «expressão temporal»)

EVENT\_DATE («evento», «expressão temporal»)

EVENT\_DATE-START («evento», «expressão temporal»)

EVENT\_DATE-END («evento», «expressão temporal»)

*O André publicou tudo na sexta-feira.*

EVENT\_TIME-CALENDAR (publicou, em a sexta-feira)

*O João esperou-a durante dois meses.*

EVENT\_TIME-DURATION (esperou, durante dois meses)

*O Ricardo visita a família duas vezes por semestre.*

EVENT\_TIME-FREQUENCY (visita, duas vezes por semestre)

*O pai do João morreu em Setembro de 1990.*

EVENT\_DATE (publicou, em a sexta-feira)

*Almada Negreiros (São Tomé e Príncipe, 1893 - Lisboa, 1970)*

EVENT\_DATE-START (lexnascimento, 1893)

EVENT\_DATE-END (lexmorte, 1970)

A ordem dos elementos sintáticos também é anotada, acrescentando os sufixos *\_PRE* e *\_POST* conforme estas ocorram antes e depois do primeiro argumento da dependência. Por exemplo, na frase que se segue, o sujeito antecede o predicado e, por conseguinte, é adicionado o sufixo *\_PRE* à dependência *SUBJ*. A expressão temporal que está a modificar o predicado, por sua vez, sucede o mesmo e, assim sendo, é adicionado o sufixo *\_POST* à relação *MOD*.

*A Rita passeou neste domingo.*

SUBJ\_PRE (passeou, Rita)

MOD\_POST (passeou, domingo)

Adicionalmente, durante o processamento de texto, são criados vários *chunks*. Por exemplo, na frase utilizada anteriormente, o artigo *a* pertence ao sujeito, pelo que se associa ao mesmo através da dependência *HEAD*. O mesmo sucede com *domingo*, que é considerado a cabeça do complemento preposicional *neste domingo*.

*A Rita passeou neste domingo.*

HEAD (Rita, a Rita)

HEAD (passeou, passeou)

HEAD (domingo, neste domingo)

SUBJ\_PRE (passeou, Rita)

MOD\_POST (passeou, domingo)

## 2.4 Sinopse

Identificaram-se dois tipos de abordagens nos sistemas analisados: sistema baseado em regras e sistemas baseados em aprendizagem automática. Enquanto os sistemas baseados em regras exigem um conhecimento sobre a língua a ser considerada, os sistemas baseados em aprendizagem automática recorrem na maioria a bases de dados lexicais (WordNet) e relacionais (VerbOcean) para complementar a informação obtida a partir dos algoritmos de aprendizagem. O sistema de Maršic (2011), mais recente, pretendeu obter o melhor de ambas as perspectivas, combinando uma abordagem baseada em regras com uma abordagem estatística. À exceção do sistema mais antigo aqui apresentado (TERSEO, de Saquete et al. 2003), os sistemas utilizam o esquema de anotação TimeML ou variantes do mesmo (no caso de TRIPS, TRIOS, XTM e o de Maršic 2011).

O TimeML, tal como foi descrito, baseia-se na relações de Allen [2] para a ordenação de eventos. Deste modo, os sistemas baseados neste esquema utilizaram esta lógica, ou derivações da mesma. A lógica de Reichenbach [3, 27, 45], no entanto, é mais simples ao considerar os eventos como pontos, usando apenas as relações de *antes* e *simultâneo*.

Relativo à ordenação de eventos, alguns dos sistemas descritos, como Chambers et al., NCSU-INDI e NCSU-JOINT, XTM e Maršic, dividiram este problema em ordenações locais e globais, o que permite aplicar estratégias mais especializadas a cada situação.



## Capítulo 3

# Identificação e Ordenação de Eventos Não Normalizados em Português

Este capítulo descreve as classes gramaticais consideradas relevantes para a identificação e ordenação de eventos não normalizados. Apresentam-se os eventos considerados no âmbito do desenvolvimento deste trabalho, verbos plenos ou nomes predicativos, assim como os tipos de conectores discursivos a utilizar para relacionar os eventos extraídos.

Em relação aos eventos, analisam-se as relações de nominalização existentes no português, assim como o significado das diversas conjugações verbais. Analisam-se os diversos valores de modo, tempo e aspeto que uma cadeia verbal pode ter, assim como as relações que se podem estabelecer entre os eventos, através da localização temporal dos mesmos.

Procede-se, então, com a descrição dos conectores considerados, introduzindo as conjunções e preposições relevantes, assim como os advérbios e locuções adverbiais. Após a análise dos mesmos, faz-se um pequeno resumo dos pontos essenciais a reter para a solução que vai ser implementada.

### 3.1 Eventos

#### 3.1.1 Nominalizações

Observam-se frequentemente relações de *nominalização* que permitem a expressão do mesmo evento sob a forma de uma construção verbal ou nominal, acompanhado ou não do verbo-suporte com que o nome predicativo se constrói.

Os verbos-suporte são um tipo particular de verbos que apoiam flexionalmente o nome predicativo. São, em geral, vazios de sentido e não estabelecem restrições de seleção sobre os outros constituintes da frase, os quais são, pelo contrário, seleccionados pelo nome predicativo, que é o núcleo do predicado semântico (ou evento) da frase. Como exemplo, considerem-se as frases:

*O Pedro apresentou o projeto ao João.*

*O Pedro fez uma apresentação do projeto ao João.*

Em ambas as frases, *Pedro* é o sujeito (agente) do evento *apresentar/apresentação*, o *projeto* é o objeto e o *João* é o interlocutor. Considera-se que entre as frases existe uma relação transformacional de *nominalização*. Este tipo de relação é permitido na segunda frase a partir do verbo-suporte, podendo-se afirmar que se está perante diferentes expressões do mesmo

predicado semântico, logo, do mesmo evento. A frase que segue, no entanto, não utiliza um verbo-suporte, sendo visível uma relação diferente da anterior, ainda que se mantenha uma configuração sintática semelhante:

*O Pedro comentou a minha apresentação do projeto do João.*

Nesta frase, as relações ou papéis semânticos de *projeto* e *João* são as mesmas (ambos são argumentos de *apresentação*) mas *Pedro* já não é argumento do nome predicativo, pois o agente deste evento é expresso pelo possessivo *minha*. Pelo contrário, *Pedro* é o sujeito (agente) de *comentar*, enquanto *comentário* é o objeto deste verbo.

O nome predicativo pode ser encontrado na cabeça de um grupo nominal, acompanhado pelos seus argumentos e associado a outro predicado, como no exemplo que se segue:

*A apresentação pública do projeto foi comentada pelo João.*

Na frase acima, o nome predicativo é acompanhado pelo argumento *projeto* e associado ao predicado *comentar*.

### 3.1.2 Verbos

#### Conjugação Verbal

A ordenação de eventos depende de forma crucial dos valores gramaticais que atualizam os predicados no eixo temporal. Os verbos podem ser marcados morfologicamente (ou conjugados/flexionados) quanto às categorias tempo, modo e aspeto [15, 36].

A classificação dos verbos segundo o tempo, assume os valores de *passado*, *presente* ou *futuro*, relativo ao momento de enunciação. Em termos de modalidade, consideramos que podem expressar uma modalidade *real* ou *irreal*. Quanto ao aspeto, os verbos podem ser do tipo *perfetivo* ou *imperfetivo*. O aspeto perfetivo pode representar eventos pontuais ou conclusivos (marcado na STRING com o traço *terminat*). O aspeto imperfetivo pode representar ações durativas, incoativas ou frequentativas (marcados na STRING com os traços *durativo*, *incoat*, *frequentat*).

#### Verbos Auxiliares

Frequentemente, estes valores gramaticais são expressos com recurso a verbos auxiliares [5]. Os verbos auxiliares aspetuais são verbos que permitem atribuir um valor aspetual ao verbo que auxiliam, como exemplificado nas frases que se seguem:

*Ele está a ler o livro.*

*Ele acabou de ler o livro.*

*Ele começou a ler o livro.*

Por seu turno, os auxiliares modais veiculam diferentes tipos de modalidade, tais como *obrigação/necessidade* ou a *possibilidade/probabilidade*:

*O Pedro pode ler o livro.*

*O Pedro tem de ler o livro.*

*O Pedro deve ler o livro.*

*O Pedro há-de ler o livro.*

Os *auxiliares temporais*, como o nome indica, veiculam diferentes valores temporais, e serão descritos com mais promenor na secção seguinte. Tradicionalmente, considera-se apenas o auxiliar *ter* + participípio passado, mas outros auxiliares podem ser também incluídos neste tipo, tal como *ir* + infinitivo:

*O Pedro tem lido o livro.*

*O Pedro vai ler o livro.*

## Modo

Nesta secção apresenta-se de forma sucinta as diferentes interpretações de base da conjugação verbal nas diversas condições de tempo-modo.

Em português, considera-se que a categoria modo, quando é morfologicamente marcada no verbo, apresenta os valores de *indicativo, conjuntivo, condicional e imperativo*.<sup>1</sup>

O *modo indicativo* representa um evento com uma modalidade real e, em geral, localiza-o temporalmente. O *modo conjuntivo* assinala a subordinação de uma oração a um verbo principal (ex: *Ele queria que eu fosse*), exprimindo, de um modo geral, a modalidade *irreal*. O *modo condicional*, em termos de modalidade, exprime uma modalidade *irreal* dependente de uma condição, tendo tendência a ser utilizado com orações subordinadas condicionais (ver Secção 3.2.2), projetando o evento no futuro do tempo de referência, expresso na condição:

*Eu iria se ele não fosse.*

O *modo imperativo* expressa um pedido ou ordem, exprimindo uma modalidade *irreal* e projetando, em geral, o evento no futuro do tempo de referência.

## Interpretação da conjugação verbal

Esta secção, foca-se na interpretação dos diversos valores de flexão da categoria *tempo*.

O *presente do indicativo* é um tempo verbal não marcado, pelo que não permite identificar, por si só, a localização temporal dos eventos. É possível, no entanto, estabelecer uma localização temporal no momento de enunciação com o auxílio de outros elementos, como os advérbios (descritos na Secção 3.3) ou os verbos aspetuais, como, por exemplo, nas frases:

*Eu chego hoje a Lisboa.*

*Eu estou a chegar a Lisboa.*

A conjugação do verbo no *pretérito perfeito do indicativo* expressa a ocorrência de um evento com início e fim no passado, como, por exemplo, na frase:

*Eu cheguei a Lisboa.*

O *pretérito perfeito composto*, à semelhança do *pretérito perfeito simples*, expressa um evento que se iniciou no passado. No entanto, pode expressar um evento iterativo que se prolonga até ao presente ou se projecta no futuro, como no exemplo que se segue:

*O Manuel tem visitado a avó todos os dias.*

---

<sup>1</sup>O chamado *modo infinitivo* é, justamente, o caso de flexão em que as categorias tempo e modo não são marcadas, dependendo dos valores do verbo principal de que o infinitivo depende.

O *pretérito mais-que-perfeito do indicativo*, simples ou composto, explicita que o evento ocorreu no passado e terminou antes de outro evento ter tido lugar, tendo um aspeto perfeitivo, utilizando um ponto de referência para se localizar temporalmente (ver Secção 1.3.3):

*Nunca ouvira tal coisa (até ouvir o que disseste/foi dito).*

*Nunca tinha ouvido tal coisa (até ouvir o que disseste/foi dito).*

A conjugação de verbos no *pretérito imperfeito do indicativo* pode indicar um evento imperfeito ou habitual com o seu início no passado (ex: *Ela comia*), mas pode muitas vezes corresponder a uma modalidade *condicional*, como na frase:

*Neste momento, já comia qualquer coisa.*

A conjugação do verbo no *futuro do indicativo* expressa um evento que ainda não se iniciou, ou seja, uma modalidade irreal, mas que pressupõe uma elevada probabilidade, como exemplificado:

*Eu chegarei a Lisboa.*

O *modo conjuntivo* pode ser conjugado no *pretérito imperfeito (que/se eu ganhasse)*, no *presente (que eu ganhe)*, *futuro (quando eu ganhar)*, *pretérito perfeito composto (que eu tenha ganho)*, *pretérito mais-que-perfeito composto (se eu tivesse ganho)* ou *futuro composto (quando eu tiver ganho)*.

Os *tempos compostos do conjuntivo* possuem um aspeto perfeitivo, devido ao uso do particípio passado. Quanto à relação temporal estabelecida, na conjugação (simples ou composta) do modo conjuntivo, apenas os tempos do futuro exprimem necessariamente uma localização temporal relativa ao momento de enunciação. O futuro simples do conjuntivo, em orações subordinadas temporais e condicionais (ver Secção 3.2.2), expressa a ocorrência do evento no futuro, como no exemplo:

*Ele (só) vai quando/se eu for.*

Neste caso, o tempo do evento expresso na oração subordinada pelo conjuntivo pode ser concomitante/simultâneo ao tempo do evento da oração principal (dependendo da conjunção subordinativa). O futuro composto do conjuntivo, por sua vez, expressa a ocorrência do evento no passado relativamente ao tempo de referência expresso na oração principal, ou seja, o evento da oração subordinada ocorre no futuro, mas é anterior ao evento da oração principal, como no exemplo:

*Ele (só) vai quando eu tiver ido.*

O verbo da oração principal pode ocorrer não só no presente do indicativo, como nos exemplos utilizados, mas também no futuro do indicativo:

*Ele (só) irá quando/se eu for.*

*Ele (só) irá quando eu tiver ido.*

O que confirma o valor futuro do emprego do presente do indicativo nos exemplos anteriores, como um tempo não marcado.

### **Relações estabelecidas através da conjugação verbal**

Em algumas situações, os tempos verbais são suficientes para se estabelecer as relações temporais entre os eventos envolvidos num par de orações principal-subordinada. Assim, e usando a conjunção subordinativa *quando* como exemplo, é possível determinar que o pretérito mais-que-perfeito da oração principal localiza o evento num tempo *anterior* ao tempo da subordinada quando o verbo desta ocorre no pretérito perfeito:

*O Rui já jantara/tinha jantado quando o João chegou a casa.*

No exemplo, do ponto de vista aspetual, o evento da oração principal é interpretado como concluído (perfectivo). Em comparação, o uso do pretérito imperfeito na oração principal implica a simultaneidade dos dois eventos:

*O Rui já jantava quando chegaste a casa.*

Estas relações podem ser estabelecidas *a priori* com base nos valores gramaticais destes diferentes tempos verbais, nomeadamente quanto ao tempo e ao aspeto e estão, aliás, na base da terminologia gramatical que é normalmente usada para os designar: um verbo conjugado no pretérito mais-que-perfeito exprime um valor perfectivo num tempo anterior a um evento de referência pretérito, enquanto um pretérito imperfeito exprime o aspeto inacabado de um evento que pode ser intersectado por outro.

### Formas Nominais

A relação estabelecida pelo infinitivo, como já referido antes, depende do tempo-modo do verbo da oração principal a que está subordinado, bem como do conector utilizado:

*Eu preparei/preparo/prepararei a mesa antes de ele fazer isso;*

*Eu preparei/preparo/prepararei a mesa depois de ele fazer isso;*

*Eu preparo/prepararei a mesa quando ele fizer isso.*

Note-se que, consoante o conector, pode haver restrições também as combinações de tempo-modo. Assim, se a conjunção for *quando*, o pretérito perfeito deixa de ser aceitável. Por outro lado, um pretérito mais-que-perfeito, dada a sua referência relativa, também é inaceitável neste contexto:

*\*Eu preparara a mesa quando ele fizer isso.*

*\*Eu preparei a mesa quando ele fizer isso.*

O gerúndio<sup>2</sup> é uma forma dita “nominal” do verbo que o subordina a uma oração principal sem, no entanto, explicitar o nexo semântico entre as duas orações. Quando tem uma interpretação temporal, exprime a concomitância ou simultaneidade dos eventos, como no exemplo que se segue:

*A professora explicou o texto lendo outra vez as passagens mais difíceis.*

Quando empregue na forma composta, é mais frequente ser interpretado com um valor causal que é (temporalmente) anterior ao evento da oração principal:

*Tendo lido o texto, percebi melhor a intenção do autor.*

O participípio passado, quando não é empregue numa conjugação perifrástica, é usado para formar as chamadas *orações reduzidas participiais*. Tem, nesses casos, um valor perfectivo, mas a sua referência temporal depende do tempo da oração principal:

*Lidos os textos, a professora explicará o assunto.*

Estas formas estão muitas vezes associadas a paráfrases introduzidas pela conjunção subordinada *uma vez*:

*Uma vez lidos os textos, a professora explicará o assunto.*

---

<sup>2</sup>Como é óbvio, não se consideram aqui as construções de verbos auxiliares formadas com o gerúndio: *estar fazendo*; *continuar fazendo*; *vir fazendo*.

Tipo de conector	Exemplos de utilização
Conjunção Coordenativa	<i>O Pedro leu e corrigiu o trabalho.</i>
Conjunção Subordinativa	<i>O Pedro leu o trabalho enquanto vinha no comboio.</i>
Advérbio	<i>O Pedro leu o trabalho posteriormente.</i>
Preposição	<i>O Pedro leu o trabalho depois do almoço.</i>

Tabela 3.1: Classes morfossintáticas consideradas na ordenação de eventos

### 3.1.3 Nomes predicativos

Os nomes predicativos são nomes abstratos que exprimem diferentes tipos de predicados semânticos. Podem designar *qualidades, ações, estados, sentimentos* ou *sensações*, como *vício, corrida, sonho, vergonha e fome*. Muitos destes nomes apresentam uma construção verbal associada e podem ser morfologicamente derivados dos respetivos verbos, como *corrida, sonho e vício* que se obtêm a partir de *correr, sonhar e viciar*.

Os nomes predicativos são auxiliados, muitas vezes, por um verbo-suporte, o qual veicula os valores gramaticais de tempo-modo, aspeto e pessoa-número que o nome, pela sua morfologia, não consegue expressar, como exemplificado na frase que se segue:

*O Pedro teve um encontro com a Ana.*

Alguns nomes predicativos não possuem uma construção verbal associada (e.g.: *crime*) ou, à semelhança do nome *aliança*, podem apresentar mais do que um significado, podendo ser utilizado como nome predicativo ou nome concreto, como exemplificado:

*O Pedro encontrou uma aliança* (concreto);

*O Pedro fez uma aliança com o João* (abstracto/predicativo).

Os nomes predicativos podem ainda ocorrer como argumentos de outros predicados, acompanhado dos seus respetivos argumentos, como na frase:

*O encontro entre o Pedro e a Ana decorreu agradavelmente.*

## 3.2 Conectores Discursivos

Para a ordenação de eventos consideram-se as classes das conjunções coordenativas, mas sobretudo as subordinativas, os advérbios e as preposições [36], como exemplificado na Tabela 3.1. Estas classes apresentam expressões que constituem palavras simples (uma única palavra gráfica) ou compostas (mais do que uma palavra gráfica). Neste último caso, são normalmente designadas como locuções conjuncionais, adverbiais e preposicionais<sup>3</sup>.

As conjunções coordenativas permitem ligar duas orações na mesma frase ou dois constituintes da mesma oração, como nos exemplos que se seguem:

*O Pedro leu e corrigiu o trabalho;*

*O Pedro e a Ana leram o trabalho.*

<sup>3</sup>Ao longo deste documento, os termos *conjunção, preposição* e *advérbio* serão utilizados com um significado mais genérico, considerando implicitamente as respetivas locuções (conjuntivas, preposicionais ou adverbiais).

As conjunções subordinativas adverbiais costumam organizar-se em classes semânticas, consoante o nexos que estabelecem entre a oração principal e a subordinada.

As conjunções coordenativas consideradas na STRING são classificadas através de traços correspondentes às funções que as conjunções desempenham: *aditivas*, *disjuntivas* ou *adversativas*. As conjunções subordinativas, por sua vez, podem ser classificadas como conjunções *temporais*, *causais*, *comparativas*, *condicionais*, *concessivas*,  *finais* ou *causais*.

### 3.2.1 Conjunções Coordenativas

As conjunções coordenativas consideradas neste trabalho são as conjunções aditivas (ou copulativas), adversativas e disjuntivas (ou alternativas).

As conjunções aditivas expressam ideia de acréscimo ou adição à oração anterior, permitindo enumerações de eventos ou conceitos. No entanto, não é possível estabelecer uma ordem precisa entre os eventos envolvidos, como no exemplo:

*Ele comeu e bebeu na festa.*

A conjunção *e*, em especial, pode ainda estabelecer uma relação de causa ou sequência temporal entre dois eventos:

*Ele disse isso e ela amou.*

As conjunções adversativas exprimem uma ideia de oposição ou contraste à oração anterior com base no conhecimento sobre o mundo. Seguem-se dois exemplos:

*Este livro é interessante apesar de caro;*

*A Ana estudou mas reprovou.*

Na primeira frase apresenta-se uma qualidade considerada positiva e outra negativa, salientando-se o contraste com o uso de uma conjunção adversativa. No segundo exemplo, era esperado que a Ana não reprovasse dado que estudou, pelo que se utiliza uma conjunção adversativa.

As conjunções disjuntivas (ou alternativas) dividem ou invalidam conceitos na presença de outros, apesar de não estabelecerem nenhuma ordenação entre os eventos envolvidos. Por outras palavras, na ocorrência de um evento, o outro não pode ocorrer, como no exemplo:

*Amanhã, ou chove ou faz sol.*

As conjunções coordenativas, de uma forma geral, permitem considerar que os eventos ocorrem num mesmo intervalo de tempo ou, no caso das disjuntivas, que apenas um deles ocorre em determinado intervalo de tempo. É possível, no entanto, ordená-los temporalmente com o auxílio de advérbios, como no exemplo:

*Ou chove amanhã, ou chove no dia a seguir.*

Numa enumeração, as vírgulas podem ser utilizadas com propriedades coordenativas (aditivas ou disjuntivas), apesar de não serem conjunções. Assumem o valor da conjunção aditiva ou disjuntiva que lhes sucede, como nos exemplos que se seguem:

*Ele comeu, bebeu e dançou na festa.*

*A culpa foi da Rita, da Inês ou do Rui.*

### 3.2.2 Conjunções Subordinativas

#### Conjunções Causais

As conjunções causais indicam que determinado evento (oração subordinada) depende de outro (oração subordinante) para ocorrer. Esta dependência estabelece uma ordem entre os eventos, em que a oração subordinada ocorre antes da oração subordinante, como exemplificado nas frases que se seguem através das conjunções *como* e *uma vez que*:

*Como estava sol, ele foi à praia.*

*Uma vez que estamos cansados, vamos ver um filme.*

Nos exemplos acima, o evento *ir à praia* só foi possível por *estar sol* e o evento *ver um filme* dependeu da ocorrência do evento *estar cansado*. Estas conjunções estão sujeitas às influências de outros elementos gramaticais, como advérbios, podendo ser alterada a ordem dos eventos, como no exemplo:

*Como vai chover amanhã, comprei um guarda-chuva.*

#### Conjunções Consecutivas

As conjunções consecutivas, expressam uma relação de consequência entre as orações. O tipo de relação estabelecido entre este tipo de conjunções é inverso ao das conjunções *causais*, ou seja, os eventos presentes na oração subordinada sucederam-se aos da oração subordinante. Seguem-se exemplos do uso de conjunções consecutivas, utilizando-se, neste caso, as conjunções: *que* e *a ponto de*.

*Ele gritou de tal forma que ficou rouco.*

*Ele gritou a ponto de ficar rouco.*

#### Conjunções Comparativas

As conjunções comparativas ligam duas orações estabelecendo uma comparação entre o evento da oração principal e o evento da oração subordinada. Por essa razão, este último é geralmente interpretado como tendo ocorrido num momento anterior ao evento da oração principal:

*Fazemos/Fizemos/Faremos isso como planeávamos.*

Repare-se que, independentemente do tempo do verbo da oração principal, o pretérito perfeito usado na subordinada comparativa remete sempre para o tempo anterior ao evento principal. O uso dos outros tempos é inaceitável:

*\*Fazemos/Fizemos/Faremos isso como planearemos;*

*\*Fazemos/Fizemos/Faremos isso como planeamos.*

Nestes dois exemplos, o presente do indicativo na oração principal remete para o futuro relativamente ao momento de enunciação. Apesar de os dois tempos não serem logicamente incompatíveis, as frases com a comparativa no presente são, em regra geral, inaceitáveis.

As orações comparativas permitem muitas vezes a redução do material repetido que, no entanto, pode ser reconstruído:



*Tal como ele [não sabia a resposta]<sub>1</sub>, ela também [não sabia a resposta]<sub>2</sub>.*

No exemplo acima, tanto a expressão delimitada com 1 como a expressão assinalada com 2 podem ser reduzidas em alternativa. Efetivamente, as comparativas pressupõem sempre repetição de material da oração principal. Mesmo no exemplo anterior é possível reconstituir uma segunda ocorrência do verbo fazer:

*Fazemos isso como planeámos [fazer isso].*

### **Conjunções Condicionais**

As conjunções condicionais introduzem um evento que é condição para que se verifique o evento da oração principal, pelo que se pressupõe a anterioridade temporal da oração subordinada à principal. Seguem-se alguns exemplos, usando as conjunções *caso* e *se*:

*Caso o João não venha, a reunião não se realizará;*

*Se o João tivesse vindo, a reunião teria corrido bem.*

### **Conjunções Concessivas**

As conjunções concessivas (e.g.: *apesar de*, *embora*), à semelhança das conjunções adversativas, exprimem um conteúdo semântico que contrasta com o que, dado o nosso conhecimento, se espera em determinada situação. Considerem-se os exemplos:

*Eu fui/vou/irei ao cinema apesar de não gostar muito do filme.*

De um modo geral, o evento da subordinada é anterior no tempo do evento da principal. Ao contrário da conjunção *apesar de*, que seleciona o modo infinitivo, a conjunção *embora* seleciona, por seu turno, o modo conjuntivo. Neste caso, verifica-se a obrigatória subordinação temporal da concessiva ao tempo da oração principal:

*Eu fui ao cinema embora não gostasse muito do filme;*

*Eu vou/irei ao cinema embora não goste muito do filme.*

A presença de certos modificadores adverbiais pode, no entanto, permitir uma ordenação temporal inversa:

*Não comprei um guarda-chuva apesar de ir chover amanhã;*

*Não comprei um guarda-chuva embora vá chover amanhã.*

### **Conjunções Proporcionais**

As conjunções proporcionais (e.g.: *à medida que*) expressam uma evolução, em simultâneo e em proporção, de acontecimentos presentes nas orações, exprimindo uma ordem entre os eventos. Os eventos da oração subordinada iniciam-se em primeiro lugar, como exemplificado na frase que se segue:

*À medida que os alunos chegam, vão recebendo o livro.*

Frase	Associação estabelecida
O João viu o filme <i>assim que</i> leu o livro.	<i>ver</i> é posterior a <i>ler</i> .
O João viu o filme <i>depois de</i> ler o livro.	<i>ver</i> é posterior a <i>ler</i> .
O João viu o filme <i>antes de</i> ler o livro.	<i>ver</i> é anterior a <i>ler</i> .
O João viu o filme <i>enquanto</i> comia pipocas.	<i>ver</i> é simultâneo a <i>comer</i> .
O João come pipocas <i>sempre que</i> vê filmes.	<i>comer</i> é simultâneo a <i>ver</i> .

