

Proposta de anotação e normalização de expressões temporais da categoria TEMPO para o HAREM II

18 de Fevereiro de 2008
(corrigida em 13 de Abril de 2008)

Caroline Hagège (Xerox)

Jorge Baptista (Universidade do Algarve, L2F – INESC-ID Lisboa)

Nuno Mamede (Instituto Superior Técnico, L2F – INESC-ID Lisboa)

SUMÁRIO:

1	Preâmbulo	2
2	Motivação da proposta	2
3	Proposta	3
3.1	Categoria TEMPO	3
3.1.1	Definição da entidade de tipo TEMPO	3
3.1.1.1	Critérios para a identificação	3
3.1.1.2	Critérios para a delimitação das EM da categoria TEMPO	6
3.2	TIPO = "TEMPO_CALEND"	9
3.2.1	SUBTIPO = "DATA"	9
3.2.2	Expressões de datas relativas: dois tipos de referências considerados	10
3.2.3	Atributo TEMPO_REF	11
3.2.4	Atributos SENTIDO e VAL_DELTA	13
3.2.5	SUBTIPO = "HORA"	15
3.2.6	SUBTIPO = "INTERVALO"	15
3.3	TIPO = "DURACAO"	15
3.4	TIPO = "FREQUENCIA"	16
3.5	TIPO = "GENERICICO"	16
3.6	Atributo VAL_NORM	17
3.6.1	Atributo VAL_NORM para expressões de subtipo DATA absoluta	17
3.6.2	Atributo VAL_NORM para expressões de tipo HORA	18
3.6.3	Atributo VAL_NORM para expressões de tipo DURACAO	19
4	Resumo das principais modificações	19
5	Alguns exemplos de anotação	20
6	Referências bibliográficas	21
7	Agradecimentos	22

1 Preâmbulo

No âmbito do Reconhecimento de Entidades Mencionadas (REM), uma das tarefas de reconhecida importância consiste no reconhecimento de expressões temporais (entidades mencionadas da categoria *TEMPO*). Esta proposta tem por finalidade acrescentar uma nova faceta a esta tarefa já na próxima campanha do HAREM: a normalização das entidades mencionadas de tipo *TEMPO*.

Para levar a cabo esta tarefa, é necessário, por um lado, completar e enriquecer a actual definição da categoria *TEMPO*, tal como se encontra em [Cardoso & Santos 06]. Por outro lado, a noção de entidade mencionada da categoria *TEMPO* tem de ser alargada à noção mais geral de expressão temporal.

2 Motivação da proposta

As motivações para esta proposta são as seguintes:

- 1) levar em conta os avanços e as direcções gerais de trabalhos recentes no âmbito do processamento de expressões temporais em textos (ver, por exemplo, TimeML em [Saurí et al. 06] e a campanha TempEval em [Verhagen et al. 07]).

Mais precisamente, considera-se que a tarefa de REM de expressões temporais pode e deve ser vista como um primeiro passo para um mais rico processamento do sistema de referências temporais em textos. Neste sentido, tem-se a convicção que a comunidade interessada em REM em língua portuguesa poderia beneficiar bastante se passasse a considerar desde já os trabalhos de PLN efectuados no domínio do processamento das expressões que denotam e estruturam as referências temporais em textos; por outro lado, parece necessário assegurar que a segmentação/delimitação e a classificação de expressões temporais preconizadas pelo HAREM sejam compatíveis com linhas de investigação já existentes neste domínio e internacionalmente estabelecidas.

- 2) enriquecer a actual categorização proposta em [Cardoso & Santos 06].

Se se aceitar a ideia de prolongar a identificação de EM temporais de modo a chegar-se à sua normalização, será necessário alargar o conceito actual de entidade temporal. Por exemplo, expressões temporais de tipo *FREQUENCIA* (i.e. repetição de eventos no tempo), cuja definição será explicitada mais adiante, não parecem estar contempladas nas directivas de anotação do primeiro HAREM.

3 Proposta

A proposta que se segue é largamente inspirada nos trabalhos recentes do TimeML (cf. www.timeml.org/site).

3.1 Categoria TEMPO

Na categoria *TEMPO*, considera-se uma grande parte das expressões que, semanticamente, denotam: (i) um **momento** no calendário (que pode ser concebido como um ponto ou como um intervalo); (ii) uma expressão de quantificação temporal que exprime uma **duração**; ou (iii) uma **repetição** de eventos no tempo; considera-se ainda (iv) o emprego **genérico** de algumas dessas expressões, geralmente associadas à noção de tempo. Nesta secção, apresenta-se, em primeiro lugar, uma definição geral das entidades mencionadas da categoria *TEMPO*. Definem-se, de seguida, os critérios que permitem determinar se uma expressão linguística pertence ou não à categoria *TEMPO*. Apresentam-se, ainda, os critérios que permitem delimitar uma expressão temporal complexa.