Tabela 3.2: Exemplos de orações subordinadas temporais

### Conjunções Finais

As conjunções finais (e.g.: *para que*; *com o objetivo de*) expressam uma relação de finalidade entre os eventos, explicitando uma ordem entre os mesmos: os eventos presentes nas orações subordinadas ocorrem depois dos eventos da oração subordinativa. Considerem-se os exemplos:

*Tudo foi planeado para que funcionasse na perfeição.*

*Tudo foi planeado com o objetivo de funcionar na perfeição.*

### Conjunções Temporais

As conjunções temporais relacionam temporalmente as orações que ligam, especificando se os eventos são concomitantes, se um ocorre depois do outro ou vice-versa. A Tabela 3.2 representa alguns exemplos deste tipo de conjunções.

### 3.2.3 Preposições

Uma preposição é uma palavra invariável que liga dois elementos numa oração, subordinando o segundo ao primeiro. As preposições podem estabelecer relações semelhantes às conjunções, podendo-se encontrar preposições do tipo *causal*, *comparativo*, *consecutivo*, *final*, *disjuntivo* e *temporal*, como exemplificado de seguida:

- (i) *O Pedro não conseguiu fazer o que queria* devido à *falta de tempo* (causal);
- (ii) *O trabalho do Pedro foi feito* à semelhança do *projetado* (comparativo);
- (iii) Em consequência do *trabalho realizado*, o *Pedro recebeu a devida nota* (consecutivo);
- (iv) *O Pedro fez as coisas* em função do *objetivo* (final);
- (v) Ao contrário do *pretendido*, a *apresentação do Pedro correu mal* (disjuntivo);
- (vi) *O Pedro acabou tudo* ao fim do *dia* (temporal).

Podem-se ainda encontrar preposições que associam eventos, utilizando um evento como referência ao outro, classificadas na STRING com o traço *topic* (e.g., *em relação a*; *acerca de*). Neste caso, verifica-se uma ordenação temporal implícita entre os eventos relacionados por estas preposições, sendo o tempo do verbo principal *anterior* ao do evento introduzido pelas preposições, como no exemplo:

*O Rui esteve a falar acerca do jantar.*

No entanto, na presença de advérbios temporais, estas preposições podem estabelecer outras relações (*simultâneo* ou *posterior*), como nos exemplos que se seguem:

*O Rui esteve a falar acerca do jantar de amanhã;*

*O Rui esteve a falar acerca da nossa situação atual.*

Frequentemente, as conjunções podem introduzir predicados nominais, sob a forma de grupos nominais, caso em que são tradicionalmente analisadas como preposições. A STRING manteve esta distinção:

*O Pedro leu o livro antes da conversa com o João.*

Ora, nestes casos, considera-se haver uma redução da oração a grupo nominal por apagamento do verbo-suporte do nome predicativo:

*O Pedro leu o livro antes de ter tido a conversa com o João.*

Outras reduções vão mais longe ainda, omitindo todo o predicado introduzido pela conjunção/preposição:

*O Pedro leu o livro antes da Ana = O Pedro leu o livro antes de a Ana o ler.*

Naturalmente, neste trabalho, casos como o anterior não irão ser tratados já que não há lá nenhum predicator/operador expresso no constituinte introduzido pela conjunção/preposição.

Por fim, as preposições que servem apenas de elementos de ligação entre constituintes não serão considerados. Tratam-se, por exemplo, das preposições utilizadas pelos verbos para introduzir os respetivos complementos (e.g. *O Pedro gosta da Ana*) ou das preposições que ligam auxiliares ao verbo principal (e.g. *O Pedro tem de ler o livro*). Contudo, muitos verbos podem selecionar, como seus argumentos, outros predicados, seja sob a forma de orações completivas, seja sob a forma de predicados nominais:

*O Pedro gosta de surfar/fazer surf/surf.*<sup>4</sup>

De um modo geral, dependendo do verbo principal e do tempo em que este se encontra conjugado, a ordenação temporal dos eventos deverá considerar, na ausência de outros modificadores específicos, que o evento do complemento ocorre antes do evento do verbo principal. Assim, na frase acima, o uso do presente do indicativo conjugado com o infinitivo da completiva-objeto leva à interpretação de que *surfar/fazer surf* é uma atividade habitual pelo que haveria simultaneidade entre este evento e o que é denotado por gostar. Ora, se a frase estiver no pretérito perfeito, a ordenação temporal dos eventos torna-se mais evidente:

*O Pedro gostou de surfar/fazer surf/surf.*

Certas formas de subordinação empregam elementos conectores que, superficialmente, parecem preposições simples. É o caso dos exemplos seguintes:

*O Pedro comia pipocas ao ver o filme;*

*O Pedro comia pipocas a ver o filme.*

Nos exemplos apresentados, há uma relação de simultaneidade entre *comer* e *ver*. Tratam-se, porém, de processos distintos, já que o subordinador *ao* pode ter um emprego causal, o que leva a uma ordenação distinta dos eventos envolvidos:

*O Pedro parou o carro ao ver o acidente.*

---

<sup>4</sup>A presença do nome de um desporto, nesta situação, tem duas interpretações: (i) pode ser analisado como o resultado da redução do verbo-suporte, como foi atrás explicado; ou (ii) designar apenas a atividade sem que o sujeito do verbo principal seja interpretado como seu sujeito. Por outras palavras, o Pedro pode gostar de surf mas não praticar a atividade. Estas distinções, no entanto, não serão consideradas.

Finalmente, refiram-se ainda as preposições/conjunções causais *de* e *por* que se ilustram nos exemplos seguintes:

*O Pedro está cansado de correr/da corrida;*

*O Pedro está contente por aquele resultado/ por ter tido aquele resultado.*

A preposição *de* pode ainda associar advérbios temporais a nomes predicativos, estabelecendo assim a sua localização no tempo:

*O Pedro falou no jogo de amanhã/ontem/hoje.*

Como noutras situações, a ordenação dos eventos está dependente tanto do tempo do verbo como dos modificadores temporais específicos dos eventos envolvidos (*falar, jogo*). Assim, com os modificadores *de ontem* e *de hoje*, o evento *jogo* é situado antes do evento *falar*, enquanto com o modificador *de amanhã*, a ordenação dos eventos é a inversa.

### 3.3 Advérbios e Locuções Adverbiais

Os advérbios, em particular os advérbios temporais e os conjuntivos, desempenham um papel importante na ordenação de eventos. Os advérbios temporais podem expressar diferentes valores: *frequências, durações* e *data* (referência relativa). Os advérbios conjuntivos, por sua vez, permitem estabelecer relações entre frases distintas. Seguem-se exemplos de frases com advérbios temporais:

*O João lê diariamente o jornal.*

*O João leu ontem um livro inteiro.*

*Antigamente, o João lia muitos policiais.*

*O João lia todo o santo dia.*

Na primeira frase, o advérbio *diariamente* exprime um valor temporal de *frequência*. Na segunda e terceira frase, os advérbios *ontem* e *antigamente* exprimem, respetivamente, uma referência relativa. No quarto exemplo, a locução adverbial *todo o santo dia* expressa apenas a duração do evento presente na frase, não estabelecendo qualquer outro tipo de relação temporal.

Os advérbios conjuntivos estabelecem uma associação entre a frase onde se encontram e a anterior, não podendo ocorrer no início absoluto de um discurso. No entanto, é possível encontrar este tipo de advérbios, devido à sua natureza conjuncional, a fazer a junção das frases que associam. Para além disso, estes advérbios demonstram ter, muitas vezes, mobilidade na frase onde se encontram. Os exemplos que se seguem são equivalentes em conteúdo semântico e demonstram estas características dos advérbios conjuntivos:

*O Pedro fez isso. Por conseguinte, também devia fazer aquilo.*

*O Pedro fez isso, por conseguinte, também devia fazer aquilo.*

*O Pedro fez isso. Também devia, por conseguinte, fazer aquilo.*

Os advérbios conjuntivos apresentam funções e relações semelhantes às conjunções, podendo expressar relações como as conjunções *aditivas* (e.g, *Além do mais*), *consecutivas* (e.g, *consequentemente, por conseguinte*) e *adversativas* (e.g, *no entanto*), entre outros, estabelecendo as ordenações à semelhança das conjunções respetivas. Seguem-se alguns exemplos:

*O Pedro fez isso. Além do mais, ainda fez aquilo;*

*O Pedro fez isso. Por conseguinte, devia ter feito aquilo;*

*O Pedro fez isso. No entanto, devia ter feito aquilo.*

Contudo, estes advérbios podem ainda estabelecer enumerações (e.g.: *primeiramente, em seguida, por último*) estabelecendo relações de posterioridade com os eventos da frase anterior a que se associam, como no exemplo:

*O Pedro fez isso. Em seguida, fez aquilo.*

É possível ainda estabelecer relações de simultaneidade entre os eventos, utilizando advérbios *equalizadores* (e.g.: *paralelamente*), *reformulativos* (ex: *em suma, resumidamente*) e *apositivos* (ex: *nomeadamente*), como exemplificado:

*O Pedro fez isso. Paralelamente, também fez aquilo;*

*O Pedro fez isso. Em suma, não fez aquilo;*

*O Pedro fez isso quando devia ter feito mais. Nomeadamente, devia ter feito aquilo.*

À semelhança das conjunções, não se pode considerar apenas o significado dos advérbios conjuntivos para relacionar e ordenar os eventos. Na frase que se segue, o advérbio *portanto* apresenta uma função consecutiva entre os eventos *visitar* e *preparar*. No entanto, devido à presença dos advérbios temporais, o evento *visitar* ocorre depois do evento *preparem*.

*Vou visitar-vos amanhã. Portanto, preparem tudo hoje.*

Muitos dos advérbios temporais, simples e compostos, são já, em grande parte, identificados pela STRING enquanto entidades mencionadas de tempo. Muito dos advérbios conjuntivos já são identificados e, no caso dos advérbios derivados e terminados em *-mente* mais frequentes, estes já foram igualmente classificados.

### 3.4 Conclusão

A identificação de eventos a partir de nomes predicativos levanta alguns problemas por existirem palavras que podem tomar mais que um significado, como *aliança*. Muitos, no entanto, estão morfologicamente associados a verbos ou apresentam-se construídos com verbos-suporte, tornando-se mais fácil identificá-los. Por vezes, os valores gramaticais de tempo, aspeto e modo são veiculados através de verbos auxiliares. Por exemplo, o evento da frase que se segue obtém a sua localização temporal, relativo ao momento de enunciação, a partir do verbo auxiliar:

*O Pedro vai ler o livro.*

Noutras situações, os valores gramaticais são veiculados ou influenciados através das conjunções, preposições ou advérbios.

As conjunções e as preposições permitem associar eventos dentro da mesma frase, mas as conjunções ligam orações e as preposições ligam elementos da oração. Algumas preposições e conjunções, no entanto, podem desempenhar funções diferentes na frase onde se inserem, dependendo dos elementos gramaticais a que se associam. A preposição *de*, por exemplo, pode ser utilizada apenas como conector entre um verbo e o seu complemento (1) ou associar uma expressão temporal a um evento (2).

1. *Ele gosta do Verão.*

2. *O filme de ontem foi muito bom.*

Os advérbios conjuntivos assemelham-se às conjunções, mas ligam eventos entre frases vizinhas (e.g.: *O evento foi um sucesso. Por conseguinte, gostava de vos agradecer*) e, à semelhança das conjunções, sujeitam-se à influência de advérbios temporais. Os advérbios temporais, quando associados a eventos, permitem localizar eventos na linha temporal, ordenando-os indiretamente, como no exemplos:

1. *Ontem ele leu o jornal.*

2. *Hoje comprou uma revista.*



## Capítulo 4

# Identificação, Relação e Ordenação de Eventos

Neste capítulo descreve-se a arquitetura da solução, assim como os problemas encontrados no desenvolvimento da solução. Representa-se uma segunda vez a arquitectura da STRING, desta vez com a introdução do novo módulo, o modo como este está estruturado e como interage com o resto do sistema, descrevendo o tipo de ficheiros de entrada e saída do mesmo.

Divide-se a solução em duas partes: a extração e classificação de relações; a interpretação das mesmas e dos eventos envolvidos, resultando na identificação da ordenação dos eventos existentes. Para a fase de extração de relações, descrevem-se as dependências introduzidas e utilizadas na resolução do problema, detalhando a solução desenvolvida. Na ordenação dos eventos, descrevem-se as dependências introduzidas para a representação da ordem, o modo como se construiu a solução, a partir da análise dos eventos, conectores e dependências previamente extraídas, e, por fim, a representação gráfica adoptada.

### 4.1 Arquitetura

Durante o decorrer deste trabalho, expandiu-se o módulo XIP de modo a possibilitar a identificação de um maior número de eventos, tanto a partir de verbos como de nomes predicativos. Procedeu-se igualmente à criação e integração de um módulo para identificar as relações entre eventos e ordená-los temporalmente. Este módulo foi introduzido na cadeia de processamento, extendendo o sistema actual, tal como está representado na Figura 4.1.

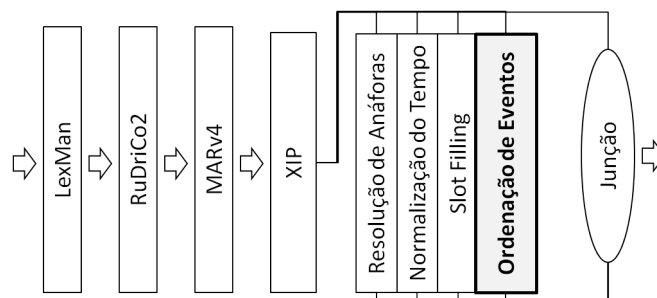


Figura 4.1: Representação da arquitetura da solução.

### 4.1.1 Estrutura do Módulo de Ordenação de Eventos

O módulo de Ordenação de Eventos, representado na Figura 4.2, recebe um ficheiro processado pelo XIP e procede à ordenação dos eventos previamente identificados, devolvendo o resultado no mesmo formato do documento de entrada e como grafo. O módulo subdivide-se em duas partes: identificação das relações entre eventos e ordenação dos eventos relacionados.

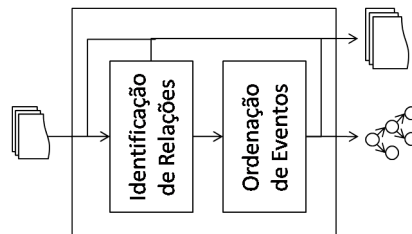


Figura 4.2: Representação da arquitetura interna do módulo de Ordenação de Eventos.

A solução inicia-se com a identificação das relações entre eventos e expressões temporais, com base na análise das dependências extraídas pelo XIP, frase a frase, descrito em mais detalhe na Secção 4.2. De seguida, utilizam-se essas relações para a identificação de novas relações, em especial, expandindo-se as relações que são comuns entre os eventos considerados agrupados, e na filtragem das relações, como um todo, de modo a evitar e a rejeitar contradições. Após este processo, obtêm-se as relações a serem utilizadas, em junção com a informação obtida do XIP, para a ordenação dos eventos.

A ordenação de eventos, na Secção 4.3, por sua vez, baseia-se na informação extraída a partir dos eventos, dos conectores e das expressões temporais, utilizando as relações detetadas previamente e devolvendo a ordenação dos eventos no formato XML (eXtensible Markup Language) [9] e em forma de grafos.

### 4.1.2 Ficheiros de Entrada e Saída

O ficheiro de entrada do módulo é um ficheiro de saída do XIP, no formato XML, que apresenta os seguintes elementos na sua estrutura <sup>1</sup>:

- XIPRESULT: que contém uma ou mais LUNITs;
- LUNIT: corresponde a uma unidade linguística; cada elemento deste tipo vai corresponder a uma frase, que pode conter um conjunto de nós (NODEs) e de dependências (DEPENDENCY);
- NODE: corresponde ao resultado da análise morfossintática. Pode conter outros NODEs ou TOKENS;
- DEPENDENCY: corresponde a relações sintáticas entre os nós. <sup>2</sup>

A interpretação do ficheiro de entrada é feita com o auxílio de uma biblioteca em java auxiliar do sistema. Após o processamento deste ficheiro, tirando proveito em especial dos atributos LUNIT, NODE e DEPENDENCY, o sub-módulo encarregue da identificação de relações extrai e adiciona novas dependências ao ficheiro de entrada. Adicionalmente, o módulo constrói a *árvore de relações* com base nas dependências identificadas e que virá a ser utilizada para a ordenação dos eventos.

A árvore de relações, exemplificado na Figura 4.3, é constituída por nós, em que cada nó representa um evento ou grupo, ligados através de relações de eventos. Um grupo é representado pelo conector, se existente, e a raiz da árvore representa a

<sup>1</sup>Note-se que cada um desses elementos apresenta ainda outros atributos, não especificados neste documento.

<sup>2</sup>As dependências são igualmente utilizadas para a representação das entidades mencionadas (NE), bem como para outro tipo de relações, de natureza semântica, tais como os papéis semânticos (*semantic roles*), que não são relevantes para o presente trabalho.



frase em análise, agrupando os nós não relacionados e implicando a criação de uma árvore por frase. No exemplo dado, como não foi possível associar os eventos *regressássemos* e *trabalho* aos restantes, o nó que representa o evento *regressássemos* foi associado à raiz.

As crianças brincavam e conversavam enquanto esperavam que regressássemos do trabalho.

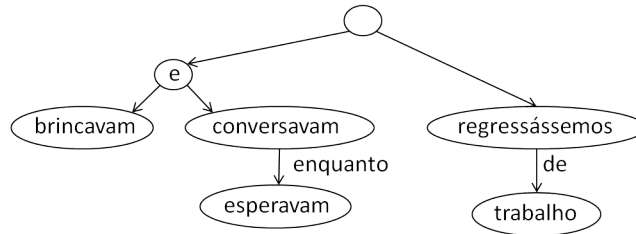


Figura 4.3: Representação da arquitetura interna do módulo de Ordenação de Eventos.

Cada nó da árvore contém as expressões informações relativas ao evento, como a classificação do tipo de evento, os traços associados, entre outros, e as temporais associados aos mesmos, obtidas a partir das dependências existentes no ficheiro de entrada.

É ainda possível unir diversas árvores numa só ao se agrupar as raízes das mesmas, o que pode ser feito ainda durante a fase de extração de relações entre os eventos da frase, e, conseqüentemente, propagar-se as dependências que se considerar necessárias. Este processo adicional ainda não foi concretizado por se encontrar fora do âmbito deste trabalho. Contudo, pretendeu-se que a estrutura criada neste momento fosse flexível tendo em vista o futuro desenvolvimento do projeto.

À saída do módulo de ordenação de eventos, adicionam-se, ao ficheiro de entrada, as dependências de ordem extraídas e cria-se um ou mais grafos com a respetiva ordem, usando a linguagem DOT [24].

## 4.2 Identificação de Relações de Eventos

### 4.2.1 Dependências

Para a identificação das relações entre eventos, criaram-se quatro tipos de dependências distintas, especificadas nas diretivas desenvolvidas no âmbito deste trabalho (ANEXO B):

- `EVENT_DIRECTRELATION («evento», «evento»)`
- `EVENT_INDIRECTRELATION («evento», «evento», «conector»)`
- `EVENT_GROUPRELATION («evento», «evento» [, «conector»])`
- `EVENT_TIMERELATION («evento», «expressão temporal»)`

Representam-se as relações de subordinação entre eventos através das dependências `EVENT_DIRECTRELATION` e `EVENT_INDIRECTRELATION`. A dependência `EVENT_INDIRECTRELATION` distingue-se da primeira pela presença da conjunção ou preposição que estabelece a relação. Adicionalmente, as relações de coordenação adversativas são também representadas por esta dependência.

*A Rita saiu de casa depois do jantar*

```
EVENT_INDIRECTRELATION(saiu, jantar, depois de)
```

*Enquanto comia,o rapaz lia o livro.*

EVENT\_INDIRECTRELATION (comia, lia, enquanto)

*Ele atrasou-se, mas ainda chegou a tempo de ver o espectáculo.*

EVENT\_INDIRECTRELATION (atrasou-se, chegou, mas)

Na ausência de um conector, as relações de subordinação são representadas pela dependência EVENT\_DIRECTRELATION.

*O trabalho publicado pelo Rui foi analisado pelo Júri.*

EVENT\_DIRECTRELATION (trabalho, publicado)

EVENT\_DIRECTRELATION (trabalho, analisado)

A dependência EVENT\_GROUPRELATION representa um agrupamento de dois eventos coordenados, associados sem conector ou através de conjunções coordenativas aditivas ou disjuntivas.

*O André comeu e bebeu bastante no último sábado.*

EVENT\_GROUPRELATION (comeu, bebeu, e)

*Ora chovia,ora vinha o sol.*

EVENT\_GROUPRELATION (chovia, vinha, ora)

As relações temporais, estabelecidas entre os eventos e as expressões temporais, são expressas através da dependência EVENT\_TIMERELATION. Esta dependência auxilia a ordenação de eventos ao complementar a informação temporal, aspetual ou modal dos eventos associados.

*Parece-me que ele acordou hoje maldispoto.*

EVENT\_TIMERELATION (acordou, hoje)

Para auxiliar a identificação das relações, consideraram-se ainda algumas dependências já extraídas pelo sistema, especificadas nas secções 2.3.4 e 2.3.5; são elas:

Identificação de eventos:

- EVENT\_LEX («evento», «tipo de evento»)
- EVENT\_TIME-CALENDAR («evento», «expressão temporal»)
- EVENT\_TIME-DURATION («evento», «expressão temporal»)
- EVENT\_TIME-FREQUENCY («evento», «expressão temporal»)
- EVENT\_DATE-START («evento», «expressão temporal»)
- EVENT\_DATE-END («evento», «expressão temporal»)

Dependências estabelecidas com os conectores existentes:

- INTROD («conector», «verbo»)
- QBOUNDARY («conector», «primeiro elemento da oração», «último elemento da oração»)
- CONNECTOR («verbo», «conector»)
- COORD («conector», «elemento coordenado»)
- PREPD («conector», «nome ou advérbio»)

Outras dependências relevantes:

- CDIR («verbo», «complemento direto»)
- CINDIR («verbo», «complemento indireto»)
- SUBJ («verbo», «sujeito»)
- MOD («verbo ou nome», «modificador»)
- PREDSUBJ («verbo», «predicativo do sujeito»)

## 4.2.2 Extração com base nas dependências existentes

A partir da combinação das dependências existentes é possível, na maioria dos casos, identificar as relações entre os eventos. Considere-se a seguinte frase com as dependências detetadas pelo XIP:

*A Rita saiu de casa depois do jantar.*

```
MOD_POST(saiu, jantar)
PREPD(jantar, depois de)
EVENT_LEX(saiu, outro)
EVENT_LEX(jantar, outro)
```

A dependência `MOD_POST(saiu, jantar)` identifica uma relação entre os dois eventos existentes na frase e, por sua vez, a dependência `PREPD(jantar, depois de)` identifica o modo como o evento *jantar* se relaciona, indicando-o como um evento subordinado, através da associação com a preposição *depois de*. Por conseguinte, é extraída a seguinte relação:

```
EVENT_INDIRECTRELATION(saiu, jantar, depois de)
```

O mesmo processo tem lugar ao identificar uma dependência do tipo `SUBJ` ou `CDIR` a ligar dois eventos. Nestes casos, porém, não há um conector a ligá-los, pelo que se extrai uma relação direta (`EVENT_DIRECTRELATION`).

*O trabalho publicado pelo Rui foi analisado.*

```
MOD_POST(trabalho, publicado)
SUBJ_PRE(analisado, trabalho)
EVENT_LEX(analisado, outros)
EVENT_LEX(publicado, outros)
EVENT_LEX(trabalho, outros)
EVENT_DIRECTRELATION(trabalho, publicado)
EVENT_DIRECTRELATION(analisado, trabalho)
```

A associação do evento subordinado com o conector pode igualmente fazer-se com base na dependência *INTROD* (de “introdutor”).

*Enquanto comia, o rapaz lia o livro.*

```
MOD_SENTENTIAL(lia, comia)
INTROD(Enquanto, comia)
EVENT_LEX(comia, outro)
EVENT_LEX(lia, outro)
EVENT_INDIRECTRELATION(lia, comia, enquanto)
```

Ainda relativamente às relações de subordinação, uma relação indireta pode também ser extraída a partir da associação das dependências *INTROD* e *CONNECTOR* através de um conector comum, como exemplificado.

*Ele esteve a comer enquanto estudava.*

```
INTROD(enquanto, estudava)
```

CONNECTOR (comer, enquanto)  
EVENT\_LEX (comer, outro)  
EVENT\_LEX (estudava, outro)  
EVENT\_INDIRECTRELATION (comer, estudava, enquanto)

Os agrupamentos (EVENT\_GROUPRELATION) são obtidos com o auxílio da dependência COORD, representando-se os eventos na ordem com que os mesmos aparecem na frase.

*O André chegou a casa, poisou a mochila e estendeu-se na cama.*

COORD (e, chegou)  
COORD (e, poisou)  
COORD (e, estendeu-se)  
EVENT\_LEX (chegou, outro)  
EVENT\_LEX (poisou, outro)  
EVENT\_LEX (estendeu-se, outro)  
EVENT\_GROUPRELATION (chegou, poisou, e)  
EVENT\_GROUPRELATION (poisou, estendeu-se, e)

Por último, as relações temporais (EVENT\_TIMERELATION) são extraídas a partir de dependências existentes, do tipo EVENT\_TIME e EVENT\_DATE, e das associações estabelecidas entre eventos e advérbios temporais, através das dependências do tipo MOD.

*Ele publicou tudo na sexta-feira.*

DETD (sexta-feira, a)  
PREPD (sexta-feira, em)  
MOD\_POST (publicou, sexta-feira)  
EVENT\_LEX (publicou, outro)  
EVENT\_TIME-CALENDAR (publicou, em a sexta-feira)  
EVENT\_TIMERELATION (publicou, em a sexta-feira)

Como exemplificado anteriormente, dependências distintas podem expressar associações semelhantes<sup>3</sup>, existindo o risco de se obter uma duplicação de relações. Para evitar estas situações, ignoram-se as dependências que expressem relações entre eventos e expressões temporais já relacionadas.

Note-se, também, que a representação da relação obtida com a dependência MOD é distinta da obtida através da dependência EVENT\_TIME-CALENDAR. Isto deve-se ao facto de a dependência MOD ser de natureza sintática e operar entre o verbo e a cabeça (representada pela dependência HEAD) de um complemento preposicional ou adverbial.

Em relação aos agrupamentos, há ainda outras relações que podem ser extraídas além das encontradas. Devido às propriedades das conjunções aditivas e disjuntivas, analisadas na Secção 3.2, os eventos relacionados pertencem ao mesmo instante temporal, a menos que seja especificado o contrário. Deste modo, se apenas um dos eventos agrupados associado a expressões temporais, as relações temporais são replicadas, como no exemplo que se segue:

---

<sup>3</sup>Trata-se de dependências que foram definidas em diferentes fases de desenvolvimento do sistema e que ainda não foram harmonizados.

**O Rui comeu e bebeu bastante no último sábado.**

COORD (e, comeu)

COORD (e, bebeu)

MOD (bebeu, em o último sábado)

EVENT\_LEX (comeu, outro)

EVENT\_LEX (bebeu, outro)

EVENT\_TIME-CALENDAR (bebeu, em o último sábado)

EVENT\_GROUPRELATION (comeu, bebeu, e)

EVENT\_TIMERELATION (comeu, em o último sábado)

EVENT\_TIMERELATION (bebeu, em o último sábado)

## 4.3 Ordenação de Eventos

### 4.3.1 Dependências

Para a ordenação dos eventos, considerou-se que um evento pode ser *anterior*, *simultâneo* ou *posterior* a outro. Quando um evento é posterior a um segundo, a ordem identificada é a inversa, ou seja, o segundo evento antecede o primeiro. Por conseguinte, criaram-se apenas duas dependências para a ordenação de eventos, especificadas em detalhe nas diretivas:

```
EVENT_ORDERBEFORE («evento anterior», «evento posterior»)
```

```
EVENT_ORDERSIMULT («evento principal», «evento subordinado/coordenado»)
```

Considerou-se também a existência de relações de interseção e inclusão, entre outros, como as existentes na lógica temporal de intervalos [2], definindo-se que:

- um evento antecede outro se foi concluído antes de o segundo se iniciar, ou seja, nas condições expressas na Figura 4.4;
- um evento é simultâneo a outro se ocorre em simultâneo, interseta ou é incluído pelo segundo, como representado na Figura 4.5.

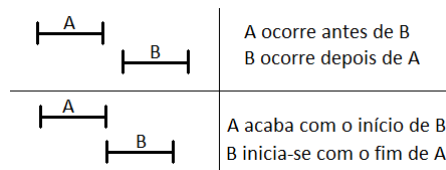


Figura 4.4: Condições necessárias para o evento A ser anterior ao B.

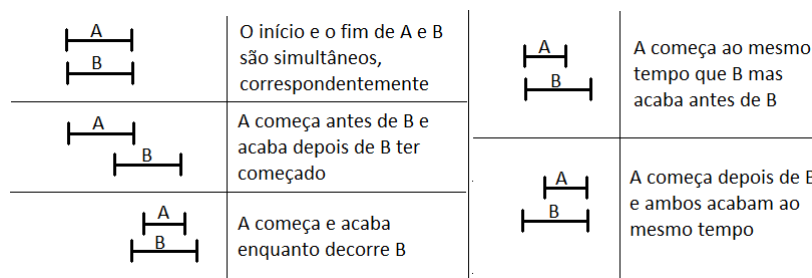


Figura 4.5: Condições necessárias para dois eventos serem simultâneos entre si.

Na análise temporal e aspetual dos eventos, é previsível que nem toda a informação esteja disponível. Assim, quando apenas é referido a conclusão de um evento (aspeto terminativo), considera-se que o seu início ocorre num passado distante; inversamente, quando apenas se identifica o início do evento (aspeto incoativo/inceptivo), a sua conclusão é deixada indefinida (ver Figura 4.6).

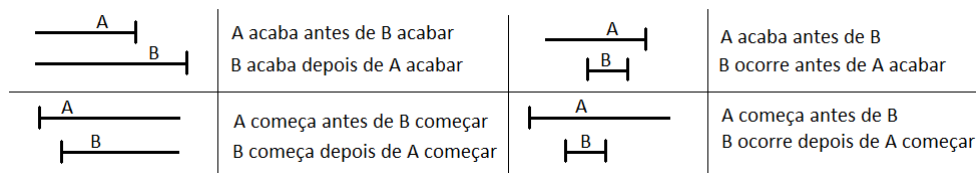


Figura 4.6: Relações adicionais a considerar em relação às fronteiras de início e fim de eventos.

### 4.3.2 Análise dos Eventos

Para auxiliar a ordenação dos eventos, pode-se recorrer à informação temporal, aspetual e modal associada aos eventos relacionados. Este tipo de informação pode provir do verbo a partir do qual se identificou o evento, dos verbos auxiliares com que este se constrói ou das expressões temporais associadas a partir da dependência `EVENT_TIMERELATION`.

Quando possível, para cada evento, extraem-se e analisam-se os traços temporais e aspetuais existentes, não se decidindo nada caso existam contradições, pressupondo-se, nessas situações, que a informação encontrada não é fiável.

Na ausência de traços aspetuais, a ordenação baseia-se apenas na informação temporal: eventos futuros são posteriores a eventos com ocorrência no passado ou presente; eventos passados são anteriores a eventos com ocorrência no presente ou futuro. Caso contrário, considera-se a possibilidade de se encontrar interseções entre os eventos relacionados, como representado na Figura 4.6, na secção anterior. Se a execução de um evento se intersejar ou for incluído por outro, os eventos passam a ser considerados simultâneos.

Para a comparação entre eventos identificados a partir de verbos, definiu-se também, com base na análise dos mesmos, apresentada na Secção 3.1.2, que:

- a conjugação do verbo no gerúndio representa simultaneidade com o evento relacionado<sup>4</sup>;

*Relendo o texto, o professor esclarece o aluno.*

`EVENT_DIRECTRELATION(esclarece, relendo)`

`EVENT_ORDERSIMULT(esclarece, relendo)`

- um verbo conjugado no particípio passado, nas chamadas orações reduzidas, expressa um evento no passado e com um aspeto perfetivo, sendo considerado anterior a qualquer evento presente ou futuro ou identificado a partir de verbos conjugados no pretérito simples.

*Revisto o problema, o rapaz vai pensar na solução mais apropriada.*

`EVENT_DIRECTRELATION(Revisto, pensar)`

`EVENT_ORDERBEFORE(Revisto, pensar)`

*Revisto o problema, o rapaz pensou na solução mais apropriada.*

`EVENT_DIRECTRELATION(Revisto, pensou)`

`EVENT_ORDERBEFORE(Revisto, pensou)`

- os verbos conjugados no pretérito perfeito e no mais-que-perfeito expressam eventos iniciados e concluídos no passado, sendo anteriores a eventos presentes ou futuros. Adicionalmente, o pretérito mais-que-perfeito localiza o evento como anterior a eventos expressos por verbos conjugados no pretérito perfeito ou imperfeito.

<sup>4</sup>Tal decisão ignora os casos em que o gerúndio tem um valor *causal*, que implica uma ordenação como *anterior* ao evento da oração principal.

*A noiva já chegara à igreja quando o noivo apareceu.*

EVENT\_INDIRECTRELATION (chegara, apareceu, quando)

EVENT\_ORDERBEFORE (chegara, apareceu)

- a conjugação do verbo no futuro simples expressa um evento iniciado num instante futuro, sendo posterior a qualquer evento que ocorra ou termine no passado ou presente.

### 4.3.3 Análise dos Conectores

Analisando os atributos dos conectores, quando existentes, e dos eventos relacionados, é possível, na maioria dos casos, definir a ordem mais apropriada a atribuir. Relativamente aos conectores, analisa-se a função que desempenham, analisando os atributos respetivos, e.g. *CAUSAL*, *PROPORTIONAL*, *CONDITIONAL*, *CONCESSIVE*, entre outros.

Estabelece-se que um evento subordinante é anterior ao subordinado se o conector tem um valor *comparativo*, *final*, *consecutivo* ou *proporcional*.

*Ele comia como um esfomeado.*

EVENT\_INDIRECTRELATION (comia, esfomeado, como)

EVENT\_ORDERBEFORE (comia, esfomeado)

*Ele esteve a rever o livro para aprofundar o conhecimento.*

EVENT\_INDIRECTRELATION (rever, aprofundar, para)

EVENT\_ORDERBEFORE (rever, aprofundar)

*Falou tanto na reunião que ficou rouco.*

EVENT\_INDIRECTRELATION (falou, ficou, tanto que)

EVENT\_ORDERBEFORE (falou, ficou)

*Ele foi falando cada vez mais baixo à medida que ia enrouquecendo.*

EVENT\_INDIRECTRELATION (falando, enrouquecendo, a a medida que)

EVENT\_ORDERBEFORE (falando, enrouquecendo)

Por outro lado, se o conector tiver um valor *condicional*, *causal*, *concessivo* ou de *tópico*, o evento subordinante é posterior ao evento subordinado. Nesta situação, extrai-se uma ordem de anterior, com a ordem dos eventos invertida em relação à ordem linear que aparecem no discurso (quando não há anteposição da subordinada).

*Se esperares por mim, eu vou aí ter contigo.*

EVENT\_INDIRECTRELATION (vou, esperares, se)

EVENT\_ORDERBEFORE (esperares, vou)

*Ele adormeceu porque estava cansado.*

EVENT\_INDIRECTRELATION (adormeceu, cansado, porque)

EVENT\_ORDERBEFORE (cansado, adormeceu)

*Embora ele não mereça, eu dou-lhe uma última oportunidade.*

EVENT\_INDIRECTRELATION (dou-lhe, mereça, embora)

EVENT\_ORDERBEFORE (mereça, dou-lhe)



*Estive a ler um livro sobre a 1ª Guerra Mundial*<sup>5</sup>.

```
EVENT_INDIRECTRELATION(Estive a ler,1ª Guerra Mundial,sobre)
EVENT_ORDERBEFORE(1ª Guerra Mundial,Estive a ler)
```

Por fim, os conectores temporais podem exprimir qualquer um dos tipos de ordem, com base no seu valor semântico: *anterior*; *simultâneo* ou *posterior*.

*Cheguei a casa muito antes de a festa dos meus tios começar.*

```
EVENT_INDIRECTRELATION(Cheguei,festa,muito antes de)
EVENT_ORDERBEFORE(Cheguei,festa)
```

*Ela sorriu quando me viu.*

```
EVENT_INDIRECTRELATION(sorriu,viu,quando)
EVENT_ORDERSIMULT(sorriu,viu)
```

*Implicou comigo assim que me viu.*

```
EVENT_INDIRECTRELATION(Implicou,viu,assim que)
EVENT_ORDERBEFORE(viu,Implicou)
```

Para auxiliar a ordenação com base nas conjunções e preposições temporais, associaram-se aos conectores, no léxico do XIP, os traços que definem o tipo de ordenação expressa pelos mesmos.

#### 4.3.4 Análise das Dependências

Para a dependências do tipo `EVENT_GROUPRELATION`, ordenam-se os eventos segundo a informação associada aos mesmos. Quando não for atribuída uma ordem com base nos eventos associados, os mesmos são considerados simultâneos.

Relativamente às dependências de subordinação, `EVENT_INDIRECTRELATION` e `EVENT_DIRECTRELATION`, dá-se prioridade à informação obtida a partir dos advérbios temporais e eventos, de seguida à dos conectores, se existirem, e, por fim, caso não se encontre uma ordem a atribuir, verifica-se o tipo de eventos envolvidos, definindo-se que:

- Se o evento subordinado foi identificado a partir de um nome predicativo, o mesmo é considerado anterior.

*Ele entregou o trabalho.*

```
EVENT_DIRECTRELATION(entregou,trabalho)
EVENT_ORDERBEFORE(trabalho,entregou)
```

*Ele fez a entrega do trabalho.*

```
EVENT_INDIRECTRELATION(entrega,trabalho, de)
EVENT_ORDERBEFORE(trabalho,entrega)
```

#### 4.3.5 Representação gráfica da ordenação

É possível representar graficamente a ordenação de eventos através de uma linha cronológica ou através de grafos. Existem vantagens e desvantagens a considerar em cada um dos métodos, sendo necessário escolher o mais apropriado ao objetivo a alcançar.

---

<sup>5</sup>Em rigor, o complemento *sobre a 1ª Guerra Mundial* é um complemento de *livro* e não de *ler*. Este problema, conhecido como o problema do *PP-attachment* ainda não está suficientemente bem tratado na cadeia `STRING`.

O uso de uma linha cronológica permite associar cada um dos eventos aos instantes temporais em que ocorrem. Deste modo, é possível identificar a ordem de ocorrência de cada evento em relação aos restantes (Figura 4.7).

*O Reino de Portugal estabeleceu-se em 1139, sendo reconhecido como independente em 1143. A monarquia só terminou com a Revolução de 1910.*

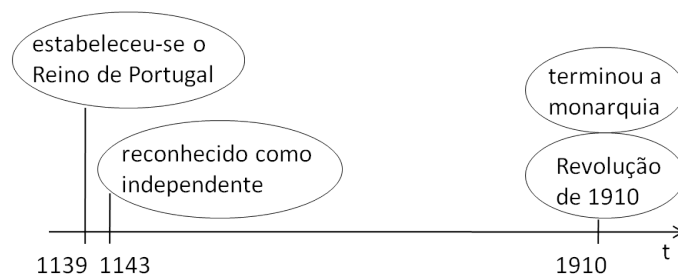


Figura 4.7: Representação de eventos através de uma linha cronológica.

No entanto, o uso deste método é desadequado quando não existem referências temporais que permitam a especificação do instante de ocorrência dos eventos. Nesse tipo de situações, como exemplificado nas frases que se seguem e na Figura 4.8, será necessário estimar a posição cronológica dos mesmos.

1. *Iremos preparar a festa de despedida no dia 30 de Março.*
2. *Ele adormeceu porque estava cansado.*

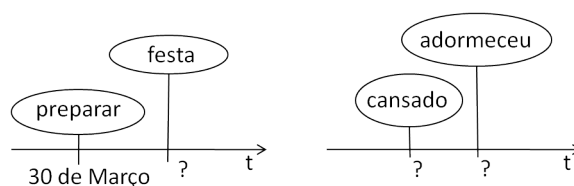


Figura 4.8: Representação da ordem dos eventos através de uma linha cronológica.

Outro problema levantado por este tipo de representação é a definição da escala, tendo esta de se manter coerente e apropriada ao tipo de textos em análise. O uso de grafos torna-se mais adequado, contornando este tipo de problemas, ao permitir a representação da ordem relativa entre os eventos, como exemplificado para as frases usadas nos exemplos anteriores (Figura 4.9).



Figura 4.9: Representação da ordem dos eventos através de um grafo.

No entanto, há também desvantagens em usar este tipo de representação: a interpretação do grafo torna-se mais difícil na presença de uma grande quantidade de eventos e, conseqüentemente, de relações. Assim, é necessário definir até que ponto se pretende especificar a informação obtida, de modo a evitar o aumento da complexidade. A Figura 4.10 representa apenas as relações de ordem entre os eventos que se associam diretamente. A representação da ordem de cada evento com todos os restantes resultaria num grafo, para a mesma frase, com 45 ligações distintas.

Selecionou-se este segundo método de representação porque o processo de ordenação se baseou bastante nas relações entre os eventos e, conseqüentemente, na ordenação relativa dos mesmos. No entanto, decidiu-se manter a informação temporal no grafo, quando existente.

Devido à presença desta população centenária, o município local decidiu iniciar um projecto para analisar a sua dieta alimentar e chegar a dados que ajudem a explicar a sua longevidade.

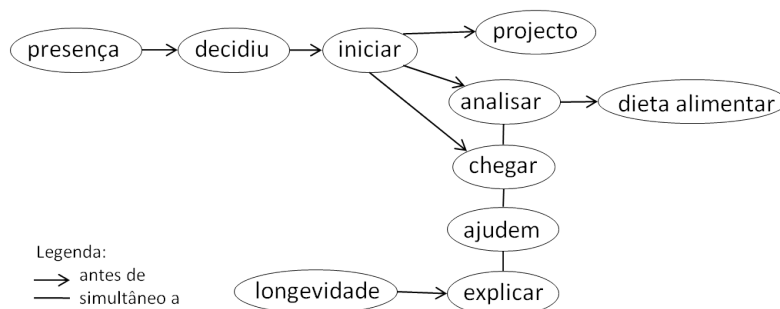


Figura 4.10: Exemplo complexo da representação da ordem dos eventos através de um grafo.

Estabeleceu-se que os eventos e as expressões temporais são representados por nós distintos (elipse e rectângulo, respetivamente) e que as associações são representadas por ligações diferentes, conforme o significado, à semelhança do exemplo anterior. O grafo de ordenação é construído diretamente a partir das dependências de ordem e das relações temporais obtidas.

Como exemplo, considere-se a frase que se segue, onde se está perante duas relações de ordem. Como representado na Figura 4.11, ainda que seja possível deduzir que o evento *foi* antecede o evento *festa*, não se representa essa relação.

*Ele foi a casa nessa tarde para se preparar para a festa dessa noite.*

```
EVENT_LEX(foi,outro)
EVENT_LEX(preparar,outro)
EVENT_LEX(festa,outro)
EVENT_TIMERELATION(foi,em essa tarde)
EVENT_TIMERELATION(festa,de essa noite)
EVENT_INDIRECTRELATION(foi,preparar,para)
EVENT_INDIRECTRELATION(preparar,festa,para)
EVENT_ORDERBEFORE(foi,preparar)
EVENT_ORDERBEFORE(preparar,festa)
```



Figura 4.11: Exemplo de representação da ordem de eventos e expressões temporais através de um grafo.

## 4.4 Implementação

Inicia-se a criação da *árvore de relações* ao percorrer a frase e extrair os conectores e os eventos. Os conectores são extraídos ao se identificar as palavras e expressões com os traços correspondentes a uma conjunção (*CONJ*), uma preposição (*PREP*) ou um advérbio conjuntivo (*ADVCONJ*). Os eventos são extraídos através das dependências do tipo *EVENT\_LEX* («evento», «tipo de evento»), criando-se um nó para cada evento. Considere-se a frase que se segue como exemplo:

*A Rita vai lanchar assim que acabar de ler.*

A *STRING* analisa a frase, etiquetando os vários elementos que a compõem, obtendo-se a estrutura representada pela Figura 4.12. Entre os principais traços atribuídos aos nós da árvore identifica-se uma conjunção, *assim que*, e uma preposição, *de*, marcadas respetivamente por *CONJ* e *PREP*. A preposição *de*, no entanto, como verificado adiante, não será considerada, apesar de se tratar de um conector, por não permitir a extração de relações entre eventos.

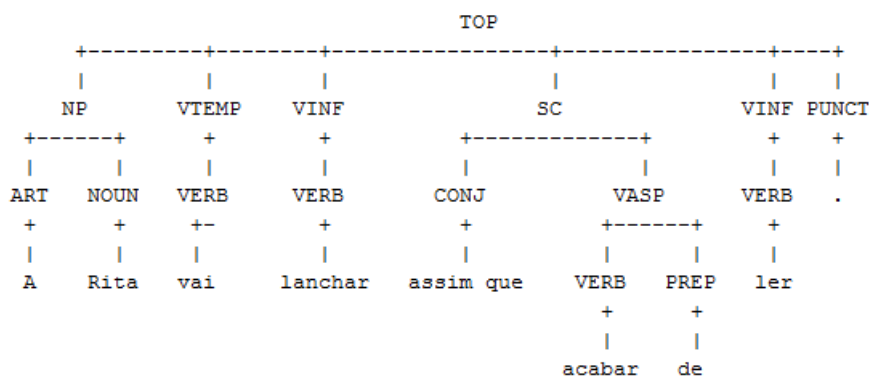


Figura 4.12: Exemplo da árvore sintática gerada pela *STRING*.

Ao se analisar a informação lexical associada aos nós da árvore sintática, extraem-se dois eventos, *lanchar* e *ler*:

*A Rita vai lanchar assim que acabar de ler.*

`EVENT_LEX(lanchar, outro)`

`EVENT_LEX(ler, outro)`

Durante a próxima fase, analisam-se as dependências com os eventos e/ou conectores identificados e procede-se à extração de relações para se obter a árvore de relações. Existem diversos tipos de dependências, descritos previamente na Secção 4.2.1, permitindo a identificação de:

1. uma relação entre dois eventos, mas sem se identificar o conector  
(e.g: MOD («verbo ou nome», «modificador»));
2. uma relação entre um evento coordenado e um conector  
(e.g: COORD («conector», «elemento coordenado»));
3. uma relação entre um evento subordinado e um conector, sem se identificar o evento subordinante  
(e.g: PREPD («conector», «nome ou advérbio»));
4. uma relação entre um evento subordinante e um conector, sem se identificar o evento subordinado  
(e.g: CONNECTOR («verbo», «conector»));

#### 5. uma relação entre um evento e uma expressão temporal

(e.g: EVENT\_TIME-CALENDAR(«evento», «expressão temporal»)).

A extração de relações é feita em duas iterações. Na primeira iteração, analisam-se as situações 1, 2 e 3.

No primeiro caso, quando se identifica uma associação entre dois eventos, cria-se uma relação sem se especificar o conector. Ao se identificar o conector associado ao evento subordinado, no segundo caso, associa-se o conector ao nó que representa o evento. Se for possível associar um conector a uma relação entre eventos, trata-se de uma relação do tipo EVENT\_INDIRECTRELATION, caso contrário, considera-se que se identificou uma dependência do tipo EVENT\_DIRECTRELATION.

Na terceira situação, quando se identifica um evento coordenado através de uma dependência (e.g: COORD), é verificado se já foi criado algum grupo identificado pelo conector. Se existir, adiciona-se o evento ao grupo, caso contrário, cria-se o nó que representa o grupo e inclui-se o evento. Para obter as dependências respectivas, EVENT\_GROUPRELATION, itera-se pelos eventos do grupo, associando cada um deles ao conector e o evento seguinte.

Em relação à frase utilizada como exemplo, obtêm-se algumas dependências consideradas relevantes para a extração de relações. No entanto, como se pode verificar, nenhuma das dependências associa diretamente um evento a outro:

*A Rita vai lanchar assim que acabar de ler.*

SUBJ\_PRE(lanchar, Rita)

SUBJ\_PRE\_ANAPH0(ler, Rita)

CONNECTOR(lanchar, assim que)

QBOUNDARY\_TEMPORAL(assim que, assim que, ler)

Como referido, nesta primeira fase, analisam-se apenas as dependências que associam dois eventos ou um evento coordenado/subordinado a um conector. As dependências SUBJ\_PRE(lanchar, Rita) e SUBJ\_PRE(lanchar, Rita) são, por conseguinte, descartadas. A dependência CONNECTOR(lanchar, assim que) associa um evento, *lanchar*, a um conector, *assim que*. Apesar desta associação ser relevante para a extração de relações, como a dependência representa uma associação de um evento subordinante com um conector, é descartada nesta fase. A dependência QBOUNDARY\_TEMPORAL(assim que, assim que, ler), por sua vez, já associa uma conjunção a um evento subordinado. Assim, associa-se a conjunção ao nó do evento *ler*.

Na segunda iteração, analisam-se as situações 4 e 5. Ao se identificar uma relação entre um evento subordinante e um conector, na situação 4, se existir algum evento subordinado com o conector associado, ligam-se os dois eventos. Caso contrário, essa dependência é ignorada. Na situação 5, como resultado, associam-se as expressões temporais aos nós dos eventos, extraindo-se as dependências do tipo EVENT\_TIMERELATION.

Recorde-se que, no exemplo anterior, foi ignorada a dependência CONNECTOR(lanchar, assim que). Durante a primeira iteração, não havia garantia que a conjunção *assim que* associava dois eventos. Ao se voltar a encontrar a dependência, nesta segunda iteração, CONNECTOR(lanchar, assim que) liga um evento subordinante a uma conjunção já previamente associada a um nó da árvore de relações, através do evento *ler*. Como resultado, obtêm-se a seguinte relação:

EVENT\_INDIRECTRELATION(lanchar, ler, assim que)

Não existem mais relações com eventos, nem com conectores ou expressões temporais, pelo que a árvore de relações é considerada terminada. Para se ter apenas uma árvore única, ao se completar a extração das relações, cria-se um nó central que agrupa as raízes das relações não interligadas. Assim, como resultado da análise da frase usada, obtém-se a árvore representada pela Figura 4.13.