3.1.1 Definição da entidade de tipo TEMPO

3.1.1.1 Critérios para a identificação

Uma expressão temporal é qualquer expressão que responde ao critério 1 e a pelo menos um dos subcritérios de 2 **ou**, então, poderá ser uma expressão temporal genérica, que responde ao critério 3. As expressões temporais poderão **não conter algarismos ou palavras em maiúsculas**. Consideram-se necessários e suficientes para uma definição de expressão temporal os seguintes critérios, ordenados como acima se referiu:

critério 1 – uma expressão temporal em contexto pode responder adequadamente a uma das interrogativas “(<prep> *quando?*”, “(<prep> *quanto tempo?*”, “(<haver> *quanto tempo?*” ou “*com que frequência?*”.

critério 2 – uma expressão temporal contém pelo menos uma unidade lexical que corresponda a um dos seguintes tipos:

2-1 - uma data numérica (por exemplo, *29-10-2008*);

2-2 - uma unidade de medida temporal (*dia, mês, trimestre, ano, século, etc.*) ou um advérbio terminado em “-mente” derivado destas expressões (*diariamente, semanalmente, mensalmente, etc.*);

2-3 - um nome correspondente à designação de uma destas unidades de medida de tempo. Isto é: nome de meses (*Setembro, Dezembro, etc.*), nome de dia (*segunda-feira, domingo, etc.*);

- 2-4 - um nome de festividade, religiosa ou não (*Natal, Páscoa, Quaresma, Entrudo*); nomes de estações do ano (*Primavera, Inverno*); nomes de festividades, que podem incluir o nome *dia* (*dia de Santo António, dia de Nossa Senhora da Conceição, dia de São Valentim, dia dos namorados, no São Martinho*, etc.);
- 2-5 - alguns advérbios de tempo (simples, não-derivados e semanticamente não ambíguos), tais como: *hoje, ontem, amanhã, outrora*; algumas locuções adverbiais de tempo (ou advérbios compostos), semanticamente não ambíguas, como, por exemplo: *antes de ontem, depois de amanhã*; excluem-se da actual campanha de avaliação os seguintes advérbios simples, sintáctica ou semanticamente ambíguos, apesar de poderem representar expressões temporais: *agora, ainda, já, sempre*.
- 2-6 - um sintagma preposicional cujo núcleo seja uma das palavras *altura, tempo, momento, período, era*, etc., quando estas palavras forem determinadas por um demonstrativo (por exemplo: *nesse tempo*), ou especificados por uma relativa (por exemplo: *na altura em que ela adoeceu*), um possessivo (por exemplo: *durante a nossa era*) ou modificado por outro sintagma preposicional introduzido por *de* (por exemplo: *durante a era dos dinossauros*) ou então por um adjectivo capitalizado (por exemplo: *durante o período Barroco, Cretáceo*, etc.);
- 2-7 - Os complementos determinativos com a forma *de Num Ntmp* de nomes predicativos, que não respondem adequadamente ao critério (1) mas que são indubitavelmente EM a anotar (e.g. *uma viagem de 5 dias*); a preposição *de* deve ser incluída na EM;
- 2-8 - expressões de frequência, como as seguintes: *de vez em quando, às vezes, de quando em quando, frequentemente*, etc.
- 2-9 - expressões da forma *Prep + <unidade de medida temporal> + que + verbo vir ou verbo passar* (por exemplo, *no ano que passou, para o mês que vem*)
- 2-10 - expressões com os verbos *fazer* ou *haver* e <unidade de medida temporal> (e.g. *há três anos, faz duas semanas*).

Notas:

- (a) Excluem-se, no critério 1, as expressões de tipo genérico como o emprego de *o inverno* em frases como *Adoro o inverno*, que serão retomadas de forma autónoma, no critério 3, abaixo.
- (b) Excluem-se também, com o critério 1, as locuções que, embora contendo expressões do conjunto identificado no critério 2, não respondem adequadamente às interrogativas de tempo. Trata-se de locuções como *de dia para dia*, que encontramos em frases como *A situação agrava-se de dia para dia*, que funciona como um circunstancial de modo; repare-se na inaceitabilidade do par pergunta-resposta: P:(*quando, com que frequência, em quanto tempo*) é que a situação se agrava? R: *de dia para dia*.
- (c) Repare-se também que, para qualquer dos pontos 2-2 a 2-9 do critério 2, se pode fazer uma definição *em extensão* dos elementos em questão. Assegura-se, assim, o problema de intersubjectividade das anotações.

critério 3 – uma expressão temporal que contém uma unidade lexical do tipo das que foram definidas no critério 2 mas para a qual o critério 1 não se aplica.

Trata-se de expressões temporais genéricas como *o mês de Julho* em exemplos como *Adoro o mês de Julho* onde *o mês de Julho* não responde à pergunta *quando?* embora contenha elementos lexicais como os que foram definidos no critério 2.

Os critérios apresentados permitem contemplar, nos exemplos que se seguem, as expressões representadas em negrito:

- ***Em 2008*** *haverá mais confiança no futuro*, que, em contexto, responde à pergunta *quando?* (critério 1) e responde ao critério 2-1;
- ***Chegou no dia 5 de Junho de 2006*** que, em contexto, responde à pergunta *quando?* (critério 1) e responde aos critérios 2-1 e 2-3;
- ***Viveu em Lisboa entre 2000 e 2003*** que responde à pergunta *quando?* e responde ao critério 2-1;
- ***De um dia para o outro*** *o restaurante mudou completamente o seu menu* que, em contexto, responde à pergunta *quando?* (critério 1) e contém a palavra *dia* (critério 2-2);
- ***Trabalhei durante dois meses*** que responde à pergunta (*durante*) *quanto tempo?* (critério 1) e que contém a palavra *mês* (critério 2-2);
- ***O padeiro vem duas vezes por semana*** que, em contexto, responde à pergunta *com que frequência?* (critério 1) e contém a palavra *semana* (critério 2-2);
- ***Vou visitar os meus pais semanalmente***, que, em contexto, responde à pergunta *com que frequência?* (critério 1) e contém um advérbio terminado em *-mente* derivado de um nome de tempo semana (critério 2-2);
- ***Chegou no dia de Natal*** que, em contexto, responde à pergunta *quando?* (critério 1) e responde aos critérios 2-2 e 2-4;
- ***De hoje em diante*** *vou trabalhar* que, em contexto, responde à pergunta (*a partir de*) *quando?* (critério 1), e contém a palavra *hoje* (critério 2-5);
- ***Esteve em Lisboa há dois anos*** que, em contexto, responde à pergunta *há quanto tempo?* (critério 1) e que responde ao critério 2-2;
- ***As vindimas fazem-se nesta altura do ano***, que em contexto, responde à pergunta *quando?* (critério 1) e que responde ao critério 2-6;
- ***Aconteceu durante a era dos dinossauros*** que, em contexto, responde à pergunta *quando?* (critério 1) e que responde ao critério 2-6;
- ***Vou à pesca de vez em quando***, que em contexto, responde à pergunta *com que frequência?* e que responde ao critério 2-8;
- ***Ficou doente dois anos mais tarde***, que em contexto, responde à pergunta *quando?* e que responde ao critério 2-2.

Consideram-se também expressões genéricas como:

- *A Primavera é a mais bela estação do ano*, em virtude do critério 3 (expressões temporais genéricas).

Pontos importantes:

Embora obedçam sempre ao critério 1 e possam às vezes obedecer a algum dos subcritérios de 2, excluem-se as orações subordinadas de tempo (por exemplo, *quando o meu pai chegar* que responde ao critério 1).

Do mesmo modo, as expressões temporais introduzidas pelas locuções prepositivas (ou preposições compostas) *antes de* e *depois de* só serão marcadas como entidades mencionadas de tempo quando introduzirem um sintagma/grupo nominal (*antes de Domingo*, *depois de 2008*) ou ou sintagma adverbial (*antes de amanhã*, *depois de ontem*), em que ocorre um dos elementos lexicais especificados nos subcritérios de 2. Pelo contrário, quando estas locuções funcionam como conjunções e introduzem orações subordinadas temporais (*antes de o meu pai chegar*, *depois de ter acabado o trabalho*), todo o complemento adverbial será ignorado.

Também se excluem expressões fixas/idiomáticas com valor temporal, tais como *Quando as galinhas tiverem dentes* ou *daqui para frente*, apesar de responderem adequadamente ao critério 1. *Quando as galinhas tiverem dentes*, é, superficialmente, uma oração subordinada, conquanto seja uma expressão fixa, e, *daqui para frente* não obedece a nenhum dos subcritérios de 2.

Note-se que do ponto de vista linguístico, não há qualquer razão para excluir estas expressões (compostas, idiomáticas, subordinadas) do processamento de expressões temporais. A exclusão deste tipo de expressões, por ora, prende-se apenas com os limites que se pretendem delinear para a tarefa de reconhecimento de entidades mencionadas.

3.1.1.2 Critérios para a delimitação das EM da categoria TEMPO

A fim de se poder anotar de maneira unívoca as entidades da categoria *TEMPO*, convém ainda definir rigorosamente critérios sintáctica e semanticamente motivados que deverão ser seguidos a fim de se delimitar com precisão as fronteiras das entidades a anotar.

Nesta proposta, considera-se que a totalidade da expressão temporal deverá ser delimitada entre as balizas `<EM ID=... CATEG="TEMPO"> e `, isto é, **incluindo a preposição** que a introduzir, no caso da expressão temporal ser um sintagma preposicional (e.g. *no ano passado*), **ou o determinante** no caso de ser um sintagma nominal (e.g. *dois dias depois*).