Prossegue-se para a ordenação dos eventos, percorrendo a árvore a partir da raiz e marcando os nós como visitados à medida que se analisa cada relação, comparando o nó pai com um nó filho. Para cada relação, analisam-se os eventos, de seguida os conectores e, por fim, o tipo de relação extraída, extraindo a ordem entre os eventos quando possível. Cada relação

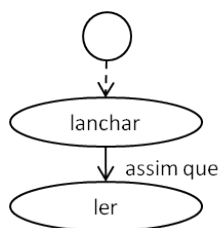


Figura 4.13: Exemplo da árvore de relações gerado pelo módulo de ordenação de eventos.

de ordem extraída corresponde a uma dependência do tipo `EVENT_ORDERBEFORE` ou `EVENT_ORDERSIMULT` e a uma ligação entre os eventos no grafo de ordenação resultante.

Na árvore de relações extraída no exemplo usado, tem-se uma relação do tipo `EVENT_INDIRECTRELATION` a associar dois eventos extraídos a partir de verbos, sendo possível identificar, pela análise sintática realizada previamente, as respectivas cadeias verbais *vai lanchar* e *acabar de ler*. É de notar que a `STRING` atribui bastantes traços que não são representados na árvore sintática. No exemplo, a cadeia verbal *vai lanchar* tem os traços `VERB`, `VTEMP` e `FUTU`, entre outros. Esses traços permitem a classificação do evento respetivo como sendo identificado a partir de um verbo (`VERB`), acompanhado de um verbo auxiliar temporal (`VTEMP`) com o valor de futuro (`FUTU`). A cadeia verbal *acabar de ler*, por sua vez, tem os traços `VERB`, `VASP` e `TERMINAT`, permitindo a classificação do evento *ler* como sendo identificado a partir de um verbo (`VERB`), acompanhado de um verbo auxiliar de valor aspetual (`VASP`) terminativo (`TERMINAT`). Assim, sabe-se que o evento *lanchar* tem a sua ocorrência no futuro, mas não se sabe o instante de ocorrência do evento *ler*. A análise dos eventos é insuficiente para a respetiva ordenação, procedendo-se à análise do conector.

Ao analisar o conector *assim que*, identificam-se os traços `CONJ`, `TEMPORAL` e `T-REF-AFTER`, entre outros, classificando-o como uma conjunção (`CONJ`) temporal (`TEMPORAL`) que expressa uma ordem posterior (`T-REF-AFTER`), ou seja, que o elemento subordinante ocorre após o elemento subordinado, neste caso, os eventos *lanchar* e *ler*. Assim, extrai-se a dependência:

```
EVENT_ORDERBEFORE(ler, lanchar)
```

Por fim, controli-se o grafo de ordenação de eventos com base nas dependências de ordem, obtendo-se a estrutura representada na Figura 4.14.

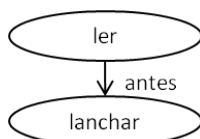


Figura 4.14: Exemplo do grafo de ordem gerado pelo módulo de ordenação de eventos.

# Capítulo 5

## Avaliação

Neste capítulo, descreve-se o corpus linguístico e a metodologia utilizada para a avaliação do sistema desenvolvido no âmbito desta dissertação. À semelhança de outros sistemas, realizaram-se também dois tipos de avaliação, uma avaliação *estrita* e *relaxada*, e consideraram-se as métricas *abrangência*, *precisão* e *medida-f*. Além disso, também se decompôs a avaliação da solução em várias fases, fazendo a diferenciação entre a avaliação da identificação dos eventos, da extração de relações e da ordenação dos eventos. Por conseguinte, foi possível analisar melhor o desempenho de cada fase da resolução da tarefa.

Apresentam-se, igualmente, os resultados obtidos, no conjunto e por cada fase da resolução do problema, analisando-se o desempenho do módulo desenvolvido em relação às metas definidas e aos sistemas descritos ao longo deste trabalho.

### 5.1 Anotação

Este trabalho não recorreu a nenhum método de aprendizagem automática, não sendo necessário a construção de um corpus para treinar o sistema. No entanto, para avaliar o sistema, construiu-se um corpus a partir da junção de vários excertos de natureza narrativa e tópicos atuais diversificados, tendo também em vista a abundância de eventos e de expressões temporais. Naturalmente, este pequeno corpus não pretende ser representativo de qualquer variedade ou uso da língua em particular: a sua função foi apenas de fornecer material para a avaliação do sistema com textos reais a fim de se aferir a dificuldade e os problemas da tarefa.

Os excertos foram extraídos a partir de artigos, contos e diversas notícias, entre os quais se inclui um excerto do conto *Adão e Eva no Paraíso*, de Eça de Queirós (1897), um artigo biográfico de Fernando Pessoa, extraído da Wikipédia<sup>1</sup>, e diversas notícias de divulgação científica (ver Anexo A).

O corpus foi anotado manualmente por dois linguistas, de acordo com as diretivas desenvolvidas ao longo desta dissertação (ver Anexo B), identificando-se cerca de 100 expressões temporais, 700 eventos e 700 relações, num total de quase 200 frases e 4.500 palavras (Tabela 5.1).

### 5.2 Métodos de Avaliação e Técnicas Utilizadas

Antes de se proceder à avaliação, é necessário considerar que a tarefa de ordenação de eventos se subdivide em várias subtarefas distintas (identificação de eventos, análise dos eventos, análise dos conectores, identificação das relações, etc.) até

---

<sup>1</sup>pt.wikipedia.org

	Frases	Palavras	Expr. Temp.	Eventos	Rel. Eventos	Rel. Temp.	Ord. Eventos
jornalístico	109	2906	77	535	373	65	385
literário	43	1301	18	190	147	27	178
enciclopédico	10	216	6	47	30	6	44
Total	162	4423	101	772	550	98	607

Tabela 5.1: Detalhes do corpus de teste utilizado para avaliação do sistema.

se proceder à respetiva ordenação, sendo possível e preferível fazer uma análise mais pormenorizada de cada etapa em vez de se avaliar o processo apenas como um todo.

Assim sendo, subdividiu-se o processo de avaliação para se realizar uma análise mais rigorosa do módulo aqui desenvolvido. Descrevem-se de seguida os tipos de avaliação, as diversas fases em que se subdividem e os parâmetros considerados na avaliação do sistema.

### 5.2.1 Avaliação Estrita e Relaxada

Analisando o problema de ordenação de eventos, é possível subdividir este processo nas seguintes tarefas:

1. Identificação e classificação de eventos
2. Identificação e classificação de relações entre eventos
3. Identificação e classificação de relações temporais
4. Ordenação de eventos

Note-se que existem bastantes tarefas envolvidas, existindo ainda uma distinção entre as tarefas de identificação e as de classificação. Uma solução é mal classificada se a dependência encontrada foi mal representada, existindo erros na representação dos elementos, na ordem dos elementos encontrados ou no tipo de dependência extraída. Esta diversidade de subtarefas e situações a avaliar possibilita a realização de dois tipos de avaliação distintos: uma avaliação *estrita* e outra *relaxada*.

A *avaliação estrita* avalia a solução em relação ao resultado ideal, ou seja, considera uma solução como correta apenas se o resultado for exatamente igual ao esperado.

A *avaliação relaxada*, por sua vez, foca-se na identificação da ordem dos eventos, tolerando alguns erros nos resultados intermédios ou secundários à ordenação, ou até mesmo erros de representação dos resultados finais. Mais especificamente, a avaliação relaxada ignora:

- erros na classificação dos eventos nas dependências:
  - tipo de evento incorreto, e.g:
    - \* `EVENT_LEX(lia,afiliação)` em vez de `EVENT_LEX(lia,outro)`.
  - representação incorreta do evento, e.g:
    - \* `EVENT_LEX(lia-se,outro)` em vez de `EVENT_LEX(lia,outro)`;
    - \* `EVENT_DIRECTRELATION(lia-se,trabalho)` em vez de `EVENT_DIRECTRELATION(lia,trabalho)`;
    - \* `EVENT_ORDERSIMULT(revelou-se,paga-se)` em vez de `EVENT_ORDERSIMULT(revelou,paga)`.



- relações mal classificadas ou com o conector incorreto:
  - ordem incorreta na representação dos eventos relacionados
    - \* `EVENT_DIRECTRELATION(lia, trabalho)` em vez de `EVENT_DIRECTRELATION(trabalho, lia)`
  - identificação incorreta do conector associado aos eventos, e.g:
    - \* `EVENT_INDIRECTRELATION(aumentado, ritmo, a)` em vez de `EVENT_INDIRECTRELATION(aumentado, ritmo, mas)`
  - tipo de relação identificado incorretamente, e.g:
    - \* `EVENT_DIRECTRELATION(revelou, paga)` em vez de `EVENT_INDIRECTRELATION(revelou, paga, que)`
  - a inversão dos eventos na dependência `EVENT_ORDERSIMULT`
    - \* `EVENT_ORDERSIMULT(revelou, paga)` em vez de `EVENT_ORDERSIMULT(paga, revelou)`

## 5.2.2 Métricas Utilizadas

Na avaliação do sistema, podendo esta ser uma avaliação estrita ou relaxada (Secção 5.2.1), um determinado resultado pode ser considerado *correto*, *espúrio* ou *em falta*.

Um resultado é *correto* (verdadeiro positivo) se cumprir os requisitos definidos pela avaliação realizada. Um resultado é considerado *espúrio*, ou falso positivo, se não existir uma correspondência ao mesmo no corpus de teste. Qualquer resultado esperado não encontrado, por sua vez, é considerado *em falta* (falso negativo).

A partir da combinação destes parâmetros obtêm-se as medidas de *abrangência*, *precisão* e *medida-f*:

- A *abrangência* mede a percentagem de entidades que o sistema conseguiu identificar corretamente. Este valor pode ser obtido a partir da quantidade de resultados *corretos* e *em falta* (eq. 5.1).

$$\text{abrangência} = \frac{\text{bem identificados}}{\text{existentes}} = \frac{n^{\circ} \text{ corretos}}{n^{\circ} \text{ em falta} + n^{\circ} \text{ corretos}} \quad (5.1)$$

- A *precisão* mede a proporção de respostas corretas em relação às respostas fornecidas pelo sistema e é obtida ao combinar o número de resultados corretos e os resultados espúrios (eq. 5.2).

$$\text{precisão} = \frac{\text{bem identificados}}{\text{identificados}} = \frac{n^{\circ} \text{ corretos}}{n^{\circ} \text{ espúrio} + n^{\circ} \text{ corretos}} \quad (5.2)$$

- A *medida-f* corresponde à média harmónica da abrangência e da precisão (eq. 5.3).

$$\text{medida-f} = \frac{2 * \text{precisão} * \text{abrangência}}{\text{precisão} + \text{abrangência}} \quad (5.3)$$

Recorre-se ainda a uma matriz de confusão para analisar os erros na classificação das dependências de relação e nas relações de ordem.

### 5.2.3 Delimitação das tarefas a avaliar

Como descrito previamente, o processo de ordenação de eventos pode ser decomposto em diversas tarefas, cujo resultado depende da conclusão das anteriores: é necessário identificar os eventos existentes antes de ser possível relacioná-los; procede-se à ordenação dos eventos com base na análise das relações e dos respetivos eventos.

Assim, após a identificação incorreta de alguns eventos, é provável que se encontre erros ao extrair algumas relações. Do mesmo modo, a presença de erros na identificação e relação de eventos prejudica a ordenação dos mesmos. No exemplo que se segue, extraiu-se a dependência `EVENT_LEX(costuma)` apesar de *costuma* não ser considerado um evento<sup>2</sup>. Como consequência, *costuma* associa-se ao evento *rever*, extraíndo-se uma relação que depois é analisada durante a ordenação de eventos, obtendo-se mais um erro na solução.

*Ele costuma rever as fotografias de antigamente.*

```
EVENT_LEX(rever, outro)
EVENT_LEX(fotografias, outro)
*EVENT_LEX(costuma, outro)
*EVENT_DIRECTRELATION(costuma, rever)
EVENT_DIRECTRELATION(rever, fotografias)
EVENT_TIMERELATION(fotografias, de antigamente)
*EVENT_ORDERBEFORE(costuma, rever)
EVENT_ORDERBEFORE(fotografias, rever)
```

Para avaliar individualmente as tarefas de identificação, relação e ordenação de eventos, o sistema de avaliação realiza duas iterações sobre os resultados obtidos após a avaliação da solução:

- A primeira iteração identifica os eventos considerados incorretos e remove os resultados relativos às relações e ordenações associadas a esses eventos;
- A segunda iteração remove os resultados obtidos, na ordenação de eventos, a partir das relações mal identificadas.

A Tabela 5.2 exemplifica este processo, exibindo os resultados antes e após cada iteração do sistema de avaliação.

Como se pode verificar através da Tabela, as iterações filtram os resultados que são consequências diretas de resultados obtidos incorretamente na tarefa anterior, independentemente de a avaliação desses resultados ser positiva ou negativa. Isto é, uma dependência pode ser extraída corretamente a partir de dependências erradas, sendo igualmente removida. Por exemplo, na frase que se segue, a relação extraída está incorreta, mas ainda foi possível obter a ordem esperada:

*A carne bovina também tem aumentado na China, mas a um ritmo mais lento.*

```
EVENT_LEX(aumentado, outro)
EVENT_LEX(ritmo, outro)
*EVENT_INDIRECTRELATION(aumentado, ritmo, a)
(resultado esperado: EVENT_INDIRECTRELATION(aumentado, ritmo, mas))
EVENT_ORDERSIMULT(aumentado, ritmo)
```

<sup>2</sup>A cadeia `STRING` considera *costumar* um verbo auxiliar de valor aspetual, e não como um evento, como alguns dos sistemas apresentados neste documento. Deste modo, *costuma* não deverá ser capturada pela dependência `EVENT_LEX`, ao contrário do exemplo descrito neste documento.

Tarefas	Avaliação	Valores Iniciais	Após 1ª iteração	Após 2ª iteração
Identificação	corretos	528	528	528
	espúrios	151	0	0
	em falta	244	0	0
Relação Temporal	corretos	63	63	63
	espúrios	87	7	0
	em falta	485	178	0
Relação entre Eventos	corretos	103	103	103
	espúrios	127	37	0
	em falta	525	228	0
Ordenação	corretos	43	43	42
	espúrios	158	89	82
	em falta	560	299	154

Tabela 5.2: Exemplo das iterações realizadas pelo sistema de avaliação sobre resultados obtidos do processo de avaliação

### 5.3 Resultados

Avaliando a solução com base nos critérios da avaliação estrita, ou seja, sem tolerância a erros, obtiveram-se os resultados descritos na Tabela 5.3. A partir dos mesmos, calcularam-se os valores das métricas de avaliação para a ordenação de eventos como um todo e para cada uma das tarefas em separado (ver Tabela 5.4).

Realizou-se igualmente a avaliação relaxada da solução, obtendo-se os valores discriminados na Tabela 5.5. Em relação às métricas, na Tabela 5.6, verificaram-se que os resultados da ordenação de eventos, no seu total, foram iguais aos obtidos na avaliação estrita. No entanto, comparando os resultados, na tarefa de relação entre eventos obteve-se uma melhoria de 17% e 13,8% nos valores de *abrangência* e *precisão*. Para a tarefa de ordenação, ainda que pequena, obteve-se uma variação no sentido oposto: perda de *abrangência* (0,6%) e de *precisão* (0,8%).

Compararam-se ainda os erros entre os valores devolvidos e os esperados pela solução, discriminando-os através da matriz de confusão apresentada na Tabela 5.7. Estes resultados mostram que existe uma clara distinção entre as tarefas que definem o processo de ordenação de eventos. Por outras palavras, não foi definida nenhuma dependência de ordem em vez de uma dependência de relação e nenhuma relação temporal no lugar de uma relação de eventos. Também demonstram que existiu alguma confusão entre a extração de uma relação direta ou indirecta/agrupamento. Estas situações indicam problemas na identificação dos conectores, dada a relação direta ser extraída quando o mesmo não é identificado apesar de extrair uma relação. A extração da relação de agrupamento ao invés da relação indirecta, por sua vez, indica uma identificação ou interpretação incorreta do conector que associa os eventos. Em relação à ordenação de eventos, é preciso notar que a extração incorreta deste tipo de dependências depende igualmente da interpretação da relação e da extração incorreta da mesma. Ao se detetar incorretamente 32 relações de eventos, o módulo de ordenação de eventos demonstrou alguma flexibilidade ao extrair apenas 15 relações incorretamente.

Na representação gráfica da solução obtiveram-se, na maioria dos casos, grafos pequenos, constituídos por dois eventos ou um evento e a expressão temporal a ele associado, revelando que existe ainda bastante fragmentação na ordenação dos eventos.

Tarefas	Avaliação	Valores Obtidos	Após 1ª it.	Após 2ª it.
Identificação	corretos	528		
	espúrios	151		
	em falta	244		
Relação Temporal	corretos	24	24	
	espúrios	45	32	
	em falta	74	52	
Relação entre Eventos	corretos	104	103	
	espúrios	171	94	
	em falta	513	228	
Ordenação	corretos	49	48	41
	espúrios	152	86	38
	em falta	556	295	65

Tabela 5.3: Resultados obtidos na avaliação estrita em termos de valores *corretos*, *espúrios* e *em falta*.

	Precisão (%)	Abrangência (%)	Medida-F (%)
Total	24,4	8,1	12,2
Por tarefa:			
- Identificação de Eventos	77,8	68,4	72,8
- Relação Temporal	42,9	31,6	36,4
- Relação entre Eventos	52,3	31,1	39,0
- Ordenação de eventos	51,9	25,6	44,3

Tabela 5.4: Resultados obtidos na avaliação estrita com base nas métricas *precisão*, *abrangência* e *medida-f*.

Tarefas	Avaliação	Valores Iniciais	Após 1ª it.	Após 2ª it.
Identificação	corretos	528		
	espúrios	151		
	em falta	244		
Relações Temporais	corretos	24	24	
	espúrios	45	32	
	em falta	74	52	
Relações entre Eventos	corretos	137	136	
	espúrios	138	61	
	em falta	513	228	
Ordenação	corretos	49	48	48
	espúrios	152	86	46
	em falta	556	295	89

Tabela 5.5: Resultados obtidos na avaliação relaxada em termos de valores *corretos*, *espúrios* e *em falta*.

	Precisão (%)	Abrangência (%)	Medida-F (%)
Total	24,4	8,1	12,2
Por tarefa:			
Identificação de Eventos	77,8	68,4	72,8
Relação Temporal	42,9	31,6	36,4
Relação entre Eventos	69,0	37,4	48,5
Ordenação de Eventos	51,1	35,0	41,6

Tabela 5.6: Resultados obtidos na avaliação relaxada com base nas métricas *precisão*, *abrangência* e *medida-f*.

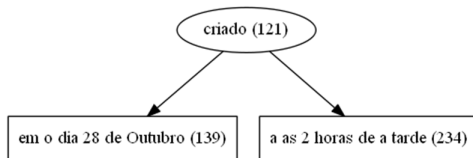
obtido / previsto	rel. direta	rel. indireta	rel. agrup.	rel. temp.	ord. ant.	ord. sim.
relação direta	47	19	5			
relação indireta	7	51	1			
relação de agrup.			7			
relação temporal				24		
ordem anterior					57	15
ordem simultâneo						11

Tabela 5.7: Matriz de confusão obtida a partir dos resultados

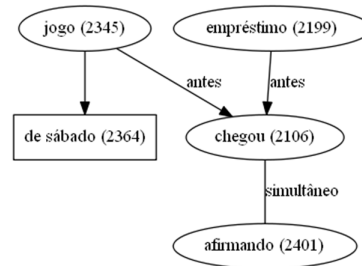
Além disso, foi necessário acrescentar no grafo um identificador para cada evento, de modo a garantir que eventos distintos, mas identificados com a mesma palavra gráfica, não fossem representados como o mesmo.

Entre os grafos de dimensão média obtidos a partir da solução, ordenando vários eventos e expressões temporais associadas aos mesmos, encontra-se grafos gerados com boa legibilidade, como os representados na Figura 5.1.

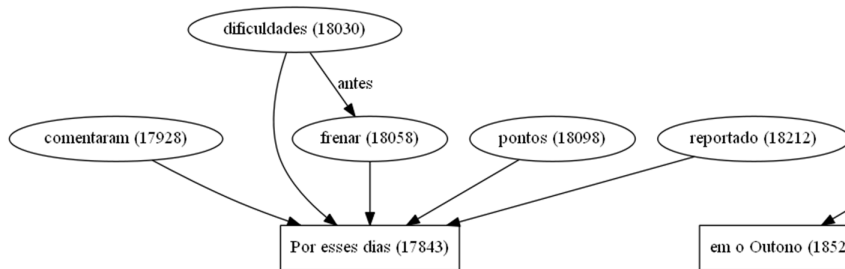
Adão, Pai dos Homens, foi criado no dia 28 de Outubro, às 2 horas da tarde...



Bruno Cortez, de 26 anos, chegou ao Benfica esta época por empréstimo do clube brasileiro que será o adversário no jogo de sábado, afirmando que vai ser especial jogar pela primeira vez no Estádio da Luz.



Por esses dias, alguns maquinistas comentaram nas redes sociais que tinham sentido dificuldades em frenar noutros pontos da rede, embora não tivessem reportado essas situações superiormente.



...a via férrea leva a cuidados redobrados no Outono, ao ponto de os operadores adaptarem os horários dos comboios...

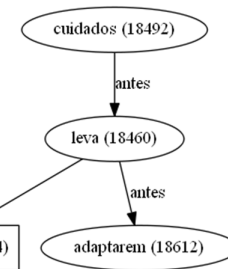


Figura 5.1: Representação de vários grafos gerados sobre as dependências extraídas da solução.

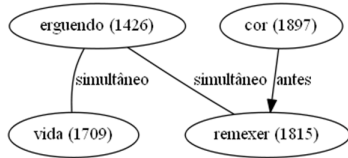
No entanto, detetaram-se problemas de legibilidade na representação de eventos *simultâneos* entre si. Quando existem muitos eventos a ocorrer em simultâneo, o grafo torna-se pouco legível, sendo complicado visualizar à primeira vista que esses eventos ocorrem no mesmo instante temporal, como exemplificado pelos grafos da Figura 5.2.

## 5.4 Discussão

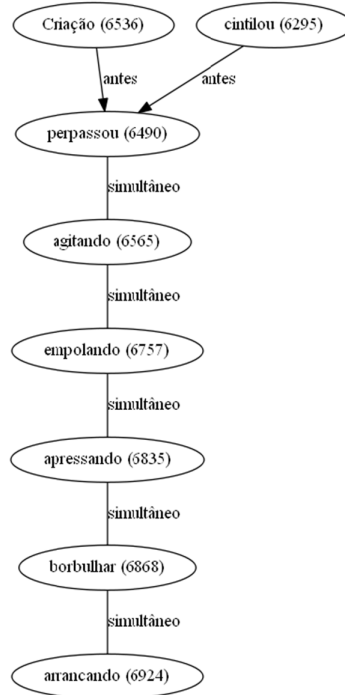
Atualmente obtém-se uma precisão e abrangência já considerável na identificação dos eventos. Ainda que a identificação de eventos não tenha sido o foco deste trabalho, nota-se que desempenha um papel crucial dado que, removidos os eventos *espúrios* ou *em falta*, aumenta-se bastante a percentagem de resultados corretos obtidos. No entanto, mesmo com os eventos corretos, obtém-se resultados poucos satisfatórios nas outras tarefas, pelo que existe ainda muito esforço a ser investido na relação e ordenação de eventos.

A tarefa de relacionamento de eventos também tem um grande impacto na ordenação dos mesmos, como demonstrado pelos resultados. Obteve-se uma abrangência e precisão baixa, mas é possível explicá-la com as decisões tomadas ao longo deste trabalho. Sendo o projecto complexo, simplificou-se a tarefa de relacionamento de eventos ao se recorrer unicamente às conjunções, preposições e advérbios identificados. Assim, não se detetou as relações estabelecidas entre eventos quando o

...erguendo, aqui e além, rígidos troncos duma só folha e dum só rebento, muito solitária, muito silenciosa, com uma vida toda escondida, apenas surdamente revelada pelo remexer de bichos obscuros, gelatinosos, sem cor e quase sem forma, crescendo no fundo dos lodos .



Quando a oitava hora cintilou e fugiu, uma emoção confusa, feita de medo e feita de glória, perpassou por toda a Criação, agitando num frémito as relvas e as frondes, arripiando o pelo das feras, empolando o dorso dos montes, apressando o borbulhar das nascentes, arrancando dos pórfiros um brilho mais vivo...



E as empresas de caminhos-de-ferro possuem até máquinas especiais que limpam os carris, pulverizando a película de gordura criada pelas folhas para que o contacto entre a roda e o carril se faça com mais eficácia.

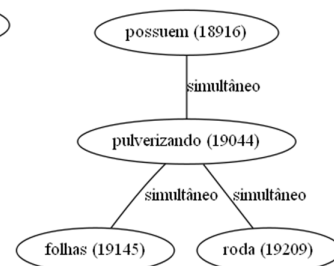


Figura 5.2: Representação de grafos com muitas relações de ordem *simultâneo*.

evento subordinado se encontrava numa oração relativa <sup>3</sup>:

*Ele ainda se não submetera à imobilidade augusta que lhe impôs...*

EVENT\_LEX (submetera, outro)

EVENT\_LEX (imobilidade, outro)

EVENT\_LEX (impôs, outro)

EVENT\_INDIRECTRELATION (submetera, imobilidade, a)

EVENT\_ORDERBEFORE (imobilidade, submetera)

(Não detetado:

EVENT\_INDIRECTRELATION (imobilidade, impôs, que)

EVENT\_ORDERBEFORE (imobilidade, impôs) )

Assim, no excerto, o pronome *que* permite a associação e ordenação dos eventos *imobilidade* e *impôs*, no entanto, o módulo desenvolvido não conseguiu extrair estas relações.