No caso de expressões complexas como *dois dias depois do Natal*, a questão que se coloca é a de se saber se esta expressão deverá ser considerada como uma só EM ou,

então, segmentada em duas subexpressões *dois dias + depois do Natal* (obedecendo tanto a expressão mais longa como ambas as subexpressões aos critérios definitórios mencionados acima).

Os critérios adoptados para a segmentação são os definidos em [Hagège & Tannier 07] e que aqui foram reproduzidos:

Uma expressão temporal complexa **deverá ser dividida** em unidades menores se se verificarem **simultaneamente** os critérios seguintes:

1 - cada expressão componente é sintacticamente válida quando combinada independentemente com o evento que modifica.

2 - cada expressão componente, combinada com o evento que modifica, está logicamente implicada na expressão complexa. Ou seja, cada combinação “evento + expressão_temporal_mínima” deve ser logicamente implicada pela combinação “evento + expressão_temporal_complexa”. Em outras palavras, o valor de verdade de todas as combinações “evento+expressão_temporal_mínima” deve poder ser deduzido do valor de verdade da combinação “evento + expressão_temporal_complexa” (ver exemplos abaixo para ilustração).

Exemplos:

Na frase:

Visitei-o dois dias nesta semana,

a expressão *dois dias nesta semana* deverá ser considerada como constituída por duas entidades, pois cada subexpressão, *dois dias* (*DURACAO*) e *nesta semana* (*DATA*), pode combinar-se separadamente com o evento (*visitei-o*) e, se considerarmos que o valor de verdade da frase é verdadeiro, ambos os valores de verdade de *visitei-o dois dias* e de *visitei-o nesta semana* são verdadeiros.

Na frase:

Visitei-o dois dias depois,

a expressão *dois dias depois* deverá ser considerada como uma só entidade (*DATA*). Com efeito, apesar de cada uma das subexpressões (*dois dias* (*DURACAO*) e *depois* (*DATA*)) poder combinar-se individualmente com o evento (*visitei-o*), verifica-se uma diferença de significado relativamente à interpretação da expressão complexa, mais concretamente, surge um novo adverbial, de *DURACAO*. O critério 2 não se verifica, pois: se se supuser que a asserção *visitei-o dois dias depois* é verdadeira, nada garante que o valor de verdade *visitei-o dois dias* seja verdadeiro.

Na frase:

Isso aconteceu dois dias depois do Natal,

a expressão *dois dias depois do Natal* deve ser considerada como uma só entidade (*DATA*). Com efeito, ao considerar as duas sub-expressões *Isso aconteceu dois dias* e *Isso aconteceu depois do Natal*, pode-se verificar que nem só a primeira subexpressão é duvidosa do ponto de vista da aceitabilidade sintáctica, mas também que, mesmo que fosse aceitável, o valor de verdade desta sub-expressão (*FREQUENCIA*) não é logicamente implicado pelo valor de verdade da frase inicial (*DATA*).

NB: Casos ambíguos como:

Vimo-nos <EM ID=... CATEG="TEMPO">*dois dias depois do Natal*.

Esta frase é ambígua e pode ser interpretada como:

- 1) *Vimo-nos do dia 27 de Dezembro*;
- 2) *Vimo-nos durante dois dias, a seguir ao 25 de Dezembro*.

Neste caso, embora a presença do segundo membro tenha tendência em ‘forçar’ a leitura complexa da expressão temporal (*DATA*), a ambiguidade será expressa na anotação (ver exemplos finais do ponto 5).

Outros exemplos de delimitação de expressões temporais complexas

Emigrou há 23 anos depois do 25 de Abril.

Neste exemplo devem ser consideradas duas expressões temporais separadas (*há 23 anos* e *depois do 25 de Abril*). Com efeito, além das duas sub-expressões *emigrou há 23 anos* e *emigrou depois do 25 de Abril* serem sintacticamente válidas, se se considerar que o valor de verdade da frase é verdadeiro, também o valor de verdade das duas sub-expressões é verdadeiro.

Pelas mesmas razões, expressões como:

durante um fim de semana em Abril,
depois das férias do Natal em 2003

devem ser consideradas como pares de expressões temporais separadas:

durante um fim de semana (DATA) em Abril (DATA),
depois das férias do Natal (DATA) em 2003 (DATA)

No entanto, uma expressão como *dois anos mais tarde* em:

Ficou doente dois anos mais tarde

deve ser considerada como uma única expressão. Com efeito embora as sub-expressões *Ficou doente dois anos* e *Ficou doente mais tarde* sejam sintacticamente válidas, o valor de verdade de *Ficou doente dois anos* (*DURACAO*) não pode ser deduzido do valor de verdade da totalidade da expressão (*DATA*). (Nada se pode dizer sobre a duração da doença se se considerar que *Ficou doente dois anos mais tarde* for verdadeiro).