<sup>3</sup>No caso das orações relativas não há um conector explícito, já que o pronome relativo funciona como um constituinte da oração subordinada. Por outro lado o pronome retoma anaforicamente um nome antecedente, o que tornaria mais complexa a forma de representação.

Noutras situações, dois eventos estão associados através de locuções preposicionais complexas, dificultando a deteção correta das relações. Por exemplo, o sistema detetou, no excerto seguinte, uma relação entre os eventos *luxo* e *características*, mas com o conector errado, considerando a preposição simples *de* em vez de considerar o conector composto *em termos de*<sup>4</sup>:

*...alterna troços que são um luxo em termos de características da via...*

EVENT\_LEX (alterna, outro)

EVENT\_LEX (luxo, outro)

EVENT\_LEX (características, outro)

(*Não detetado:*

EVENT\_INDIRECTRELATION (alterna, luxo, que)

EVENT\_ORDERSIMULT (alterna, luxo) )

\*EVENT\_INDIRECTRELATION (luxo, características, de)

\*EVENT\_ORDERBEFORE (características, luxo)

(*Resultado esperado:*

EVENT\_INDIRECTRELATION (luxo, características, em termos de)

EVENT\_ORDERSIMULT (luxo, características) )

Também se verificou que os excertos em que se extraiu menos relações, menos de metade das existentes, são os excertos em que se utiliza bastantes delimitadores do discurso (como a vírgula). Estes permitem a segmentação de frases grandes ou complexas, mas dificultam a análise das mesmas, não sendo possível extrair as relações pelos mesmos métodos utilizados em frases simples. Assim, por exemplo, considere-se a frase:

*Agora, durante os dias genesisacos de 26 e 27, toda ela se completara, se abastecera e se enfeitara, para acolher condignamente o Predestinado que vinha.*

Nesta frase, verifica-se um agrupamento de eventos coordenados (*completara, abastecera, enfeitara*), que, por sua vez, se associam através da conjunção subordinativa final *para* ao evento *acolher*. O sistema ainda não é capaz de analisar o texto de modo a detectar de forma completamente adequada estas situações e será necessário realizar uma análise aprofundada sobre os delimitadores antes de se prosseguir para uma solução apropriada. Do mesmo modo, a extração de relações temporais também foi afectada pelo uso frequente dos delimitadores de discurso. Como consequência, na frase anterior, o sistema não conseguiu associar o advérbio *agora* a nenhum dos eventos identificados. Adicionalmente, ainda existem algumas expressões temporais complexas que não são devidamente identificadas, impedindo a extração das respectivas relações. A expressão *durante os dias genesisacos de 26 e 27* representa um destes casos, tratando-se de uma expressão temporal complexa. Assim, obtém-se o seguinte resultado:

*Agora, durante os dias genesisacos de 26 e 27, toda ela se completara, se abastecera e se enfeitara, para acolher condignamente o Predestinado que vinha.*

EVENT\_LEX (completara, outro)

EVENT\_LEX (abastecera, outro)

EVENT\_LEX (enfeitara, outro)

EVENT\_LEX (acolher, outro)

EVENT\_LEX (vinha, outro)

---

<sup>4</sup>A expressão *em termos de* é ambígua, pelo que não foi lexicalizada na STRING.



EVENT\_INDIRECTRELATION (enfeitara, acolher, para)

EVENT\_ORDERBEFORE (enfeitara, acolher)

(*Não detetado:*

EVENT\_GROUPRELATION (completara, abastecera, e)

EVENT\_GROUPRELATION (abastecera, enfeitara, e)

EVENT\_TIMERELATION (completara, agora)

EVENT\_TIMERELATION (abastecera, agora)

EVENT\_TIMERELATION (enfeitara, agora)

EVENT\_TIMERELATION (completara, durante os dias genesíacos de 26 e 27)

EVENT\_TIMERELATION (abastecera, durante os dias genesíacos de 26 e 27)

EVENT\_TIMERELATION (enfeitara, durante os dias genesíacos de 26 e 27)

EVENT\_INDIRECTRELATION (completara, acolher, para)

EVENT\_INDIRECTRELATION (abastecera, acolher, para)

EVENT\_INDIRECTRELATION (acolher, vinha, que)

EVENT\_ORDERSIMULT (completara, abastecera)

EVENT\_ORDERSIMULT (abastecera, enfeitara)

EVENT\_ORDERBEFORE (completara, acolher)

EVENT\_ORDERBEFORE (abastecera, acolher)

EVENT\_ORDERBEFORE (acolher, vinha )

A ordenação dos eventos depende, em grande parte, da interpretação do significado dos conectores e eventos. No entanto, por vezes, a ordem que foi estabelecida entre os eventos baseou-se no caso comum, isto é, tom-se decisões com base no que se conhece, sob o risco de errar algumas situações mais específicas. Uma das decisões envolveu a ordenação de eventos obtidos a partir de construções nominais, que não possuem informação temporal e aspetual, desde que o evento subordinante fosse de natureza verbal e possuísse esse tipo de informação. Por exemplo, apesar do evento *obra* ser obtido a partir de uma construção nominal, ao ser um evento subordinado ao evento *apresentava*, deteta-se a relação de ordem:

*A Inês apresentava a obra.*

EVENT\_LEX (apresentava, outro)

EVENT\_LEX (obra, outro)

EVENT\_INDIRECTRELATION (apresentava, obra, a)

EVENT\_ORDERBEFORE (apresentava, obra)

No entanto, em certas situações, apesar de os eventos estarem relacionados, devido ao significado dos mesmos, não existe uma relação de ordem:

*Os cientistas puderam medir com precisão...*

EVENT\_LEX (medir, outro)

EVENT\_LEX (precisão, outro)

EVENT\_INDIRECTRELATION (medir, precisão, com)

\*EVENT\_ORDERBEFORE (precisão, medir)

O sistema identificou, no excerto apresentado, uma relação entre os eventos *medir*, obtido a partir de uma construção verbal auxiliada pelo verbo *poder*, e *precisão*, obtido a partir de uma construção nominal, e ligado ao verbo *medir* através da preposição *com*. Após identificada a relação, o sistema prosseguiu para a ordenação, no entanto, o evento *precisão* funciona aqui como um complemento circunstancial de maneira/modo, não sendo correto considerar uma ordem entre os eventos (excepto, talvez, simultaneidade).

Noutras situações, existe realmente uma relação de ordem, mas a mesma é mal identificada por falta de informação semântica sobre os eventos envolvidos:

...Benfica vai “*mostrar a sua força*”...

\*EVENT\_LEX(vai, outro)

EVENT\_LEX(mostrar, outro)

EVENT\_LEX(força, outro)

EVENT\_DIRECTRELATION(mostrar, força)

\*EVENT\_ORDERBEFORE(força, mostrar)

(Resultado Esperado:

EVENT\_ORDERSIMULT(mostrar, força) )

Na frase apresentada, o evento *mostrar* é simultâneo ao evento *força*, mas o sistema considera que o evento *força* é anterior a *mostrar*. Ou seja, apresenta o mesmo comportamento que tem ao processar os eventos *avaliar* e *força* na frase, de estrutura semelhante:

...Benfica vai *avaliar a sua força*...

\*EVENT\_LEX(vai, outro)

EVENT\_LEX(avaliar, outro)

EVENT\_LEX(desempenho, outro)

EVENT\_DIRECTRELATION(avaliar, força)

EVENT\_ORDERBEFORE(força, avaliar)

Por último, focou-se pouco o trabalho na análise e resolução do problema a partir das relações temporais. A ordenação de eventos baseada apenas em *passado*, *presente* e *futuro* torna-se desapropriada quando se lida com expressões temporais, dado o potencial existente no uso de datas (e.g.: *3 de Abril*; *no final de 2001*) e outras referências temporais (e.g.: *no Natal*; *na quarta-feira de manhã*).

A frase seguinte exemplifica uma situação em que, mesmo sem conhecer o significado dos eventos *lexnascimento* e *lexmorte*, subentendidos na estrutura da frase, teria sido possível ordená-los ao explorar os valores representados pelas expressões temporais.

Fernando António Nogueira Pessoa (*Lisboa, 13 de Junho de 1888 — Lisboa, 30 de Novembro de 1935*), mais conhecido como Fernando Pessoa, foi um poeta, filósofo e escritor português .

EVENT\_LEX(lexnascimento, nascimento)

EVENT\_LEX(lexmorte, morte)

EVENT\_TIMERELATION(lexmorte, 30 de Novembro de 1935)

EVENT\_TIMERELATION(lexnascimento, 13 de Junho de 1888)

(Não detetado: EVENT\_ORDERBEFORE (lexnascimento, lexmorte) )

Na frase apresentada, associaram-se corretamente as datas *13 de Junho de 1888* e *30 de Novembro de 1935* aos respectivos eventos, *lexnascimento* e *lexmorte*, mas não foi possível extrair nenhuma relação de ordem entre os mesmos.

Em relação à representação gráfica, verificam-se, como se disse atrás, alguns problemas de legibilidade quando existem muitos eventos *simultâneos* entre si. Grande parte deste problema deve-se ao alinhamento dos nós, definido por omissão na geração do grafo. Idealmente, para uma melhor legibilidade, todos os eventos *simultâneos* entre si, e apenas esses, deveriam realmente ficar alinhados na horizontal, como exemplificado pela Figura 5.3. A resolução deste problema, no entanto, não é simples devido à linguagem utilizada para gerar os grafos.

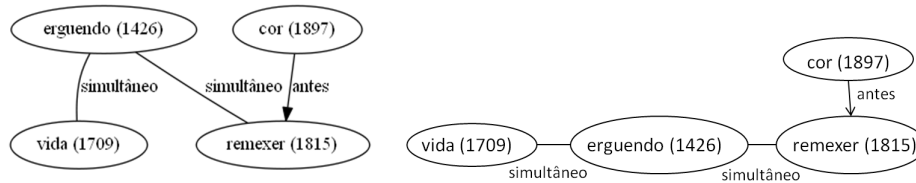


Figura 5.3: Comparação entre a representação adotada na solução, à esquerda, com o resultado de alinhar os eventos *simultâneos*.

Adicionalmente, é possível duplicar as relações de ordem entre os eventos *simultâneos*. Por exemplo, ao se detetar que um evento *A* é *simultâneo* a *B* e o evento *B* é *anterior* a *C*, deduz-se que o evento *A* é *anterior* a *C*, como representado na Figura 5.4. Este tipo de raciocínio sobre as relações de ordem entre os eventos pode ser aplicada no aperfeiçoamento da solução, tornando a representação da mesma mais legível.

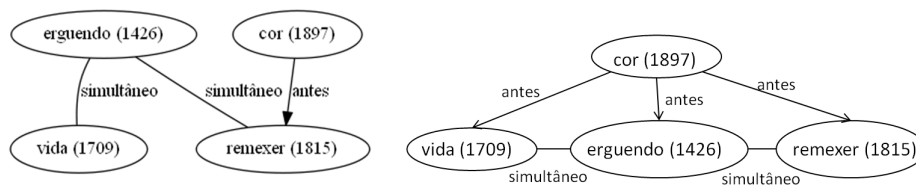


Figura 5.4: Comparação entre a representação gráfica da solução atual, à esquerda, com a solução ideal, à direita.



## Capítulo 6

# Conclusão e Trabalhos Futuros

O problema de identificação e ordenação de eventos é bastante complexo e extenso para ser totalmente resolvido num único passo, pelo que, neste estudo, se reduziu o conjunto de relações a tratar e se simplificou o problema de identificação de eventos. Uma das simplificações refere-se ao facto de se considerar a relação e ordenação apenas dentro de uma mesma frase e, ao contrário de alguns sistemas analisados, considerar apenas as relações locais para a decisão quanto à ordem a extrair, em vez de se considerar todas as relações existentes em simultâneo. Simplificou-se, também, o trabalho ao se considerar apenas as relações de ordem *antes* e *simultâneo*. Ainda assim, estruturou-se a solução de forma a possibilitar a uma futura conversão para a lógica baseada em intervalos, mais completa e mais utilizada, segundo os sistemas analisados durante este trabalho.

Expandiu-se igualmente a identificação de eventos não normalizados e obteve-se uma solução com base na análise dos conectores, expressões temporais e verbos associados aos eventos. Ao avaliar a solução, como descrito no capítulo anterior, detetaram-se algumas situações que exigem mais esforço e tempo para se resolver. Inclusive, nos problemas mais complexos, a análise humana recorre ao valor semântico das frases para obter o sentido, não sendo isto inteiramente possível para o módulo. O módulo demonstra problemas na ordenação dos eventos quando existe pouca informação disponível, como a presença de construções verbais não marcadas ou a ausência de conectores, e problemas mais complexos, como o uso de conectores compostos ou frases bastante fragmentadas pelo uso de vírgulas, ou quando a solução depende do valor semântico da frase.

A utilização de estatísticas ou de um método probabilístico pode auxiliar a resolução destas situações. Inclusive, esse tipo de abordagem já foi realizada com sucesso em sistemas anteriores, nomeadamente em TRIPS e TRIOS (2010) [56], NCSU-INDI (2010) [26] e Maršić (2011) [35]. Em contrapartida, será necessário utilizar um corpus de grande dimensões para teste, sendo uma opção pouco eficaz, caso contrário.

Para esta solução, como descrito anteriormente, utilizou-se uma abordagem baseada em regras, à semelhança do TERSEO (2005) [49, 50] e Maršić (2011) [35], a qual exige uma análise bastante cuidada da língua. Contudo, como exemplificado pelos sistemas analisados, com mais investimento na análise linguística é possível obter-se resultados melhores. Uma sugestão de melhoria, seguindo esta abordagem, será a de investir na interpretação de expressões temporais, na determinação do momento textual e de enunciação, e na análise dos delimitadores do discurso.

Outra situação, que outros sistemas tiveram em conta, tal como Chambers et al. (2008) e Maršić (2011), foi a realização da análise global das relações entre eventos. Chambers et al. realizaram várias abordagens e verificaram um aumento global da precisão e da abrangência ao abordar o problema tanto como local como globalmente.

Uma alternativa à abordagem baseada em regras passa por uma estratégia de aprendizagem automática, como em Chambers et al. (2008) [12], ou por uma solução híbrida. Ambas exigem a utilização de um corpus de treino e de teste de grande dimensão para se obter bons resultados, mas revelaram-se recompensadores do esforço dispendido na sua construção. O sistema Maršić é um bom exemplo desta abordagem híbrida, utilizando a abordagem baseada em regras em conjunto com métodos de

aprendizagem automática ou resultados de estatísticas já existentes, na tentativa de combinar o melhor dos dois mundos.

# Referências

- [1] Ait-Mokhtar, S., Chanod, J., and Roux, C. (2002). Robustness beyond shallowness: incremental dependency parsing. *Natural Language Engineering*, 8:121–144.
- [2] Allen, J. F. and Ferguson, G. (1994). Actions and Events in Interval Temporal Logic.
- [3] Areces, C. and Blackburn, P. (2005). Reichenbach, Prior and Montague: A semantic get-together. pages 77–88.
- [4] Baptista, J., Cabarrão, V., and Mamede, N. J. (2012). Classification directives for Events and Relations Extraction between Named Entities in Portuguese Texts. Technical report, L2F-Spoken Language Laboratory.
- [5] Baptista, J., Mamede, N., and Gomes, F. (2010). Auxiliary verbs and verbal chains in european portuguese. In Pardo, T., Branco, A., Klautau, A., Vieira, R., de Lima Tiago Pardo, V. L. S., Branco, A., Klautau, A., Vieira, R., and de Lima, V. L. S., editors, *Computational Processing of the Portuguese Language*, number 6001 in Lecture Notes in Computer Science / Lecture Notes in Artificial Intelligence, Berlin. PROPOR 2010, Springer.
- [6] Baptista, J., Mamede, N., Hagège, C., and Maurício, A. (2011). Time Expressions in Portuguese Guidelines for Identification, Classification and Normalization. Technical report, L2F-Spoken Language Laboratory.
- [7] Berger, A. L., Pietra, S. A. D., and Pietra, V. J. D. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, 22:39–71.
- [8] Boguraev, B., Pustejovsky, J., Ando, R., and Verhagen, M. (2007). TimeBank evolution as a community resource for TimeML parsing. *Language Resources and Evaluation*, 41(1):91–115.
- [9] Bray, T., Paoli, J., Sperberg-McQueen, C. M., and Maler, E. (2000). Extensible Markup Language (XML) 1.0 (Second Edition).
- [10] Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2:121–167.
- [11] Casati, R. and Varzi, A. (2010). Events. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2010 edition.
- [12] Chambers, N. and Jurafsky, D. (2008). Jointly Combining Implicit Constraints Improves Temporal Ordering. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 698–706, Honolulu, Hawaii. Association for Computational Linguistics.
- [13] Chambers, N., Wang, S., and Jurafsky, D. (2007). Classifying Temporal Relations Between Events. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 173–176, Prague, Czech Republic. Association for Computational Linguistics.

- [14] Chklovski, T. and Pantel, P. (2004). VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. In Lin, D. and Wu, D., editors, *Proceedings of EMNLP 2004*, pages 33–40, Barcelona, Spain. Association for Computational Linguistics.
- [15] Costa, M. H. and Xavier, M. F. (1991). *Sintaxe e Semântica do Português*. Universidade Aberta.
- [16] de Moura Rodrigues, D. J. G. (2007). Uma evolução do sistema ShRep. Optimização, interface gráfica e integração de mais duas ferramentas. Master’s thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- [17] Diniz, C., Mamede, N. J., and Pereira, J. C. S. D. (2010). RuDriCo2 - A faster disambiguator and segmentation modifier. In *II Simpósio de Informática (INForum)*, pages 573–584, Universidade do Minho, Portugal.
- [18] Diniz, C. F. P. (2010). Um Conversor baseado em regras de transformação declarativas. Master’s thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- [19] Fellbaum, C., editor (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, Cambridge, MA, illustrated edition.
- [20] Felzenszwalb, P. F. and Mcallester, D. (2007). The generalized A\* architecture. *Journal of Artificial Intelligence Research*, 29:153–190.
- [21] Ferro, L., Gerber, L., Hitzeman, J., Lima, E., and Sundheim, B. (2004). ACE Time Normalization (TERN) 2004 English Training Data v 1.3. Philadelphia, Pa. Linguistic Data Consortium.
- [22] Ferro, L., Gerber, L., Mani, I., Sundheim, B., and Wilson, G. (2005). TIDES 2005 Standard for the Annotation of Temporal Expressions. Technical report, MITRE Corporation.
- [23] Fletcher, T. (2009). Support Vector Machines Explained. <http://www.tristanfletcher.co.uk/>. Última visita em: 2014-10-02.
- [24] Gansner, E. R., Koutsofios, E., and North, S. (2009). Drawing graphs with DOT. <http://www.graphviz.org/Documentation/dotguide.pdf>. Última visita em: 2014-10-02.
- [25] Garside, R., Leech, G. N., and McEnery, T. (1997). *Corpus annotation: linguistic information from computer text corpora*. Longman.
- [26] Ha, E. Y., Baikadi, A., Licata, C., and Lester, J. C. (2010). NCSU: Modeling Temporal Relations with Markov Logic and Lexical Ontology. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval ’10*, pages 341–344, Uppsala, Sweden. Association for Computational Linguistics.
- [27] Hackmack, S. (2007). Reichenbach’s Theory of Tense and it’s Application to English. <http://www.fb10.uni-bremen.de/khwagner/verb/pdf/Reich.pdf>. Última visita em: 2014-10-02.
- [28] Hagège, C. and Tannier, X. (2007). XRCE-T: XIP temporal module for TempEval campaign. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval ’07*, pages 492–495, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [29] Harabagiu, S. M., Miller, G. A., and Moldovan, D. I. (1999). WordNet 2 - A Morphologically and Semantically Enhanced Resource. In *SIGLEX99: Standardizing Lexical Resources*, pages 1–8.
- [30] Hart, P. E., Nilsson, N. J., and Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4(2):100–107.



- [31] Huang, L. (2008). *Forest-based algorithms in Natural Language Processing*. PhD thesis, University of Pennsylvania, Philadelphia, PA, USA.
- [32] Jurafsky, D. and Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall, Pearson Education International, Englewood Cliffs, NJ, 2 edition.
- [33] Loureiro, J. (2007). Reconhecimento de Entidades Mencionadas (Obra, Valor, Relações de Parentesco e Tempo) e Normalização de Expressões Temporais. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa. MSc Dissertation.
- [34] Mamede, N., Baptista, J., and Cláudio, D. (2012). STRING - An Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. In Springer, editor, *PROPOR 2012*.
- [35] Maršić, G. (2011). *Temporal Processing of News: Annotation of Temporal Expressions, Verbal Events and Temporal Relations*. PhD thesis, University of Wolverhampton, Wolverhampton, UK.
- [36] Mateus, M. H. M., Brito, A. M., Duarte, I., and Faria, I. H. (2003). *Gramática da Língua Portuguesa*. Editorial Caminho, 3<sup>rd</sup> edition.
- [37] Maurício, A. (2011). Identificação, Classificação e Normalização de Expressões Temporais. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.
- [38] Miller, G. A. (1995). WordNet: A Lexical Database for English. In *Communications of the ACM*, volume 38, pages 39–41.
- [39] Och, F. J., Ueffing, N., and Ney, H. (2001). An Efficient A\* Search Algorithm for Statistical Machine Translation. In *In Data-Driven Machine Translation Workshop*, pages 55–62.
- [40] Pustejovsky, J., Castañó, J. M., Ingria, R., Sauri, R., Gaizauskas, R. J., Setzer, A., Katz, G., and Radev, D. R. (2003a). TimeML: Robust Specification of Event and Temporal Expressions in Text. In Maybury, M. T., editor, *New Directions in Question Answering*, pages 28–34. AAAI Press.
- [41] Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., and Lazo, M. (2003b). The TIMEBANK corpus. In *Proceedings of Corpus Linguistics 2003*, pages 647–656, Lancaster.
- [42] Pustejovsky, J. and Verhagen, M. (2009). SemEval-2010 task 13: evaluating events, time expressions, and temporal relations (TempEval-2). In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions*, DEW '09, pages 112–116, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [43] Pustejovsky, J., Verhagen, M., Nianwen, X., Gaizauskas, R., Hepple, M., Schilder, F., Katz, G., Sauri, R., Saquete, E., Caselli, T., Calzolari, N., Lee, K., , and Im, S. (2009). TempEval-2: Evaluating Events, Time Expressions, and Temporal Relations. <http://www.timeml.org/tempeval2/tempeval2-proposal.pdf>. Última visita em: 2014-10-02.
- [44] Ramage, D. (2007). Hidden Markov Models Fundamentals. <http://cs229.stanford.edu/section/cs229-hmm.pdf>. Última visita em: 2014-10-02.
- [45] Reichenbach, H. (1947). The Tenses of Verbs. In *Elements of Symbolic Logic*, pages 287–298. Macmillan, New York.
- [46] Ribeiro, R. (2003). Anotação Morfossintáctica Desambiguada do Português. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.

- [47] Richardson, M. and Domingos, P. (2006). Markov Logic Networks. *Machine Learning*, 62(1-2).
- [48] Roth, D. and Yih, W. (2004). A Linear Programming Formulation for Global Inference in Natural Language Tasks. In *In Proceedings of CoNLL-2004*, pages 1–8.
- [49] Saquete, E., Muñoz, R., and Martínez-Barco, P. (2003). TERSEO: Temporal Expression Resolution System Applied to Event Ordering. In Matoušek, V. and Mautner, P., editors, *Text, Speech and Dialogue*, volume 2807 of *Lecture Notes in Computer Science*, pages 220–228. Springer Berlin Heidelberg.
- [50] Saquete, E., Muñoz, R., and Martínez-Barco, P. (2006). Event ordering using TERSEO system. *Data Knowl. Eng.*, 58(1):70–89.
- [51] Saurí, R., Goldberg, L., Verhagen, M., and Pustejovsky, J. (2009). Annotating Events in English TimeML Annotation Guidelines. <http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/EventGuidelines-050409.pdf>. Última visita em: 2014-10-05.
- [52] Saurí, R. and Pustejovsky, J. (2009). TimeML in a Nutshell. <http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/introToTimeML-052809.pdf>. Última visita em: 2014-10-05.
- [53] Setzer, A. (2001). *Temporal Information in Newswire Articles: An Annotation Scheme and Corpus Study*. PhD thesis, University of Sheffield.
- [54] TimeML Working Group (2009). Guidelines for Temporal Expression Annotation for English for TempEval 2010. <http://www.timeml.org/tempeval2/tempeval2-trial/guidelines/timex3guidelines-072009.pdf>. Última visita em: 2014-10-05.
- [55] UzZaman, N. and Allen, J. (2010a). TRIOS-TimeBank Corpus: Extended TimeBank corpus with help of Deep Understanding of Text. In Chair, N. C. C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- [56] UzZaman, N. and Allen, J. F. (2010b). TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval '10*, pages 276–283, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [57] Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., and Pustejovsky, J. (2007). SemEval-2007 Task 15: TempEval Temporal Relation Identification. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07*, pages 75–80, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [58] Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Moszkowicz, J., and Pustejovsky, J. (2009). The TempEval challenge: identifying temporal relations in text. *Language Resources and Evaluation*, 43(2):161–179.
- [59] Vicente, A. M. F. (2013). LexMan: um Segmentador e Analisador Morfológico com Transdutores. Master's thesis, Instituto Superior Técnico.
- [60] Wallach, H. M. (2004). Conditional random fields: An introduction. Technical report, University of Pennsylvania.
- [61] Weston, J. and Ben-Hur, A. (2010). A User's Guide to Support Vector Machines. In Carugo, O. and Eisenhaber, F., editors, *Data Mining Techniques for the Life Sciences*, volume 609 of *Methods in Molecular Biology*, pages 223–239. Humana Press.
- [62] Wilt, C. M., Thayer, J. T., and Ruml, W. (2010). A comparison of greedy search algorithms. In Felner, A. and Sturtevant, N. R., editors, *SOCS*. AAAI Press.