Qualquer expressão temporal deverá ser anotado por `<EM ID=... CATEG="TEMPO">` e possuir o atributo obrigatório *TIPO*.

O atributo *TIPO* é o único atributo obrigatório do elemento EM de categoria *TEMPO*.

Os diferentes valores do atributo *TIPO* são:

TEMPO_CALEND (tempo calendário),
DURACAO (duração),
FREQUENCIA (frequência)
GENERICO (genérico).

Cada um deste tipo é detalhado nas secções seguintes.

3.2 TIPO = "*TEMPO_CALEND*"

As entidades de tipo *TEMPO_CALEND* são expressões que permitem inserir o predicado que elas modificam numa linha temporal (como um ponto ou um intervalo).

Correspondem aos seguintes subtipos:

- **datas** sejam elas **absolutas** (fórmulas contendo três campos *ANO-MES-DIA*, na qual até dois campos no máximo podem ser omitidos) ou **referenciais** (expressões temporais cuja resolução implica conhecer ou o momento da enunciação, ou outra data de um evento que funciona como referência).
- **intervalos** (expressões denotando uma duração no tempo e que têm explicitamente dois limites)
- **horas** (expressões temporais com valor de *DATA* mas com granularidade inferior à unidade *dia*).

3.2.1 SUBTIPO = "*DATA*"

As expressões deste subtipo podem representar *datas absolutas* ou *datas relativas* (que são referências). No primeiro caso, a expressão contém a informação necessária para localizar essa data num calendário (e.g. na expressão *em 23 de Outubro de 2007*, a informação está totalmente especificada em relação aos 3 campos; pelo contrário, nas expressões *em 23 de Outubro* e *em 2007*, a informação está parcialmente especificada em relação aos 3 campos).

Também são consideradas como abrangidas pelo subtipo *DATA* as expressões que exprimem *datas relativas*, isto é, para as quais é necessário determinar um ponto de

referência para poder localizá-las na linha temporal (e.g. *dois dias mais tarde, na quinta-feira passada, ontem, na próxima terça-feira*, etc.).

Apresentam-se de seguida alguns exemplos de expressões temporais do tipo *TEMPO_CALEND* e subtipo *DATA*:

- *Vou viajar* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">**no dia 19 de Outubro de 2007**. [Data absoluta completa \(campos dia, mês e ano preenchidos\)](#);
- *Trabalhei em Londres* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">**em 1998**. [Data absoluta incompleta \(campos dia e mês não preenchidos\)](#);
- *Vou a Lisboa* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">**no próximo dia 22**. [Data relativa](#);
- *Vai haver uma festa* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">**na próxima terça-feira**. [Data relativa](#);
- *Fui a Lisboa* <EM ID=... CATEG="TEMPO" TIPO="DATA">**na semana passada**. [Data relativa](#);
- *A Joana nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">**no Inverno** (exemplo do guia de anotação do Mini-HAREM). [Data relativa](#);
- *Vou a Londres* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">**no próximo Inverno** (exemplo do guia de anotação do Mini-HAREM). [Data relativa](#).

NOTA IMPORTANTE: No último exemplo, existe uma diferença relativamente às directivas do guia de anotação [Cardoso & Santos 06]. Com efeito, no Mini-HAREM, a expressão *Inverno* era considerada como uma duração (*PERIODO* segundo a terminologia utilizada no guia). É importante sublinhar que se considera que entidades de subtipo *DATA* não implicam um valor aspectual pontual mas que podem ser representados por intervalos com uma granularidade variável.

Assim, a expressão *no próximo inverno*, referida acima, é de subtipo *DATA*, pois mesmo que seja uma expressão que subentende uma certa duração, a ida a Londres mencionada na frase pode ser ancorada num calendário mediante o conhecimento da data da enunciação, que permite resolver se se trata do Inverno do ano de 2007/2008 ou de outro inverno qualquer (se a data de enunciação for o presente, tratar-se-á do tempo calendário entre 21 de Dezembro de 2007 e 20 de Março de 2008).

3.2.2 Expressões de datas relativas: dois tipos de referências considerados

Fez-se no ponto anterior a distinção entre expressões temporais correspondente a uma data absoluta (isto é, que permite, sem recurso a nenhum contexto, localizar na linha do

tempo o evento ao qual a data está associada) e expressões temporais relativas que são referenciais.

Explicitam-se agora os dois tipos de expressões temporais relativas consideradas: expressões temporais relativas que fazem referência ao tempo da enunciação e expressões temporais relativas cuja referência está introduzida no discurso.

Um exemplo típico desta distinção pode ser dado através do exemplo seguinte:

Chegou ontem

Chegou no dia anterior.

Nestes dois exemplos está-se na presença de expressões temporais que podem permitir localizar no calendário o evento associado (*TIPO="TEMPO_CALEND"*). Na medida em que não se trata de um intervalo de tempo com limites explícitos, nem da expressão de uma hora, pode-se associar a estas expressões o valor *SUBTIPO="DATA"*. Mas não se trata aqui de expressões correspondente a uma data absoluta, mas sim a datas relativas. Como datas relativas, estas expressões são referenciais. Com efeito, será necessário se se quiser localizar o evento *Chegou* na linha do tempo, ter em conta uma referência.

No primeiro exemplo, esta referência é o momento da enunciação.

Com efeito, se a asserção *Chegou ontem* for produzida no dia 4/12/2007, pode-se inferir que o evento *Chegou* ocorreu no dia 3/12/2007. O tempo no qual ocorre o evento neste exemplo é função do tempo do momento da enunciação (tempo_enunciação – 1 dia).

Fala-se neste caso de expressão temporal referencial relativa ao momento da enunciação.

No segundo exemplo, embora também se trate de uma data referencial, a referência não é o momento da enunciação. A localização temporal de *chegou* é independente do momento em que for produzida a asserção. Neste caso, a referência é outra data/evento que aparece no contexto textual ou discursivo.

Por exemplo:

O barco só devia chegar ao porto no dia 25 de Novembro, no entanto chegou no dia anterior

Vê-se, contextualizando o exemplo, que a referência da expressão *no dia anterior* é o evento da chegada do barco ao porto que ocorreu no dia 25/11. Conhecendo esta referência pode-se então deduzir que o evento *chegou* ocorreu no dia 24/11. Assim, neste caso está-se em presença de uma expressão referencial textual.

Esta distinção entre data absoluta, data referencial relativa ao momento de enunciação e data relativa com referência textual é formalizada, na próxima sub-secção, através do atributo *TEMPO_REF*.

3.2.3 Atributo *TEMPO_REF*

O atributo *TEMPO_REF*, diz apenas respeito às expressões temporais de *TIPO="TEMPO_CALEND" SUBTIPO="DATA"* (ver sub-secção 3.2.1).

No caso de datas absolutas, o valor do atributo *TEMPO_REF* é *ABSOLUTO*.

No caso de datas referenciais, conforme o tipo da referência (ver sub-seção 3.2.2) o valor do atributo *TEMPO_REF* é respectivamente *ENUNCIACAO* ou *TEXTUAL*.

Por exemplo, na frase *Partiu no dia 3 de Novembro de 2007*, a expressão temporal de tipo *DATA* *no dia 3 de Novembro de 2007* permite determinar sem ambiguidade que o evento da partida ocorreu no intervalo entre 3/11/2007 00:00 e 3/11/2007 24:00.

No caso das expressões com valor temporal relativo (e.g. *dois dias depois, na próxima sexta-feira*), a expressão temporal por si só, não é suficiente para poder situar o evento num calendário. No primeiro exemplo trata-se de um caso de referência textual (*TEMPO_REF="TEXTUAL"*), no segundo caso de uma referência ao momento da enunciação (*TEMPO_REF="ENUNCIACAO"*).

Os seguintes exemplos ilustram o uso do atributo *TEMPO_REF* e dos seus possíveis valores:

- *Nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ABSOLUTO">**a 3 de Janeiro de 1986**.
- *Nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL">**dois dias depois do Natal**.

No exemplo acima, note-se primeiro que, conforme os critérios de segmentação que definimos (ver ponto 3.1.1.2), esta expressão complexa tem de ser considerada como um todo. Para poder situar no calendário o evento (o nascimento) que a expressão localiza temporalmente, é necessário conhecer um tempo de referência correspondente ao *Natal*. Sendo esta referência textual, o valor de *TEMPO_REF* é *TEXTUAL*

- *Nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ENUNCIACAO">**na sexta-feira passada**.

Para poder calendarizar o evento *Nasceu* do exemplo acima, é necessário conhecer a data na qual foi enunciada a frase. A partir desta data de enunciação é que se poderá calcular o dia que corresponde à sexta-feira anterior à esta data de enunciação. O valor de *TEMPO_REF* neste caso é *ENUNCIACAO*.

- *Nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL">**dois dias depois**.

No exemplo acima, para poder localizar o evento no tempo, é necessário conhecer um tempo de referência que terá sido introduzido previamente no discurso. Neste caso, o valor de *TEMPO_REF* é *TEXTUAL* (mesmo que a referência não apareça explicitamente na frase).

NOTA IMPORTANTE:

Em caso de expressões como *dois dias depois de o meu pai chegar*, em conformidade com os critérios explicitados nos pontos anteriores, só a subexpressão *dois dias* será anotada entre as balizas *TEMPO*. Esta expressão é de tipo *TEMPO_CALEND* *SUBTIPO="DATA"* e o valor de *TEMPO_REF* é *TEXTUAL*, sendo o tempo de referência neste caso a data a que se refere a oração subordinada temporal (i.e. *depois de o meu pai chegar*).