# **Appendix A**

## **Corpus**

Em Portugal poderia ocorrer um acidente igual ao da Galiza? CARLOS CIPRIANO 03/08/2013 - 00:00 Portugal só começou a modernizar as suas linhas nos anos 90 e tem, por isso, um dos sistemas mais seguros do mundo ADRIANO MIRANDA Portugal tem um dos melhores sistemas de segurança ferroviários do mundo. E também um dos mais obsoletos. Mas um acidente como o de há dez dias em Espanha seria quase impossível de acontecer Cerca de 1300 quilómetros da rede ferroviária de passageiros portuguesa têm um sistema de segurança superior ao ASFA espanhol, instalado no troço onde se deu o acidente a 24 de Julho junto a Santiago de Compostela, na Galiza. Mas há 630 quilómetros - um terço da rede - que dependem exclusivamente de meios humanos, num sistema praticamente igual ao que existia no séc. XIX e no qual só se substituiu o telégrafo pelo telefone. Em todo o caso, um acidente como o de Santiago é muito improvável em Portugal. Em primeiro lugar, porque a CP não dispõe de muitos comboios que atinjam os 190km/hora - só os pendulares e os Intercidades é que circulam nesse patamar de velocidade. E, em segundo lugar, porque existe o sistema Convel (Controlo de Velocidade), que monitoriza o andamento da composição, impedindo-a de circular a uma velocidade superior à permitida em cada momento. Nas vias férreas portuguesas há imensos locais onde, em pouco tempo, o maquinista também é obrigado a reduzir a velocidade de 200 para 80 km/hora, tal como na curva da Linha Ourense-Santiago onde há dez dias ocorreu um dos piores desastres ferroviários das últimas décadas em Espanha. Basta dizer que a Linha do Norte, como não foi totalmente modernizada, alterna troços que são um luxo em termos de características da via com outros mais desgastados, em que se tem de circular mais devagar. E isso faz de uma viagem entre Lisboa e Porto um autêntico rally ferroviário, pois há pontos em que se circula abaixo de 100 km/hora e outros onde a velocidade é de 220 km/hora. Só que, ao contrário do ASFA espanhol, o Convel português "seguro" o comboio e obriga-o a respeitar a curva de frenagem, se o maquinista for distraído. No limite, o Convel substitui-se ao condutor do comboio e faz parar a composição em plena via numa frenagem de emergência, se este não reduzir a velocidade por forma a entrar na secção seguinte de acordo com o limite estabelecido. Este sistema luso é considerado um topo de gama em termos de segurança no tráfego ferroviário, ombreado com o que há de melhor a nível mundial. E deve-se ao facto de Portugal só ter começado a modernizar as suas linhas nos anos 90, assimilando assim a tecnologia mais importante da época e que ainda hoje é mais avançado do que o sistema alemão, inglês, espanhol ou suíço. Por exemplo, o choque frontal entre dois comboios na Suíça que ocorreu na semana passada não teria sido possível naquelas condições em Portugal, porque o Convel não deixaria que um dos comboios saísse da estação sem o outro chegar. O Convel permite ainda a condução em "piloto automático", na qual o maquinista dá indicações ao computador de bordo sobre a velocidade a que se pode circular naquele troço e o próprio comboio acelera ou freia - os comboios não travam porque não têm travões, mas sim freios - consoante esteja a subir ou a descer, de maneira a manter-se dentro do limite imposto. O que falta modernizar Mas há uma outra geografia ferroviária portuguesa que vive ainda a um ritmo não muito diferente do século XIX. São as linhas que ainda não foram modernizadas e cuja exploração continua dependente do cantonamento telefónico - um sistema no qual o chefe da estação telefona para a estação seguinte a pedir o avanço do comboio. Neste caso, a circulação ferroviária está inteiramente dependente de meios humanos. A segurança é reforçada com um conjunto de procedimentos redundantes que os agentes da Refer têm de cumprir e que se destinam a evitar erros. Há ainda a supervisão de um posto regulador em Lisboa ou no Porto, mas, na prática, o cantonamento telefónico é um mero upgrade do tempo em que a marcha dos comboios era assinalada pelo telégrafo entre estações. Na história recente, só dois acidentes ferroviários foram provocados por uma falha humana - Alcafache, em 1985, e Lousã, em 2002 -, em que duas composições chocaram de frente em via única. Ao contrário do ambiente tecnológico em que o posto regulador sabe sempre onde se encontra o comboio, os maquinistas portugueses trabalham, por outro lado, em linhas onde, por vezes, estão isolados. Em certas horas do dia, um comboio que saía de Meleças (Cacém) para as Caldas da Rainha percorre 84 quilómetros sem que o posto regulador saiba onde ele se encontra. O mesmo acontece entre o Pocinho e a Régua, em que o cantão (distância entre estações guardadas com pessoal) é de 68 quilómetros. À margem da modernização permanecem assim as linhas do Minho, do Douro, do Oeste e parte das do Algarve e do Alentejo. A Refer tenciona concorrer a fundos comunitários para electrificar e dotar de sinalização automática algumas delas.

Fonte: <http://www.publico.pt/portugal/jornal/em-portugal-poderia-ocorrer-um-acidente-igual-ao-da-galiza-26912538> [2013-08-06]

Governo recusa divulgar relatório do acidente ferroviário em Alfarelos 03/08/2013 - 00:00 Colisão deu-se a 21 de Janeiro O relatório ao acidente ferroviário do passado mês de Janeiro em Alfarelos, no distrito de Coimbra, ainda não foi divulgado publicamente, apesar de já estar terminado há vários meses. O PÚBLICO solicitou-o ao IMT (Instituto da Mobilidade e dos Transportes), que nem sequer respondeu, pelo que foi apresentada queixa à Comissão de Acesso aos Documentos Administrativos. As reticências do Governo devem-se ao facto de o relatório ser inconclusivo quanto às causas que levaram o Intercidades procedente de Lisboa a embater na cauda de um regional que estava parado na estação de Alfarelos em 21 de Janeiro deste ano. Todos os sistemas funcionaram. A sinalização estava operacional e o Convel também actuou, pois, perante a incapacidade do maquinista em reduzir a velocidade para o limite estabelecido, foi automaticamente accionado o freio de emergência que bloqueou todos os rodados da composição. Por sua vez, os freios também estavam em boas condições. Ou seja: não foram detectadas falhas na infra-estrutura nem nos comboios acidentados.

O Intercidades deslizou ao longo da via com um maquinista incrédulo ao ver que o sistema de frenagem, simplesmente, não fazia efeito. Uma simulação da mesma frenagem, realizada dias depois pela comissão de inquérito com um comboio idêntico e em condições parecidas (de noite e com chuva), comprovou que tudo funcionou sem problemas. A explicação poderá estar nas folhas de árvores que tinham caído na linha durante o temporal que assolara o país na véspera, tal como o PÚBLICO já adiantou. Esta possibilidade pode fazer sorrir muita gente, mas a verdade é que está comprovado que as folhas pisadas sobre o carril criam uma película de gordura que faz com que, em circunstâncias excepcionais, um comboio "escorregue", mesmo quando as rodas estão bloqueadas. Em Alfarelos, também a automotora regional que seguia à frente do Intercidades tivera dificuldades em frenar à entrada da estação, não conseguindo o maquinista evitar que ultrapassasse o sinal vermelho. Por esses dias, alguns maquinistas comentaram nas redes sociais que tinham sentido dificuldades em frenar noutros pontos da rede, embora não tivessem reportado essas situações superiormente. Em Inglaterra e no Canadá, o problema das folhas caídas sobre a via férrea leva a cuidados redobrados no Outono, ao ponto de os operadores adaptarem os horários dos comboios à velocidade mais reduzida com que têm de circular nessa época. E as empresas de caminhos-de-ferro possuem até máquinas especiais que limpam os carris, pulverizando a película de gordura criada pelas folhas para que o contacto entre a roda e o carril se faça com mais eficácia. C.C.

Fonte: <http://www.publico.pt/portugal/jornal/governo-recusa-divulgar-relatorio-do-acidente-ferroviario-em-alfarelos-26912546> [2013-08-06]

Adão, Pai dos Homens, foi criado no dia 28 de Outubro, às 2 horas da tarde... Assim o afirma, com majestade, nos seus "Annales Veteris et Novi Testamenti", o muito douto e muito ilustre Usseus, Bispo de Meath, Arcebispo de Armagh, e Chanceler-Mor da Sé de S. Patrício. A Terra existia desde que a Luz se fizera, a 23, na manhã de todas as manhãs. Mas já não era essa Terra primordial, parda e mole, enopada em águas barrentas, abafada numa névoa densa, erguendo, aqui e além, rígidos troncos duma só folha e dum só rebento, muito solitária, muito silenciosa, com uma vida toda escondida, apenas surdamente revelada pelo remexer de bichos obscuros, gelatinosos, sem cor e quase sem forma, crescendo no fundo dos lodos. Não! agora, durante os dias genesíacos de 26 e 27, toda ela se completara, se abastecera e se enfeitara, para acolher condignamente o Predestinado que vinha. No dia 28 já apareceu perfeita, "perfecta", com as provisões e alfaias que a Bíblia enumera, as ervas verdes de espiga madura, as árvores providas do fruto entre a flor, todos os peixes nadando nos mares resplandecentes, todas as aves voando pelos ares aclarados, todos os animais pastando sobre as colinas viçosas, e os regatos regando, e o fogo armazenado no seio da pedra, e o cristal, e o ónix, e o oiro muito bom do país de Hevilath... Nesses tempos, meus amigos, o Sol ainda girava em torno da Terra. Ela era moça e formosa e preferida de Deus. Ele ainda se não submetera à imobilidade augusta que lhe impôs mais tarde, entre amuados suspiros da Igreja, mestre Galileu, estendendo um dedo do fundo do seu pomar, rente aos muros do Convento de S. Mateus de Florença. E o sol, amorosamente, corria em volta da Terra, como o noivo dos "Cantares", que, nos lascivos dias da ilusão, sobre o outeiro de mirra, sem descanso e pulando mais levemente que os gamos de Gaalad, circundava a Bem-Amada, a cobria com o fulgor dos seus olhos, coroado de sal-gema, a faiscar de fecunda impaciência. Ora desde essa alvorada do dia 28, segundo o cálculo majestático de Usseus, o Sol, muito novo, sem sardas, sem rugas, sem falhas na sua cabeleira flamante, envolvera a terra, durante oito horas, numa contínua e insaciada carícia de calor e de luz. Quando a oitava hora cintilou e fugiu, uma emoção confusa, feita de medo e feita de glória, perpassou por toda a Criação, agitando num frémito as relvas e as frondes, arripiando o pelo das feras, empolando o dorso dos montes, apressando o borbulhar das nascentes, arrancando dos pórfiros um brilho mais vivo... Então numa floresta muito cerrada e muito tenebrosa, certo Ser, despreendendo lentamente a garra do galho de árvore onde se empoleirara toda essa manhã de longos séculos, escorregou pelo tronco comido de hera, pousou as duas patas no solo que o musgo afofava, sobre as duas patas se firmou com esforçada energia, e ficou erecto, e alargou os braços livres, e lançou um passo forte, e sentiu a sua dissemelhança da Animalidade, e concebeu o deslumbrado pensamento do que era, e verdadeiramente "foi"! Deus, que o amparara, nesse instante o criou. E vivo, da vida superior, descido da inconsciência da árvore, Adão caminhou para o Paraíso. Era medonho. Um pelo crespo e luzidio cobria todo o seu grosso, maciço corpo, rareando apenas em torno dos cotovelos, dos joelhos rudes, onde o coiro aparecia curtido e da cor de cobre fosco. Do achatado, fugidio crânio, vincado de rugas, rompia uma guedelha rala e ruiva, tufando sobre as orelhas agudas. Entre as rombas queixadas, na fenda enorme dos beiços trombudos, estirados em focinho, as presas reluziam, afiadas rijamente para rasgar a febra e esmigalhar o osso. E sob as arcadas sombriamente fundas, que um felpe hirsuto orlava como um silvado orla o arco duma caverna, os olhos redondos, dum amarelo de âmbar, sem cessar se moviam, tremiam, esgazeados de inquietação e de espanto... Não, não era belo, nosso Pai venerável, nessa tarde de Outono, quando Jeová o ajudou com carinho a descer da sua Árvore! E todavia, nesses olhos redondos, de fino âmbar, mesmo através do tremor e do espanto, rebrilhava uma superior beleza—a Energia Inteligente que o ia tropeçadamente levando, sobre as pernas arqueadas, para fora da mata onde passara a sua manhã de longos séculos a pular e a guinchar por cima dos ramos altos. Mas (se os Compêndios de Antropologia nos não iludem) os primeiros passos humanos de Adão não foram logo atirados, com alacridade e confiança, para o destino que o esperava entre

os quatro rios do Éden . Entorpecido, envolvido pelas influências da Floresta, ainda despega com custo a pata de entre o folhoso chão de fetos e begónias, e gostosamente se roça pelos pesados cachos de flores que lhe orvalham o pêlo, e acaricia as longas barbas de Líquen branco, pendentes dos troncos de roble e de teca, onde gozara as doçuras da irresponsabilidade . Nas ramagens que tão generosamente, através tão longas idades, o nutriram e o embalaram, ainda colhe as bagas sumarentas, os rebentões mais tenros . Para transpor os regatos, que por todo o bosque reluzem e sussurram depois da sazão das chuvas, ainda se pendura duma rija liana, entrelaçada de orquídeas, e se balança, e arqueia o pulo, com pesada indolência . E receio bem que quando a aragem restolhasse pela espessura, carregada com o cheiro morno e acre das fêmeas acoradas nos cimos, o Pai dos Homens ainda dilatasse as ventas chatas e soltasse do peito felpudo um grunhido rouco e triste. Mas caminha... As suas pupilas amarelas, onde faísca o Querer, sondam, esbugalhadas, através da ramaria, procuram para além o mundo que deseja e receia, e a que sente já a zoadá violenta, como toda feita de batalha e rancor . E, à maneira que a penumbra das folhagens clareia, vai surgindo, dentro do seu crânio bisonho, como uma alvorada que penetra numa toca, o sentimento das Formas diferentes e da Vida diferente que as anima . Essa rudimentar compreensão só trouxe a nosso Pai venerável turbação e terror .

Será preciso aumentar em 70% a produção de carne para alimentar o mundo em 2050. ALEXANDRA PRADO COELHO 06/08/2013 - 00:00. A produção mundial de aves cresceu 700% em 40 anos. Os humanos sempre consumiram carne, mas nunca na quantidade e à velocidade com que o fazem hoje . Com a população mundial a aumentar e o consumo de carne a crescer, a máquina de produção global foi-se transformando para dar resposta a uma procura crescente . Isto significa que grande parte da produção de cereais que utiliza as terras férteis do planeta destina-se a alimentar animais para que os humanos possam comê-los . Quanto aos animais, são, na sua grande maioria, criados em regime intensivo e alimentados com rações. Este sistema permite (ainda) colocar nos talhos e supermercados muita carne barata, sobretudo suína ou de aves . Mas este é, alertam os especialistas em alimentação, um sistema profundamente ineficaz e insustentável num futuro já relativamente próximo. As previsões das Nações Unidas apontam para que em 2050 a população mundial seja de 9000 milhões de pessoas . Prevê-se, a par disso, que o consumo de carne continue a aumentar (a organização das Nações Unidas para a Agricultura e Alimentação/FAO calcula um aumento de 73). Aliás, nas últimas décadas, a China ultrapassou os Estados Unidos e transformou-se já no maior consumidor de carne a nível mundial, embora se trate aqui sobretudo de carne de porco (o consumo de carne bovina também tem aumentado na China, mas a um ritmo mais lento). Mas os dados mostram um mundo ainda muito desigual no que diz respeito ao consumo de carne per capita . Números da FAO indicam que, em 2009, na Índia, o consumo não chegava aos 5 quilos por pessoa (o que se explica em grande parte por razões culturais), enquanto na China era já de 58 quilos, nos Estados Unidos ultrapassava os 120 quilos e em Portugal rondava os 93,5 . Juntamente com o Brasil e a Índia, a China é um dos países em que a produção de aves está a crescer mais rapidamente . Entre 1967 a 2007, essa produção aumentou, a nível mundial, mais de 700. Tudo isto tem custos ambientais enormes - basta saber que para se ter um quilo de carne de vaca são necessários 15.400 litros de água . O Environmental Working Group, organização norte-americana que criou o Meat Eater's Guide to Climate Change+Health, calculou as emissões de gases com efeito de estufa provocados pela produção de carne e concluiu que a carne de vaca produz 27 quilos de gases por cada quilo de carne consumida, um valor que é o dobro da de porco, quatro vezes mais do que a de galinha e 13 vezes superior às emissões produzidas por proteínas vegetais como o feijão ou as lentilhas . O queijo também tem um peso considerável: são 13,5 quilos de gases por cada quilo consumido . Isto faz com que a produção de gado seja responsável por 18%. É neste cenário de um mundo cada vez mais carnívoro que surge o projecto de Mark Post para criar carne artificial - esta, afirma o investigador holandês, poderá reduzir em 60% a preocupação com esta pegada ambiental tem, aliás, vindo a aumentar . Um relatório elaborado há cerca de dois meses por deputados britânicos apela à população do Reino Unido para que consuma menos carne para ajudar a reduzir as crises de alimentos no mundo. O enorme consumo de carne a que assistimos hoje no mundo desenvolvido contribui para o aumento do preço dos cereais (usados na alimentação dos animais), para a desflorestação e o esgotamento das terras férteis, além de agravar a epidemia da obesidade. Bastaria que as pessoas nos países mais ricos deixassem de comer carne um ou dois dias por semana para se registar um impacto positivo global, defende o relatório. Não é a primeira vez que surgem apelos neste sentido - a campanha internacional Meatless Monday promove precisamente essa ideia . O hambúrguer desenvolvido por Mark Post e ontem apresentado seria, para os carnívoros mais irredutíveis, uma alternativa a essa opção.

Fonte: <http://www.publico.pt/destaque/jornal/sera-preciso-aumentar-em-70-a-producao-de-carne-para-alimentar-o-mundo-em-2050-26922157>[2013-08-06 12:29]

Dupla hélice de ADN fotografada directamente pela primeira vez . Ana Gerschenfeld, 11/01/2013 18:25 . Fig. 1 . A dupla hélice do ADN tornou-se um ícone da ciência do século XX, mas só agora foi possível vê-la mesmo como ela é . As voltas sucessivas da hélice estão assinaladas por setas vermelhas . Foto: Enzo di Fabrizio . Há muito que as imagens da dupla hélice do ADN fazem parte do nosso quotidiano . Mas os pormenores da molécula que contem os genes só eram visualizáveis através de fotografias feitas por raios

X, críticas para os leigos . Agora, quase 60 anos após a descoberta da icónica estrutura, uma equipa italiana oferece-nos a primeira visualização da dupla hélice tal como ela é. A fotografia que ilustra este texto, obtida por microscopia electrónica por Enzo di Fabrizio, do Instituto Italiano de Tecnologia, e colegas, acaba de ser revelada num artigo na revista Nano Letters. Como explica a revista Nature da última quinta-feira, os cientistas espalharam gotículas que continham o ADN de um vírus que infecta as bactérias à superfície de pequenas “pastilhas” de silício . As pastilhas tinham sido previamente “gravadas”, de forma a ficarem peçadas de micrométricos buracos e de “pilares” cilíndricos. Quando as gotas secaram, as fibras de ADN ficaram bem esticadas entre os diversos “pilares” e suspensas por cima dos buracos . E os cientistas puderam então fotografar o ADN espreitando com um microscópio electrónico através dos buracos do silício. Ainda não é uma imagem da dupla hélice no seu meio natural, fisiológico, uma vez que o ADN adopta uma forma algo diferente ao ser desidratado . Mas é a primeira vez que se consegue ver as voltas que a molécula dá, à maneira de uma escada de cordas enrolada sobre si própria. Os cientistas puderam também medir com precisão a distância entre voltas consecutivas da hélice (assinaladas pelas setas vermelhas na fotografia): 2,7 milionésimos de metro, que se sabe corresponder à dita configuração “seca” do ADN.

Fonte: <http://www.publico.pt/ciencia/noticia/moleculadeadnfotografadadirectamentepelaprimeiravez1580392> [2013/08/04]

Fernando António Nogueira Pessoa (Lisboa, 13 de Junho de 1888 — Lisboa, 30 de Novembro de 1935), mais conhecido como Fernando Pessoa, foi um poeta, filósofo e escritor português . É considerado um dos maiores poetas da Língua Portuguesa, e da Literatura Universal, muitas vezes comparado com Luís de Camões. O crítico literário Harold Bloom considerou a sua obra um "legado da língua portuguesa ao mundo" . Por ter sido educado na África do Sul, para onde foi aos seis anos em virtude do casamento de sua mãe, Pessoa aprendeu perfeitamente o inglês, língua em que escreveu poesia e prosa desde a adolescência. Das quatro obras que publicou em vida, três são na língua inglesa. Fernando Pessoa traduziu várias obras inglesas para português e obras portuguesas (nomeadamente de António Botto e Almada Negreiros) para inglês . Ao longo da vida trabalhou em várias firmas comerciais de Lisboa como correspondente de língua inglesa e francesa . Foi também empresário, editor, crítico literário, jornalista, comentador político, tradutor, inventor, astrólogo e publicitário, ao mesmo tempo que produzia a sua obra literária em verso e em prosa. Como poeta, desdobrou-se em múltiplas personalidades conhecidas como heterónimos, objeto da maior parte dos estudos sobre sua vida e sua obra . Centro irradiador da heteronímia, auto-denominou-se um "drama em gente" .

[http://pt.wikipedia.org/wiki/Fernando\\_pessoa](http://pt.wikipedia.org/wiki/Fernando_pessoa) [2013-06-13]

Bruno Cortez diz que Benfica vai “mostrar a sua força” no jogo com o São Paulo . LUSA 02/08/2013 - 15:23 . Defesa esquerdo vai defrontar a ex-equipa e estrear-se a jogar no Estádio da Luz . TÓPICOS: Futebol nacional, SL Benfica . O defesa brasileiro Bruno Cortez afirmou que a equipa de futebol do Benfica vai “mostrar a sua força” no jogo particular de sábado com o São Paulo, da 6.ª edição da Eusébio Cup . “Vai ser um jogo muito bom, com duas excelentes equipas, mas estamos em casa e temos o apoio de todos os nossos adeptos . Tenho a certeza de que vamos dar o nosso melhor e mostrar a nossa força”, disse o defesa, citado no site oficial do Benfica . Bruno Cortez, de 26 anos, chegou ao Benfica esta época por empréstimo do clube brasileiro que será o adversário no jogo de sábado, afirmando que vai ser especial jogar pela primeira vez no Estádio da Luz . “É um jogo especial, no sentido em que vou pisar, pela primeira vez, o Estádio da Luz juntamente com a equipa do Benfica . É muito importante entrarmos determinados e focados no que o treinador nos pede, para fazermos um bom jogo”, afirmou . O lateral esquerdo deixou ainda a garantia de que a equipa vai mostrar “entrega, força e determinação” na partida frente ao São Paulo, agendada para as 19h45 de sábado, no Estádio da Luz .

Fonte: <http://www.publico.pt/desporto/noticia/bruno-cortez-diz-que-benfica-vai-mostrar-a-sua-forca-no-jogo-com-o-sao-paulo-1602061> [2013-08-06]

Morreu o homem que viveu mais tempo PÚBLICO e AGÊNCIAS 12/06/2013 - 16:32 Jiroemon Kimura tinha 116 anos e era também a pessoa mais velha do mundo na actualidade. A pessoa mais velha do mundo, Jiroemon Kimura, morreu na manhã desta quarta-feira aos 116 anos, após mais de um mês hospitalizado devido a uma pneumonia . O japonês, que detinha o título no Livro de Recordes Guinness desde Dezembro, tinha celebrado o último aniversário em Abril. Jiroemon Kimura foi reconhecido como a pessoa mais velha do mundo no final do ano passado, quando morreu uma norte-americana do Iowa aos 115 anos . Apesar da longevidade, Kimura não ficará no Livro de Recordes Guinness como a pessoa que mais anos viveu . Esse recorde pertence até hoje à francesa Jeanne Calment, que morreu em 1997 aos 122 anos. Kimura celebrou o 116.º aniversário em Abril último e nesse dia recebeu um vídeo do primeiro-ministro japonês, Shinzo Abe, a congratulá-lo. Um ano antes, no dia em que comemorou 115 anos, Kimura dizia que mantinha a mente activa a aprender inglês . Questionado sobre a sua longevidade, o japonês disse que o truque era apanhar ar fresco e a luz do sol . “Estou sempre a olhar o céu . É assim que sou”. Pai de sete filhos, avô de 14 crianças, bisavô de 25 outras e tetravô de 15, trabalhou durante 40 anos num posto de correios . Depois de se reformar, dedicou-se à agricultura, actividade que manteve até aos seus 90 anos.

O Japão tem mais de 50 mil centenários, segundo dados recolhidos pelo Governo em 2011. Na cidade onde Kimura vivia, Kyotango, estão registadas dezenas de outras pessoas com 100 ou mais anos . Devido à presença desta população centenária, o município local decidiu iniciar um projecto para analisar a sua dieta alimentar e chegar a dados que ajudem a explicar a sua longevidade.

<http://www.publico.pt/mundo/noticia/morreu-a-pessoa-mais-velha-do-mundo-1597181> [2013-06-13]

Papa paga “religiosamente” quotas do clube San Lorenzo LUSA 05/08/2013 - 22:35 O cartão de sócio de Jorge Mario Bergoglio AFP TÓPICOS: Futebol internacional, Futebol, Papa Francisco O papa Francisco paga “religiosamente” a quota mensal do clube de futebol argentino de que é sócio desde a infância, o San Lorenzo de Almagro, revelou nesta segunda-feira o vice-presidente do clube e estrela televisiva, Marcelo Tinelli. “O papa paga a quota religiosamente todos os meses”, disse Tinelli na Fox Sports. “Eu questionava-me ‘pagará agora’”. “Nem se deve lembrar...” “Mas não, confirmaram-me que está em dia e que todos os meses paga religiosamente a quota”, acrescentou. O vice-presidente do San Lorenzo afirmou que se emociona com tudo o que Francisco faz e que espera ansiosamente por falar com ele antes do jogo que a seleção argentina disputará contra Itália a 14 de Agosto, em Roma, em honra do novo Papa. Francisco é um fervoroso adepto do San Lorenzo de Almagro e o clube atribuiu-lhe o cartão de sócio honorário número 88.235 em Março de 2008, depois de o então arcebispo de Buenos Aires ter oficiado uma missa na capela do clube. A eleição do arcebispo Jorge Bergoglio como novo pontífice da Igreja Católica em Março passado fez disparar a popularidade do clube argentino e os jogadores vestiram camisolas com a imagem do seu sócio mais famoso no primeiro jogo disputado após a sua escolha.