Outros exemplos:

- *O Pedro chegou* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ENUNCIACAO">**ontem**.
- *O Pedro partiu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL">**na semana seguinte**.
- *O Pedro chegou a Paris* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ENUNCIACAO">**no domingo**. <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL">**Dois dias depois** foi para Grenoble.

3.2.4 Atributos SENTIDO e VAL_DELTA

No caso de expressões temporais referenciais (*TEMPO_REF* tem o valor *TEXTUAL* ou *ENUNCIACAO*), e dois novos atributos podem estar presentes na anotação: o atributo *SENTIDO* e o atributo *VAL_DELTA*.

O atributo *SENTIDO* permite dar uma informação complementar que tem por finalidade a normalização de expressões temporais referenciais. Mais precisamente, vai indicar se o seu valor temporal se situa cronologicamente *antes* ou *depois* do tempo de referência. Os possíveis valores do atributo *SENTIDO* são, pois, *ANTERIOR*, *POSTERIOR*, *SIMULT*, *ANTERIOR_OU_SIMULT*, *POSTERIOR_OU_SIMULT*. Estes valores correspondem respectivamente aos casos em que o valor temporal denotado pela expressão de data relativa referencial se situa antes, depois, ao mesmo tempo, antes ou ao mesmo tempo, ao mesmo tempo ao depois do valor temporal da referência.

Quanto ao atributo *VAL_DELTA*, ele tem por valor uma expressão que indica a distância temporal entre o tempo do evento denotado pela expressão temporal e o momento de referência (seja este o tempo da enunciação ou outro) quando esta distância temporal aparece explicitamente no texto. No caso desta distância temporal não ser explícita, o valor de *VAL_DELTA* é omitido.

No caso da distância temporal ser explícita, o valor de *VAL_DELTA* corresponde ao valor temporal que se deve incrementar ou subtrair a partir do tempo de referência para obter o valor temporal do evento associado à expressão temporal a anotar.

Os valores possíveis de *VAL_DELTA* são representados da maneira seguinte:

A<digitos>M<digitos>S<digitos>D<digitos>H<digitos>M<digitos>S<digitos>

Onde:

- as letras *A*, *M*, *S*, *D*, *H*, *M*, *S* são constantes que devem aparecer nesta ordem e que correspondem respectivamente ao valores de Anos, Meses, Semanas, Dias, Horas, Minutos e Segundos.
- os *<digitos>* à direita das letras constantes correspondem ao número de Anos, Meses, Semanas, Dias, Horas, Minutos e Segundos que se devem adicionar ou diminuir à data de referência para obter o valor temporal da expressão anotada.

Por exemplo:

- *Apareceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL" SENTIDO="PÔSTERIOR" VAL_DELTA="A0M0S2D0H0M0S0">**duas semanas** *depois da festa*.

Para proceder à normalização da expressão *duas semanas* é necessário conhecer um tempo de referência (*TEMPO_REF="TEXTUAL"*) que corresponde aqui à data da *festa*. O valor do atributo *SENTIDO* é *POSTERIOR* na medida em que a data do evento (*apareceu*) teve lugar após esta data de referência (festa); o valor de *VAL_DELTA* indica que esta distância temporal entre a data do evento e a data de referência corresponde a duas semanas: os valores de todos os campos são 0 excepto para o campo S (semana), em que se indica 2.

Outros exemplos:

- *Veio* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ENUNCIACAO" SENTIDO="ANTERIOR" VAL_DELTA="A0M0S0D1H0M0S0">**ontem**;
- *O Pedro partiu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL" SENTIDO="PÔSTERIOR" VAL_DELTA="A0M0S1D0H0M0S0">**na semana seguinte** ;
- *Nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL" SENTIDO="PÔSTERIOR" VAL_DELTA="A0M0S0D2H0M0S0">**dois dias depois do Natal**;

Nota-se que em falta de informação explícita no texto, *VAL_DELTA* poderá ser omitido ter um valor indefinido que será representado pela cadeia vazia "".

Por exemplo:

- *O Pedro chegou* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL" SENTIDO="PÔSTERIOR" VAL_DELTA="">**depois**.

No exemplo acima, trata-se de uma expressão de tipo *DATA* referencial. O ponto de referência é um evento ou uma data que não está presente na própria frase mas que foi introduzido anteriormente no discurso (*TEMPO_REF="TEXTUAL"*). O evento da chegada do Pedro ocorre a seguir a este momento de referência (*SENTIDO="POSTERIOR"*). No entanto, não se tem explicitamente a distância temporal

entre o evento da chegada do Pedro e o momento de referência. Por esta razão, o valor de *VAL_DELTA* será reduzido à cadeia vazia (*VAL_DELTA=""*). Poderá também ser omitido.