Fonte: <http://www.publico.pt/desporto/noticia/papa-paga-religiosamente-quotas-do-clube-san-lorenzo-1602303#0> [2013-08-06]



## **Appendix B**

### **Diretivas de classificação e anotação**



*Diretivas de classificação e anotação de corpora  
para a extração de relações entre eventos*

Viviana Cabrita, Jorge Baptista, Nuno Mamede

Instituto Superior Técnico, INESC-ID Lisboa  
Rua Alves Redol, 9 – Lisboa – Portugal

July 12, 2013

## Table of Contents

1	Introdução .....	1
1.1	Eventos .....	1
1.2	Conectores .....	2
1.2.1	Coordenação .....	2
1.2.2	Subordinação .....	3
1.3	Modificadores temporais .....	4
2	Extração das Relações .....	5
2.1	Relações consideradas .....	5
2.2	Relações directas e indirectas entre eventos .....	8
2.3	Relação temporal .....	9
2.4	Relação de agrupamento .....	10
3	Ordenação .....	12
3.1	Tipos de ordenação considerados .....	12
3.2	Critério de ordenação .....	13
3.3	Interpretação das relações .....	14

## 1 Introdução

Este documento apresenta o conjunto de conceitos-chave, os critérios linguísticos e o formalismo adoptado na tarefa de anotação de relações entre eventos a serem extraídas pela cadeia de processamento de língua natural STRING [5][4], desenvolvida pelo Laboratório de Sistemas de Língua Falada (L<sup>2</sup>F) do Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID).

A tarefa de extração de eventos é desenvolvida na cadeia de processamento, mas não como parte deste trabalho, considerando-se que as dependências relativas à identificação dos mesmos já são extraídas *à priori* [2], em maioria, na caracterização de entidades mencionadas [3][7]. Deste modo, neste documento, define-se uma vez mais o conceito de evento e procede-se à descrição das relações a extrair entre os eventos já identificados. As relações a anotar são as relações de coordenação e subordinação estabelecidas entre os eventos e as relações de ordem obtidas através da análise dos mesmos e das dependências. Pretende-se ainda anotar as relações estabelecidas entre os eventos e os modificadores temporais por estas auxiliarem a deteção da ordem dos eventos envolvidos.

Nesta secção descreve-se, de seguida, os eventos, conetores e modificadores temporais a considerar para a anotação. Na secção 2, procede-se com a descrição e os critérios de anotação das relações que auxiliam a ordenação dos eventos, descritos na secção 3.

### 1.1 Eventos

Considera-se *evento* uma expressão linguística que, num texto, representa um predicado semântico, isto é, seleciona um certo número de argumentos e impõe restrições distribucionais (sintáticas e semânticas) ao preenchimento lexical dessas posições argumentais. Assim, um *verbo pleno* (ou distribucional), como como *entrevistar*, é um *evento*.

(1a) O jornalista *entrevistou* o jogador

Do mesmo modo, também para os *nomes predicativos*, como *entrevista*, considera-se que se trata de *eventos*, tanto nas estruturas em que aparecem construídos com os respectivos verbos-suporte ((2a), neste caso com o verbo-suporte *fazer*), como nas expressões em que o nome aparece como argumento de outro predicado, tanto acompanhado dos respectivos argumentos (2b), como isolado (2c):

(2a). O jornalista *fez uma entrevista* ao jogador

(2b). O Pedro *leu a entrevista* do jornalista ao jogador

(2c). O Pedro *leu a entrevista*

Chama-se a atenção para o facto de a noção de *evento* aqui utilizada não corresponder precisamente ao conceito corriqueiro de “evento/acontecimento”, já que abrange não apenas esse tipo de predicados, mas também outros, que expressam variados significados, como, por exemplo, as expressões de medida (e.g. *pesar*, *medir*, *valer*), estados psicológicos (*irritar*, *alegrar*, *entristecer*), causativos (*implicar*, *causar*, *resultar*), etc.

As construções adjetivais predicativas são igualmente eventos, no sentido em que aqui se usa o termo. Contudo, na presente campanha de anotação de corpora (2013), os *adjetivos predicativos* não serão considerados para efeitos de anotação.

No caso dos *advérbios*, enquanto predicados que modificam outros elementos, estes entram igualmente na definição de eventos acima apresentada. Contudo, há que distinguir, nesta classe, diferentes tipos de advérbios – e respectivas diferenças na função sintática que desempenham. Por essa razão, na presente campanha de anotação de corpora, os *advérbios* não serão considerados para efeitos de anotação. Ressalve-se, porém, (a) os chamados *advérbios conjuntivos* (*contudo*, *por conseguinte*, etc.), que estabelecem uma ligação entre duas frases ou orações e que, por essa razão, são integrados no conjunto dos *conetores*,

apresentados na secção 1.2; e, ainda, (b) os *advérbios de tempo*, tratados como *modificadores temporais*, e descritos na secção 1.3.

No caso das construções preposicionais locativas (e.g. *sobre*, *em cima de*, *através de*), estas são consideradas eventos de acordo com a definição dada acima, conquanto não serão considerados para efeitos de anotação na presente campanha.

## 1.2 Conectores

Diversos elementos lexicais podem servir de elementos conectores entre eventos:

- as conjunções coordenativas aditivas (e.g. *e*) e disjuntivas (e.g. *ou*)
- as conjunções coordenativas adversativas (e.g. *mas*)
- as conjunções subordinativas (e.g. *enquanto*)
- as preposições (e.g. *de*)
- advérbios conjuntivos (e.g. *por conseguinte*)

Os conectores discursivos são elementos linguísticos que asseguram a coesão inter- e intrafrásica, ligando frases, orações (3a) ou elementos da mesma oração (3b). Para esta campanha, no entanto, apenas são considerados os conectores que ligam orações ou elementos *dentro da mesma frase*. Nesse sentido, o *advérbio conjuntivo por conseguinte* apenas é considerado em (3c), mas não em (3d),

- (3a). *Estava mau tempo e decidimos ficar em casa.*
- (3b). *O trabalho do Pedro foi feito à semelhança do projetado.*
- (3c). *Estava mau tempo esta manhã, por conseguinte, decidimos ficar em casa.*
- (3d). *Estava mau tempo esta manhã. Por conseguinte, decidimos ficar em casa.*

### 1.2.1 Coordenação

No caso de coordenação de eventos, formam-se *agrupamentos*, permitindo depois relacionar esses agrupamentos com outros eventos presentes na frase. Assim, por exemplo:

- (4a). *O trabalho do Pedro e a apresentação da Ana foram corrigidos pelo Rui*

Considera-se que há um agrupamento dos eventos coordenados *trabalho* e *apresentação* e que, por sua vez, este agrupamento está relacionado com o evento *corrigidos*.

Não se considera os conectores quando estes não permitem associar dois ou mais eventos, como em (5a), em que se verifica a coordenação de dois nomes próprios; nem em (5b), em que se dá a coordenação de dois predicados adjetivais – excluídos da presente campanha de anotação; nem, ainda, em (5c), em que se pode analisar a frase como o resultado da elipse de uma segunda ocorrência do verbo *falar*:

- (5a). *Vi o João e a Maria.*
- (5b). *Ele é traquinas mas amoroso.*
- (5c). *Os miúdos falaram bem e [falaram] depressa.*

A coordenação de eventos por meio de conjunções coordenativas aditivas (e.g. *e*) e disjuntivas (e.g. *ou*) permitem que os eventos assim ligados possam ter argumentos idênticos (correferentes) que, por essa razão, se podem omitir (elipse). Como a coordenação pode ocorrer em qualquer nível de análise sintática da frase, tal torna-a um fenómeno particularmente difícil de tratar. Eis alguns exemplos ilustrativos da complexidade do problema:

- (6a). *O Pedro discutiu o trabalho e a apresentação.*
- (6b). *O Pedro analisou e apresentou o trabalho.*
- (6c). *O Rui vai ter mesmo de fazer o trabalho ou arranjar quem lho faça.*

(6d). Não sei *se*, esta manhã, a Rita já sabia [o que te contei ontem] *ou* [se] acabou por descobrir mais tarde o que te contei ontem.

Nestas frases observam-se várias reduções:

- Em (6b): o sujeito de *apresentou* reduz-se a zero, por ser correferente do sujeito do primeiro verbo da coordenação, *apresentou* (anáfora zero); por outro lado, apenas o complemento direto (v.g. *o trabalho*) do verbo *apresentou* se encontra expresso, sendo, porém, possível inferir a omissão de uma primeira ocorrência desse mesmo constituinte como complemento direto do verbo *analisou* (catáfora).
- Em (6c): a coordenação dos dois verbos por *ou* leva à redução da cadeia de auxiliares *vai ter de* diante do segundo, *arranjar*;
- Em (6d): Na coordenação disjuntiva (*ou*) das duas orações integrantes, dependentes de *sei* e introduzidas por *se*, a segunda conjunção subordinativa integrante *se* encontra-se reduzida; tal como em (6b), também se observa a redução da primeira ocorrência de *o que te contei* dependente de *sabia*.

Outro caso complexo de coordenação resulta do emprego de conjunções formadas por duas expressões descontínuas, e.g. *não só...mas também*; (7a). Neste caso, o elemento conetor que figura na relação **GROUPRELATION** deveria ser a conjunção (complexa e descontínua). Note-se que em (7b), a coordenação liga duas orações subordinadas causais, introduzidas por *porque*.

(7a). *Não só* a Rita já sabia disso *como também* andou a contá-lo a toda a gente.

(7b). A Rita já sabia disso *não só porque* eu lho contei *mas também porque* toda a gente lera a carta.

De um modo geral, a cadeia STRING [5] aborda o problema da coordenação de uma perspectiva estritamente local, estabelecendo apenas a dependência **COORD** entre NPs e alguns casos de PPs e verbos que se encontram diretamente ligados à conjunção coordenativa [6]. Contudo, independentemente do problema de análise (parsing), na anotação dever-se-á assinalar no corpus esses casos de agrupamento, se ocorrerem, a fim de poderem constituir um marco a alcançar.

Nesse sentido, definem-se as seguintes orientações de anotação:

- tanto os nomes predicativos (6a) como os verbos plenos coordenados (6b) deverão constituir agrupamentos; no caso dos verbos, estes agrupamentos permitirão, numa segunda fase, propagar aos respectivos argumentos (eventos) reduzidos as relações elididas;
- uma vez que apenas se considera a cabeça de uma cadeia verbal, os casos de redução de auxiliares em verbos coordenados (6c) serão tratados como se fossem verbos simples;
- no caso das conjunções descontínuas (7a), representa-se no campo **<conetor>** da dependência de agrupamento ambas as partes da conjunção, separadas por underscore ‘\_’;
- no caso da coordenação de orações subordinadas, independentemente de haver apenas uma conjunção subordinativa expressa (6d) ou estar também presente uma segunda ocorrência da mesma conjunção (7b), a relação de agrupamento é representada tomando como dois argumentos suplementares, para cada uma das conjunções subordinativas

Nem todos os casos de coordenação são, contudo, tão complexos como os exemplos anteriores. Uma das situações mais frequentes de coordenação é o caso das *enumerações*. Numa enumeração, como em (8a), pode considerar-se que cada elemento é ligado ao seguinte pela conjunção *e*, mas apenas a última conjunção coordenativa aditiva permanece explícita, enquanto todas as anteriores se reduzem, o que é assinalado na escrita por meio de vírgulas. A vírgula pode, pois, ser considerada nestes casos como um conetor:

(8a). *Cheguei, vi e venci*

### 1.2.2 Subordinação

Consideram-se, nesta campanha, as conjunções subordinativas *causais* (e.g. *porque, por*), *comparativas* (*como*), *condicionais* (*desde que, se*), *concessivas* (*embora, conquanto, apesar de*), *proporcionais* (*à medida*

que), *consecutivas* (que, em consequência de),  *finais* (para, a fim de), *topic* (em relação a, sobre), *temporais* (quando, enquanto, antes de) e *completivas* (ou integrantes) (que, se).

Recorde-se que a distinção entre preposições e conjunções nem sempre é evidente, podendo as preposições (simples e compostas) funcionar muitas vezes como conjunções, ligando duas orações na mesma frase, como sucede em (9a), onde *de* tem um valor *causal*. Por outro lado, esta preposição pode ligar expressões temporais a nomes de eventos, como acontece em (9b); neste caso, *de* funciona efetivamente como um elemento de ligação, que estabelece a modificação temporal, mas, por ser desprovido de significado, não é considerado um conetor.

(9a). Ele está cansado *de* correr.

(9b). Ele faltou ao jogo *de* ontem.

Já em (10a) trata-se de uma preposição que liga o verbo principal ao seu argumento oracional (regência); enquanto que em (10b) *de* é apenas de um mero elemento gramatical de ligação entre o verbo auxiliar modal *ter* e o verbo principal da oração *ler*. Nestes dois casos, a preposição é desprovida de significado, razão pela qual não deve ser considerada como um conetor de eventos.

(10a). Ele gostou *de* ler o livro.

(10b). Ele teve *de* ler a acta.

### 1.3 Modificadores temporais

O terceiro conjunto de elementos abrangidos nesta campanha é formado pelos advérbios de tempo, que funcionam como modificadores temporais de um evento e que podem exprimir diferentes tipos de modificação, nomeadamente, (i) localizar um evento na linha temporal (11a), (ii) quantificar a duração de um evento (duração) (11b), (iii) definir o valor aspetual de um evento (11c), (iv) em particular a sua frequência (11d):

(11a). Ele anda *atualmente* muito contente

(11b). Ele fez isso *durante* muito tempo

(11c). Ele está *constantemente* a interromper-me.

(11d). Ele faz ginástica *semanalmente*.

Para a ordenação de eventos, apenas os modificadores do primeiro tipo são relevantes, conquanto os modificadores aspetuais e de frequência possam contribuir para relacionar eventos entre si.

Como já se disse atrás, certos advérbios e outras expressões de tempo podem modificar diretamente um verbo (12a) ou ligar-se por meio da preposição *de* a um nome de evento (12b):

(12a). Ele faltou ao jogo ontem.

(12b). Ele faltou ao jogo *de* ontem.

Apesar de o significado das duas frases acima ser semelhante, as relações de modificação temporal são distintas, com *ontem* a modificar o verbo *faltar* em (12a), enquanto que em (12b) modifica o nome de evento *jogo*.

Finalmente, certos modificadores adverbiais de tempo, desde que de natureza semântica diferente, podem modificar simultaneamente o mesmo predicado, como sucede em (13a)-(13b):

(13a). Ele acordou hoje cedo.

(13b). Ele acordou cedo hoje.

Neste caso, está-se perante dois adverbiais distintos, autónomos, que deverão, por isso, constituir diferentes relações de modificador temporal do predicado *acordar*, na linha dos critérios de delimitação/segmentação das expressões temporais [1] <sup>1</sup>.

<sup>1</sup> Apesar desta análise, a cadeia STRING ainda não faz de forma perfeita a segmentação das expressões temporais, agregando muitas vezes estes adverbiais num único chunk.

Tal situação é diferente dos casos em que um advérbio modifica outro adverbial, como, por exemplo, em (14a)-(14b):

(14a). Ele acordou muito cedo.

(14b). Ele acordou logo de manhã.

No primeiro exemplo, o advérbio de tempo *cedo* é quantificado pelo advérbio de quantidade *muito*. No segundo caso, o advérbio de tempo *de manhã* encontra-se sob o escopo do advérbio focalizador *logo*. Em ambos os casos, estas combinatórias foram um único complemento adverbial e é esse constituinte complexo que modifica o verbo *acordar*. Do ponto de vista da representação, essa modificação exerce-se entre a cabeça do adverbial (*cedo* e *de manhã*) e o verbo.

## 2 Extração das Relações

### 2.1 Relações consideradas

São consideradas três formas de relação entre eventos:

- relação direta (`DIRECTRELATION`);
- relação indireta (`INDIRECTRELATION`); e
- relação de agrupamento (`GROUPRELATION`);

a que se acrescenta:

- relação temporal (`TIMERELATION`), que liga um evento a uma expressão temporal;

#### Relação direta (`DIRECTRELATION`)

Este tipo de dependência, na ausência de *conectores*, é descrita como uma *relação direta* (`DIRECTRELATION`) e, considerando que um dos eventos é subordinado em relação ao outro, o evento principal, pode ser representada do seguinte modo:

`EVENT_DIRECTRELATION(<evento principal >, <evento subordinado >)`

Por exemplo, uma oração subordinada integrante infinitiva (15a) ou um evento expresso por um nome predicativo (15b) como complemento direto de um verbo estão ligados *diretamente* ao predicado de que dependem:

(15a). O Pedro preferia jantar cedo.

```
EVENT_LEX(preferia, outro)
EVENT_LEX(jantar, outro)
EVENT_DIRECTRELATION(preferia, jantar)
EVENT_TIMERELATION(jantar, cedo)
```

(15b). O Pedro preferiu a leitura de um bom livro.

```
EVENT_LEX(preferiu, outro)
EVENT_LEX(leitura, outro)
EVENT_DIRECTRELATION(preferia, leitura)
```

(15c). O rapaz cujo livro roubei foi ao supermercado.

```
EVENT_LEX(roubei, outro)
EVENT_LEX(foi, outro)
EVENT_DIRECTRELATION(foi, roubei)
```

(15d). O rapaz, sabendo da notícia, foi ao supermercado.

```
EVENT_LEX(sabendo, outro)
EVENT_LEX(foi, outro)
EVENT_DIRECTRELATION(foi, sabendo)
```

(15e). O livro que o Pedro comprou foi-lhe dado pela Maria.

```
EVENT_LEX(dado, outro)
EVENT_LEX(comprou, outro)
EVENT_DIRECTRELATION(dado, comprou)
```



### Relação indireta (`INDIRECTRELATION`)

Noutros casos, este tipo de dependência é expressa de forma indireta (`INDIRECTRELATION`) por meio de *conectores* tais como conjunções ou preposições. Uma vez que estes conectores explicitam o nexos semântico entre os eventos que relacionam, é necessário representá-los, o que se faz do seguinte modo:

```
EVENT_INDIRECTRELATION(<evento principal>,<evento subordinado>, <conector>)
```

Assim, por exemplo, nas frases (16a) e (16b), as conjunções subordinativas *embora* (adversativa) e *porque* (causal), respetivamente, deverão ser extraídas na representação das relações entre os eventos da oração subordinante e da subordinada:

(16a). O Pedro almoçava tarde *embora* jantasse cedo.

```
EVENT_LEX(almoçava,outro)
EVENT_LEX(jantasse,outro)
EVENT_INDIRECTRELATION(almoçava,jantasse,embora)
EVENT_TIMETRELATION(almoçava,tarde)
EVENT_TIMETRELATION(jantasse,cedo)
```

(16b). O Rui comentou a entrevista *porque* detestava o fulano.

```
EVENT_LEX(comentou,outro)
EVENT_LEX(entrevista,outro)
EVENT_LEX(detestava,outro)
EVENT_DIRECTRELATION(comentou,entrevista)
EVENT_INDIRECTRELATION(comentou,detestava,porque)
```

As *conjunções coordenativas adversativas*, e.g. *mas*, não estabelecem relações de subordinação mas, para efeitos de anotação, serão incluídas neste grupo (17a)

(17a). O Pedro deita-se muito cedo *mas* levanta-se de madrugada.

```
EVENT_LEX(deita,outro)
EVENT_LEX(levanta,outro)
EVENT_INDIRECTRELATION(deita,levanta,mas)
EVENT_TIMETRELATION(deita,cedo)
EVENT_TIMETRELATION(levanta,de madrugada)
```

### Relação de agrupamento (`GROUPRELATION`)

Pelo contrário, considera-se que as conjunções *coordenativas aditivas*, e.g. *e*, e as *disjuntivas*, e.g. *ou*, permitem a associação de eventos a dependências que lhes são comuns. Assim, considera-se a existência de dependências do tipo *agrupamento* (`GROUPRELATION`) e indica-se os eventos pela ordem em que os mesmos aparecem na frase:

```
EVENT_GROUPRELATION(<1° evento>, <2° evento>, <conjunção>)
```

Nas frases (18a) e (18b), temos exemplos de agrupamento de eventos:

(18a). Ele *almoçava e jantava* a horas certas.

```
EVENT_LEX(almoçava,outro)
EVENT_LEX(jantava,outro)
EVENT_GROUPRELATION(almoçava,jantava,e)
EVENT_TIMETRELATION(deita,a horas certas)
```

(18b). O Rui *comentou e elogiou* a entrevista publicada.

```
EVENT_LEX(comentou,outro)
EVENT_LEX(elogiou,outro)
EVENT_LEX(entrevista,outro)
EVENT_LEX(publicada,outro)
```

```

EVENT_GROUPRELATION(comentou,elogiou,e)
EVENT_DIRECTRELATION(elogiou,entrevista)
EVENT_DIRECTRELATION(publicada,entrevista)

```

Em (18a), o advérbio *a horas certas* modifica ambos os eventos *almoçava* e *jantava*. Em (18a), os eventos *comentou* e *elogiou*, têm ambos como argumento o evento *entrevista*. Em rigor, considera-se que há uma instância de *entrevista* sob *comentou* (objeto) e de *Rui* sob *elogiou* (sujeito) repetidos sob a coordenação e que, por essa razão, foram reduzidos. Neste sentido, a criação de agrupamentos, simplifica a descrição, descrevendo-a de forma mais próxima à da organização textual dos elementos envolvidos.

Note-se que este tipo de associação não impede que se identifique posteriormente uma relação de ordem entre os eventos agrupados. Considere-se, por exemplo, as frases (19a) e (19b):

- (19a). *Vamos hoje ao cinema ou vemos o filme amanhã?*  
(19b). *O Luís foi à exposição e comprou um postal.*

Na frase (19a), o evento *vamos* é identificado como *anterior* ao evento *vemos* devido ao facto de que cada um dos advérbios *hoje* e *amanhã* modifica de forma independente um destes verbos. Em (19b), por outro lado, embora também se trate de um agrupamento, está-se perante uma *sequência* de eventos, existindo uma ordenação entre os mesmos com base na ordem com que os eventos aparecem na frase.

### Relação temporal (TIMERELATION)

Finalmente, cada evento pode também ser associado aos advérbios temporais que o modificam/localizam no eixo temporal (TIMERELATION), representando-se este tipo de dependência do seguinte modo:

```

EVENT_TIMERELATION(<evento>, <advérbio temporal>)

```

No exemplo (19a), deverão, pois, ser extraídas duas relações temporais, ligando cada advérbio ao verbo que modifica.

Por último, independentemente das dependências consideradas, quando o evento é identificado a partir de uma cadeia verbal, extrai-se apenas o verbo principal. Assim, por exemplo, nas frases seguintes (20a)-(20b):

- (20a). *O Pedro começou a ler o livro hoje.*  

```

EVENT_LEX(ler,outro)
EVENT_TIMERELATION(ler,hoje)

```

(20b). *O Rui adora ficar a preguiçar na cama de manhã.*  

```

EVENT_LEX(preguiçar,outro)
EVENT_TIMERELATION(preguiçar,de manhã)

```

deverão ser extraídos os eventos *ler* e *preguiçar*, já que estes são os verbos principais das cadeias verbais *começou a ler* e *ficar a preguiçar*, respetivamente.

Descreve-se nas secções seguintes, o formalismo adoptado e os critérios de anotação para os diferentes tipos de relações, começando-se pelas relações direta e indireta 2.2, passando depois às relações temporais 2.3 e, finalmente, as relações de agrupamento, que será apresentada em 2.4. Recorde-se a representação das dependências a considerar:

```

EVENT_DIRECTRELATION(<evento principal>, <evento subordinado>)
EVENT_INDIRECTRELATION(<evento subordinante>, <evento subordinado>, <conector>)
EVENT_GROUPRELATION(<1º evento>, <2º evento>, <conector>)
EVENT_TIMERELATION(<evento>, <expressão temporal>)

```

## 2.2 Relações directas e indirectas entre eventos

Relativo à dependência `EVENT_DIRECTRELATION`, considera-se que dois eventos apresentam este tipo de relação directa quando:

- Ambos os eventos são identificados a partir de formas verbais, uma finita e outra no infinitivo, considerando-se a primeira o evento principal e sendo a segunda ou o sujeito (21a) ou o complemento directo da primeira (21b):

(21a). *Irrita-me solenemente ir ao cinema à noite.*

```
EVENT_LEX(irrita,outro)
EVENT_LEX(ir,outro)
EVENT_DIRECTRELATION(preferir,ir)
EVENT_TIMERELATION(ir,à noite)
```

(21b). *O Pedro preferiu ir ao cinema à noite.*

```
EVENT_LEX(preferiu,outro)
EVENT_LEX(ir,outro)
EVENT_DIRECTRELATION(preferir,ir)
EVENT_TIMERELATION(ir,à noite)
```

- Um dos eventos é identificado a partir de uma forma nominal (um nome predicativo) e ou outro a partir de uma forma verbal participial. Nesse caso, pode tratar-se ou de uma construção passiva (22a), em que o participio se encontra auxiliado por um verbo copulativo (recorde-se que os verbos auxiliares não são considerados eventos); ou do resultado da redução de uma construção passiva, em que o participio se encontra em posição adnominal (22b), isto é, junto do nome de evento que modifica. Em ambos os casos, o participio é o evento principal e o nome predicativo o evento subordinado:

(22a). *A entrevista foi publicada ontem.*

```
EVENT_LEX(entrevista,outro)
EVENT_LEX(publicada,outro)
EVENT_DIRECTRELATION(publicada,entrevista)
EVENT_TIMERELATION(publicada,ontem)
```

(22b). *A entrevista publicada ontem era falsa.*<sup>2</sup>

```
EVENT_LEX(entrevista,outro)
EVENT_LEX(publicada,outro)
EVENT_DIRECTRELATION(publicada,entrevista)
EVENT_TIMERELATION(publicada,ontem)
```

Quando se está perante conjunções ou preposições, extrai-se a dependência `EVENT_INDIRECTRELATION`. O evento que é introduzido pela conjunção ou a preposição é considerado o evento subordinado, independentemente da ordem linear por que cada oração aparece no discurso. Assim, por exemplo, a inversão da ordem das orações nas frases seguintes não influencia a ordem dos eventos na extração da dependência:

(23a). *O Rui lia um livro enquanto a Joana escrevia uma carta.*

```
EVENT_LEX(lia,outro)
EVENT_LEX(escrevia,outro)
EVENT_INDIRECTRELATION(lia, escrevia, enquanto)
```

(23b). *Enquanto a Joana escrevia uma carta, o Rui lia um livro.*

```
EVENT_LEX(lia,outro)
EVENT_LEX(escrevia,outro)
EVENT_INDIRECTRELATION(lia, escrevia, enquanto)
```

<sup>2</sup> Note-se que não se extrai uma dependência a partir do adjetivo predicativo *falsa* presente na frase (22b) (ver secção 1.1).

Recorde-se que as conjunções coordenativas adversativas (e.g. **mas**), ainda que façam parte dos processos de coordenação, são extraídas como se de uma subordinação se tratasse, não dando origem à formação de agrupamentos (secção 2.4).

(24a). **O carro bateu contra o muro, mas ninguém se feriu.**

```
EVENT_LEX(bateu,outro)
EVENT_LEX(feriu,outro)
EVENT_INDIRECTRELATION(bateu, feriu, mas)
```

### 2.3 Relação temporal

Numa dependência temporal é necessário identificar o evento e o advérbio temporal associado ao mesmo, sendo representada do seguinte modo:

```
EVENT_TIMERELATION(<evento>, <advérbio>)
```

Os advérbios temporais podem aparecer:

- como modificador de toda a frase (25a);
- como modificadores de um predicado verbal (25b);
- como modificadores de um predicado nominal, e ligados a ele por meio de preposição (25c);

(25a). **Atualmente, as pessoas preferem ver um filme a ir ao cinema, mas este estudo diz o contrário**

(25b). **O Rui comentou ontem a publicação da entrevista.**

(25c). **O Rui comentou a entrevista de ontem.**

Quando um advérbio modifica toda a frase, extraem-se relações temporais entre o advérbio e *todos* os eventos identificados a partir dos *verbos* ou nomes predicativos presentes na frase (26a). Nos restantes casos, o modificador temporal é extraído com o evento (verbo ou nome predicativo) que modifica (26b)-(26c):

(26a). **Atualmente, as pessoas preferem ver um filme a ir ao cinema, mas este estudo diz o contrário**

```
EVENT_LEX(preferem,outro)
EVENT_LEX(ver,outro)
EVENT_LEX(ir,outro)
EVENT_LEX(estudo,outro)
EVENT_LEX(diz,outro)
EVENT_DIRECTRELATION(preferem, ver)
EVENT_INDIRECTRELATION(preferem, ir, a)
EVENT_INDIRECTRELATION(preferem, diz, mas)
EVENT_DIRECTRELATION(diz, estudo)
EVENT_TIMERELATION(preferem, Atualmente)
EVENT_TIMERELATION(ver, Atualmente)
EVENT_TIMERELATION(ir, Atualmente)
EVENT_TIMERELATION(estudo, Atualmente)
EVENT_TIMERELATION(diz, Atualmente)
```

(26b). **O Rui comentou ontem a publicação da entrevista.**

```
EVENT_LEX(comentou,outro)
EVENT_LEX(publicação,outro)
EVENT_LEX(entrevista,outro)
EVENT_DIRECTRELATION(comentou, publicação)
EVENT_INDIRECTRELATION(publicação, entrevista, de)
EVENT_TIMERELATION(comentou, ontem)
```

(26c). **O Rui comentou a entrevista de ontem.**

```
EVENT_LEX(comentou,outro)
EVENT_LEX(entrevista,outro)
EVENT_DIRECTRELATION(comentou, entrevista)
EVENT_TIMERELATION(entrevista, ontem)
```

## 2.4 Relação de agrupamento

Recorde-se que as conjunções coordenativas copulativas e as disjuntivas permitem que se formem agrupamentos entre os eventos envolvidos, representando-se um agrupamento do seguinte modo:

```
EVENT_GROUPRELATION(<1° evento>, <2° evento>, <conjunção>)
```

Estes agrupamentos indicam que os eventos assim relacionados estão num mesmo nível de análise sintática, podendo por isso estar sujeitos ou impôr o mesmo conjunto de dependências. É essa redundância que explica, pois, a possibilidade de redução de elementos repetidos sob a coordenação:

(27a). *Ele almoçava e jantava a horas certas.*

(27b). *O Rui comentou e elogiou a entrevista.*

Por exemplo, na frase (27a), o advérbio *a horas certas* modifica sintaticamente o verbo *jantava*. Contudo, porque este se encontra coordenado com *almoçava* (e porque este último verbo não tem já nenhum outro modificador da mesma natureza), é possível interpretar a frase como apresentando igualmente uma modificação temporal entre *almoçava* e *a horas certas*. De modo semelhante, na frase (27b), a coordenação dos dois verbos, *comentou* e *elogiou*, e o facto de o primeiro elemento que poder funcionar como complemento direto, permitem analisar esta frase como o resultado da coordenação de duas frases e posterior redução (elipse) de material repetido, nomeadamente, do complemento direto do primeiro verbo (catáfora). Finalmente, em ambos os casos, tratando-se da coordenação de dois verbos dos quais apenas o primeiro apresenta um sujeito explícito, uma vez que o segundo não se apresenta construído com nenhum elemento que desempenhe as funções de sujeito, considera-se que o sujeito deste verbo é correferente do sujeito do primeiro verbo e que, por isso, se reduziu (anáfora).

Contudo, repare-se que em cada uma destas frases, as relações sintáticas de modificador (no caso do advérbio *a horas certas*) e de complemento direto (*a entrevista*) só se estabelecem entre estes elementos e o verbo que ocorre imediatamente à sua esquerda.

É justamente com base nos agrupamentos e nas dependências existentes entre cada um dos eventos agrupados, que se procede à propagação e extração das dependências restantes em falta, nomeadamente, extraíndo uma relação temporal entre o evento *almoçava* e o advérbio *a horas certas* e uma relação direta entre os eventos *comentou* e *entrevista*.

(28a). *Ele almoçava e jantava a horas certas.*

```
EVENT_LEX(almoçava,outro)
EVENT_LEX(jantava,outro)
EVENT_GROUPRELATION(almoçava, jantava, e)
EVENT_TIMERELATION(almoçava, a horas certas)
EVENT_TIMERELATION(jantava, a horas certas)
```

(28b). *O Rui comentou e elogiou a entrevista.*

```
EVENT_LEX(comentou,outro)
EVENT_LEX(elogiou,outro)
EVENT_GROUPRELATION(comentou, elogiou, e)
EVENT_DIRECTRELATION(elogiou, entrevista)
EVENT_DIRECTRELATION(comentou, entrevista)
```

Note-se que, em certas situações, a vírgula assume a mesma função da conjunção coordenativa encontrada, permitindo que se agrupe mais que um evento. Em (29a), a relação temporal existente entre *dançou* e *durante a noite de ontem* deverá ser replicada pelos três eventos (29b).

(29a). *Ele comeu, bebeu e dançou bastante durante a noite de ontem.*

```
EVENT_LEX(comeu,outro)
EVENT_LEX(bebeu,outro)
EVENT_LEX(dançou,outro)
EVENT_GROUPRELATION(comeu, bebeu, e)
```

- (29b). **Ele comeu, bebeu e dançou bastante durante a noite de ontem.**
- ```

EVENT_GROUPRELATION(bebeu, dançou, e)
EVENT_LEX(comeu,outro)
EVENT_LEX(bebeu,outro)
EVENT_LEX(dançou,outro)
EVENT_GROUPRELATION(comeu, bebeu, e)
EVENT_GROUPRELATION(bebeu, dançou, e)
EVENT_TIMERELATION(dançou, durante a noite de ontem)
EVENT_TIMERELATION(bebeu, durante a noite de ontem)
EVENT_TIMERELATION(comeu, durante a noite de ontem)

```

Nalgumas situações, os eventos estão associados individualmente a outros eventos ou advérbios, independentemente de estarem ou não agrupados. Quando se identifica que existe dependências distintas entre dois eventos agrupados, considera-se que essas relações são relações específicas, obtando-se por não as propagar.

- (30a). **Vamos ao cinema hoje ou vemos o filme amanhã?**
- ```

EVENT_LEX(vamos,outro)
EVENT_LEX(vemos,outro)
EVENT_GROUPRELATION(vamos, vemos, ou)
EVENT_TIMERELATION(vamos, hoje)
EVENT_TIMERELATION(vemos, amanhã)

```
- (30b). **Ele avaliou a publicação dela e elogiou o trabalho.**
- ```

EVENT_LEX(avaliou,outro)
EVENT_LEX(publicação,outro)
EVENT_LEX(elogiou,outro)
EVENT_GROUPRELATION(avaliou, elogiou, e)
EVENT_DIRECTRELATION(avaliou, publicação)
EVENT_DIRECTRELATION(elogiou, trabalho)

```

Finalmente, considerem-se os casos mais complexos em que uma relação de se encontra associada a outra relação. Os exemplos seguintes ilustram esta situação:

- (31a). **O Pedro lia o jornal ou via televisão enquanto os miúdos brincavam na piscina.**
- ```

EVENT_LEX(lia,outro)
EVENT_LEX(via,outro)
EVENT_LEX(brincavam,outro)
EVENT_GROUPRELATION(lia, via, ou)
EVENT_INDIRECTRELATION(lia, brincavam, enquanto)
EVENT_INDIRECTRELATION(via, brincavam, enquanto)

```
- (31b). **O partido vai votar contra Orçamento e abster-se quanto ao projeto-lei não obstante o recuo do Governo nas suas propostas.**
- ```

EVENT_LEX(votar,outro)
EVENT_LEX(abster-se,outro)
EVENT_LEX(recuo,outro)
EVENT_LEX(propostas,outro)
EVENT_GROUPRELATION(votar, abster-se, e)
EVENT_INDIRECTRELATION(votar, recuo, não obstante)
EVENT_INDIRECTRELATION(abster-se, recuo, não obstante)
EVENT_INDIRECTRELATION(recuo, propostas)

```

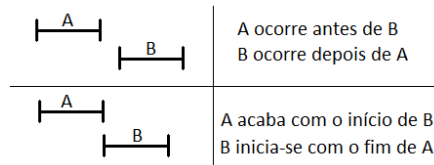
### 3 Ordenação

#### 3.1 Tipos de ordenação considerados

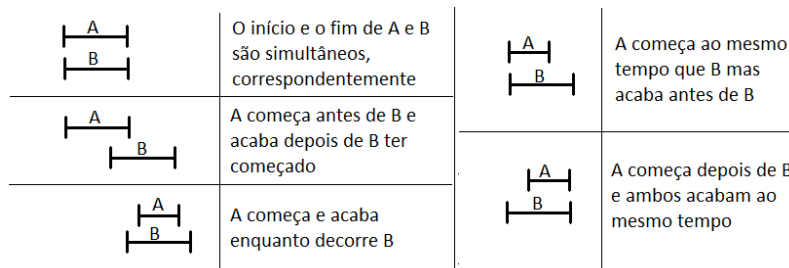
A ordem entre dois eventos é extraída com base nas dependências consideradas na secção anterior e na informação semântica associada aos conectores envolvidos (sobretudo conjunções, preposições), na informação temporal, modal e aspetual presente na conjugação (flexão) verbal e nos advérbios de tempo envolvidos. As relações de ordem consideradas são *anterior* (**ORDERBEFORE**) e *simultâneo*, (**ORDERSIMULT**) expressas do seguinte modo:

```
EVENT_ORDERBEFORE(<evento anterior>, <evento posterior>)
EVENT_ORDERSIMULT(<evento principal>, <evento subordinado/coordenado>)
```

Entre cada par de eventos, considera-se que um evento antecede outro se foi concluído antes de o segundo se iniciar, ou seja, nas condições expressas na figura 1. Se um evento ocorre em simultâneo, intercepta ou é incluído por outro, o mesmo é considerado simultâneo ao segundo, como representado na figura 2.



**Fig. 1.** Condições necessárias para o evento A ser anterior ao B.



**Fig. 2.** Condições necessárias para dois eventos serem simultâneos entre si.

Ainda que se pudesse considerar que a relação de ordem *simultâneo* fosse não-orientada, opta-se indicar o sentido em que foi identificada a dependência entre os eventos associados, como se ilustra nas frases seguintes:

- (32a). **A Maria come pipocas sempre que vê filmes.**  
 EVENT\_LEX(come,outro)  
 EVENT\_LEX(vê,outro)  
 EVENT\_INDIRECTRELATION(come, vê, sempre)  
 EVENT\_ORDERSIMULT(come, vê)
- (32b). **A Maria vê filmes sempre que come pipocas.**  
 EVENT\_LEX(come,outro)  
 EVENT\_LEX(vê,outro)  
 EVENT\_INDIRECTRELATION(vê, come, sempre)  
 EVENT\_ORDERSIMULT(vê, come)

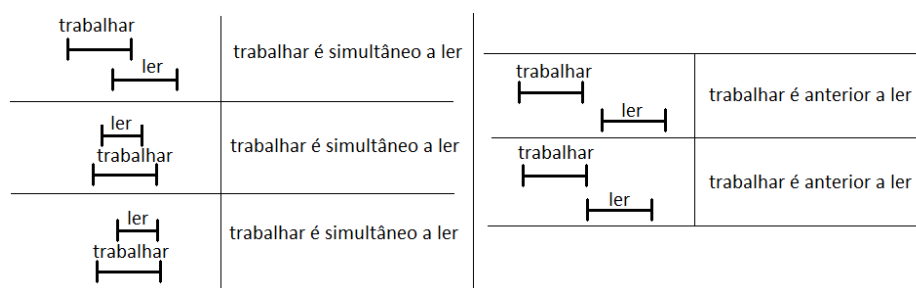
### 3.2 Critério de ordenação

Existem algumas situações ambíguas em que não é possível identificar perfeitamente as fronteiras de início e de fim para cada um dos eventos relacionados. Considere-se, por exemplo, as seguintes frases:

(33a). O Rui começou a ler esse livro depois de a Rita começar a trabalhar.

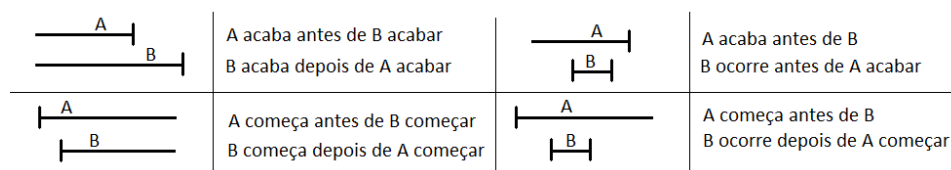
(33b). O Rui acabou de ler esse livro antes de a Rita acabar de estudar.

Em (33a), identifica-se que o evento **ler** se inicia depois de **trabalhar**, mas não se sabe qual dos eventos termina primeiro, existindo várias interpretações possíveis (ver figura 3). O problema é semelhante ao encontrado em (33b), onde é possível identificar o evento que termina em primeiro lugar, mas não a ordem com que se iniciaram.



**Fig. 3.** Interpretações possíveis para a frase *O Rui começou a ler esse livro depois de a Rita começar a trabalhar*.

Para a resolução destas ambiguidades, define-se que o início de um evento ocorre num passado distante quando se refere apenas a sua conclusão (aspeto terminativo); inversamente, quando apenas se identifica o início do evento (aspeto incoativo/incetivo), a sua conclusão é deixada indefinida (ver figura 4).



**Fig. 4.** Relações adicionais a considerar em relação às fronteiras de início e fim de eventos

Como resultado, os eventos das frases (33a) e (33b), ao se intercetarem, são simplesmente considerados como simultâneos e representando-se como se ilustra em (34a) e (34b).

(34a). O Rui começou a ler esse livro depois de a Rita começar a trabalhar.

```
EVENT_LEX(ler,outro)
EVENT_LEX(trabalhar,outro)
EVENT_INDIRECTRELATION(ler, trabalhar, depois de)
EVENT_ORDERSIMULT(ler, trabalhar)
```

(34b). O Rui acabou de ler esse livro antes de a Rita acabar de estudar.

```
EVENT_LEX(ler,outro)
EVENT_LEX(estudar,outro)
EVENT_INDIRECTRELATION(ler, estudar, antes de)
```



`EVENT_ORDERSIMULT(ler, estudar)`

Pressupõe-se que um evento sem informação aspetual associada tem uma interpretação *pontual*, isto é, inicia-se e finaliza-se no mesmo instante. Tal é particularmente relevante no caso de eventos em que o verbo (ou o verbo-suporte do nome predicativo) se encontram no presente do indicativo ou no infinitivo. Compare-se as frases (35a) e (35b).

(35a). O Rui chega a casa antes de a Rita sair do trabalho

(35b). O Rui chegou a casa antes de a Rita sair do trabalho.

Nestes exemplos, o tempo (não marcado) do infinitivo *sair* depende do tempo do verbo da oração principal. Contudo, o uso do presente do indicativo do verbo *chegar* em (35a) faz com que seja possível atribuir uma leitura *habitual-frequentativa* ao processo, equivalente a “costuma chegar a casa”, ao passo que o uso do pretérito perfeito em (35b) leva a uma leitura pontual. Neste caso, portanto, força-se a leitura pontual de (35a).

Recorde-se, ainda, que a ordenação de eventos depende de muitos fatores, podendo existir ambiguidades ou conflitos na ordem expressa entre as várias dependências detetadas. As frases que se seguem são exemplos deste tipo de conflitos:

(36a). O Rui esteve a falar acerca do jantar.

(36b). O Rui esteve a falar acerca do jantar de amanhã.

No exemplo (36a) é detectado uma relação entre os eventos *falar* e *jantar* a partir da preposição *acerca de*, a qual tem associado o traço semântico de *tópico* (*topic*), sugerindo a ocorrência do segundo evento antes da do primeiro. Por conseguinte, identificar-se-ia o evento *jantar* como *anterior* a *falar*. Por sua vez, o emprego do pretérito perfeito no auxiliar *esteve* de *falar* conduz à localização do evento num momento anterior ao momento da enunciação.

No entanto, em (36b), na presença do advérbio *amanhã* a modificar o evento *jantar*, este último é localizado no futuro, enquanto o tempo verbal de *falar* o remete para o passado. Deste modo, a ordem dos eventos é contrária à sugerida pela preposição *acerca de*: o evento *falar* ocorre *antes* do evento *jantar*, como se ilustra abaixo.

(37a). O Rui esteve a falar acerca do jantar

```
EVENT_LEX(falar,outro)
EVENT_LEX(jantar,outro)
EVENT_INDIRECTRELATION(falar, jantar, acerca de)
EVENT_ORDERBEFORE(jantar, falar)
```

(37b). O Rui esteve a falar acerca do jantar de amanhã.

```
EVENT_LEX(falar,outro)
EVENT_LEX(jantar,outro)
EVENT_INDIRECTRELATION(falar, jantar, acerca de)
EVENT_TIMERELATION(jantar, amanhã)
EVENT_ORDERBEFORE(falar, jantar)
```

### 3.3 Interpretação das relações

Quanto à ordenação particular entre os eventos que se relacionam, os conetores (conjunções subordinativas e preposições) *causais*, *comparativos* e *concessivos* exprimem, em geral, o sentido *anterior* (*ORDERBEFORE*), ou seja, considera-se que o evento subordinante ocorre *antes* do evento subordinado. Os conetores *condicionais*, *consecutivos*,  *finais* e de *tópico*, por sua vez, exprimem geralmente o sentido contrário, ou seja, que o evento subordinante ocorre *depois* do subordinado. As conjunções e preposições *proporcionais* (e.g. *à medida que*) exprimem a simultaneidade ou a interseção (*ORDERSIMULT*) dos eventos.

(38a). Ele gritou a ponto de enrouquecer.

- ```

EVENT_LEX( gritou, outro)
EVENT_LEX( enrouquecer, outro)
EVENT_INDIRECTRELATION( gritou, enrouquecer, a ponto de)
EVENT_ORDERBEFORE( gritou, enrouquecer)
(38b). Ele foi falando cada vez mais baixo à medida que ia enrouquecendo.
EVENT_LEX( falando, outro)
EVENT_LEX( enrouquecendo, outro)
EVENT_INDIRECTRELATION( falando, enrouquecendo, à medida que)
EVENT_ORDERSIMULT( falando, enrouquecendo)

```

Naturalmente, a natureza semântica dos eventos pode determinar a inversão desta ordenação básica dos eventos. Compare-se:

- ```

(39a). A viagem que nos contam é interessante.
EVENT_LEX( viagem, outro)
EVENT_LEX( contam, outro)
EVENT_INDIRECTRELATION( viagem, contam, que)
EVENT_ORDERBEFORE( viagem, contam)
(39b). A viagem que planeamos é interessante.
EVENT_LEX( viagem, outro)
EVENT_LEX( planeamos, outro)
EVENT_INDIRECTRELATION( viagem, planeamos, que)
EVENT_ORDERBEFORE( planeamos, viagem)

```

A ordem expressa pelas conjunções *temporais* varia conforme o significado das mesmas, como exemplificado nas frases (40a) e (40b).

- ```

(40a). Eu preparo a mesa antes de ele chegar.
EVENT_LEX( preparo, outro)
EVENT_LEX( chegar, outro)
EVENT_INDIRECTRELATION( preparo, chegar, antes de)
EVENT_ORDERBEFORE( preparo, chegar)
(40b). Eu preparo a mesa depois de ele chegar.
EVENT_LEX( preparo, outro)
EVENT_LEX( chegar, outro)
EVENT_INDIRECTRELATION( preparo, chegar, depois de)
EVENT_ORDERBEFORE( chegar, preparo)

```

Quando um dos eventos é expresso por um nome predicativo, na ausência de relações que permitam definir uma ordem explícita para esses eventos, considera-se, em geral, que o nome predicativo é *anterior* ao evento de que depende. No entanto, note-se que a natureza semântica dos eventos pode também determinar a inversão da ordem.

- ```

(41a). O Rui vai publicar a entrevista.
EVENT_LEX( publicar, outro)
EVENT_LEX( entrevista, outro)
EVENT_DIRECTRELATION( publicar, entrevista)
EVENT_ORDERBEFORE( entrevista, publicar)
(41b). O Rui preparou a entrevista de amanhã.
EVENT_LEX( preparou, outro)
EVENT_LEX( entrevista, outro)
EVENT_DIRECTRELATION( preparou, entrevista)
EVENT_TIMERELATION( entrevista, amanhã)
EVENT_ORDERBEFORE( preparou, entrevista)
(41c). O Rui está a preparar a entrevista.

```

```

EVENT_LEX(preparar,outro)
EVENT_LEX(entrevista,outro)
EVENT_DIRECTRELATION(preparar, entrevista)
EVENT_ORDERBEFORE(preparar, entrevista)

```

Define-se que as conjunções adversativas (e.g. **mas**) estabelecem, por omissão, uma ordem entre os eventos da dependência, em que o primeiro evento ocorre antes do segundo. A conjunção copulativa **e**, ao permitir uma enumeração de eventos, principalmente se o sujeito da ação for o mesmo, pode também estabelecer uma ordem de anterioridade entre o primeiro e o segundo evento.

(42a). **Eu avisei-a e tentei fazê-la compreender, mas ela não me ligou nenhuma.**

```

EVENT_LEX(avisei,outro)
EVENT_LEX(fazê,outro)
EVENT_LEX(compreender,outro)
EVENT_LEX(ligou,outro)
EVENT_GROUPRELATION(avisei, fazê, e)
EVENT_DIRECTRELATION(fazê, compreender)
EVENT_INDIRECTRELATION(avisei, ligou, mas)
EVENT_ORDERBEFORE(avisei, fazê)
EVENT_ORDERBEFORE(fazê, compreender)
EVENT_ORDERBEFORE(avisei, ligou)

```

## Bibliography

- [1] Baptista, J., Mamede, N., Hagège, C., Maurício, A.: Time expressions in portuguese guidelines for identification, classification and normalization. Tech. rep., L2F-Spoken Language Laboratory (September 2011)
- [2] Baptista, Jorge; Vera Cabarrão, N.J.M.: Classification directives for events and relations extraction between named entities in portuguese texts. Tech. rep., L2F-Spoken Language Laboratory (2012)
- [3] Loureiro, J.: Reconhecimento de Entidades Mencionadas (Obra, Valor, Relações de Parentesco e Tempo) e Normalização de Expressões Temporais. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa (2007), mSc Dissertation
- [4] Mamede, N.: String - a cadeia de processamento de língua natural do l2f (February 2011), slides from the Seminar at NILC/ICMC/USP
- [5] Mamede, N., Baptista, J., Cláudio, D.: String - an hybrid statistical and rule-based natural language processing chain for portuguese. In: Springer (ed.) PROPOR 2012 (2012)
- [6] Mamede, N., Baptista, J., Hagège, C., et al.: Nomenclature of chunks and dependencies in portuguese xip grammar (v.4). Tech. rep., L2F/INESC-ID Lisboa, Portugal (June 2013)
- [7] Romão, L.: Reconhecimento de Entidades Mencionadas em Língua Portuguesa: Locais, Pessoas, Organizações e Acontecimentos. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa (2007), mSc Dissertation