3.2.5 SUBTIPO = "HORA"

Trata-se de expressão temporais com valor de *DATA* mas com granularidade inferior à unidade *dia*. Do ponto de vista da sua definição, mantém-se a proposta do primeiro HAREM.

Exemplo:

- *O Pedro está disponível* <EM ID=--- CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="+-----T15--E-LMA">**às 15:00** .

3.2.6 SUBTIPO = "INTERVALO"

Corresponde a uma expressão complexa, isto é, composta por duas expressões temporais elementares/simples, mas que, semanticamente, forma um única entidade mencionada e que tem **explicitamente** dois *limites temporais* (*limite inicial* e *limite final*).

Exemplos:

- *Trabalhei em Londres* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="INTERVALO">**entre 2000 e 2003** .
- *Trabalhei em Londres* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="INTERVALO">**de Outubro a Dezembro de 2007** .
- *Vai demorar* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="INTERVALO">**de 3 a 6 meses** .

NOTA 1: Embora certas expressões do *SUBTIPO="DATA"* (ver 3.2.1) ou de *TIPO="DURACAO"* (ver 3.3 a seguir) definam implicitamente um intervalo de tempo (e.g. *todo o inverno* define implicitamente um intervalo entre 21 de Dezembro e 20 de Março), no tipo *INTERVALO*, os limites **têm de ser explícitos** para se poder considerar o subtipo *INTERVALO*.

3.3 TIPO = "DURACAO"

Corresponde a uma expressão *TEMPO* que refere uma duração de tempo contínuo. Mas, ao contrário das datas, trata-se de expressões que não exprimem propriamente a localização de um evento (calendarização do evento), mas sim quantificação temporal, sendo constituídas por nomes de unidades de medida de tempo e determinantes com

função de quantificadores (e.g.. numerais). Podem, por vezes, ser introduzidas, facultativamente, pela preposição *durante* e respondem adequadamente à interrogativa (*prep*) quanto tempo?.

Exemplos:

- *Fiquei* <EM ID=... CATEG="TEMPO" TIPO="DURACAO">**dois meses** em Lisboa.
- *O urso fica* <EM ID=... CATEG="TEMPO" TIPO="DURACAO">**todo o inverno** na toca.

3.4 TIPO = "FREQUENCIA"

O tipo *FREQUENCIA* corresponde às expressões *TEMPO* que exprimem uma repetição no tempo. Estas expressões respondem às interrogativas do tipo *com que frequência?*.

Exemplos:

- *Vou ver os meus pais* <EM ID=... CATEG="TEMPO" TIPO="FREQUENCIA">**diariamente** .
- *Vou ver os meus pais* <EM ID=... CATEG="TEMPO" TIPO="FREQUENCIA">**todos os dias**;
- *Vou ver os meus pais* <EM ID=... CATEG="TEMPO" TIPO="FREQUENCIA">**duas vezes por semana**;
- *Vou ver os meus pais* <EM ID=... CATEG="TEMPO" TIPO="FREQUENCIA">**dia sim dia não**.

3.5 TIPO = "GENERICICO"

Trata-se de expressões *TEMPO* que não se referem um data específica embora a expressão linguística seja composta por unidades lexicais que denotam elementos temporais. Estas expressões obedecem ao critério 3 definido em 3.1.1.

Exemplos:

- *Adoro* <EM ID=... CATEG="TEMPO" TIPO="GENERICICO">**o verão**.
- <EM ID=... CATEG="TEMPO" TIPO="GENERICICO">**Fevereiro** é o mês mais curto do ano.

3.6 Atributo *VAL_NORM*

O atributo *VAL_NORM* será apenas atribuído à algumas entidades *TEMPO*. Pretende ser um primeiro passo para a normalização de expressões temporais. Este atributo vai estar presente exclusivamente para as seguintes entidades temporais:

```
<EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA"
TEMPO_REF="ABSOLUTO" />
```

```
<EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="HORA" />
```

```
<EM ID=... CATEG="TEMPO" TIPO="DURACAO" />
```

3.6.1 Atributo *VAL_NORM* para expressões de subtipo *DATA* absoluta

Recorde-se que as expressões de subtipo *DATA* têm um atributo *TEMPO_REF* que poderá ter um dos seguintes valores: *ABSOLUTO*. Só é neste caso que se vai calcular a data absoluta correspondente a expressões temporais. (por outras palavras, não se vai tentar normalizar as datas referenciais no âmbito desta proposta).

O valor de *VAL_NORM* obedece ao seguinte formato:

```
<Era><Ano><Mes><Dia>T<Hora><Minuto>E<ESTACAO>LM<limite_aberto>
```

Onde:

- <Era> corresponde a 1 carácter que é + ou – conforme a data seja depois ou antes da nossa era;
- <Ano> corresponde a 4 caracteres de tipo dígito que representam o valor do ano ou então a subsequência "----";
- <Mes> corresponde a 2 caracteres de tipo dígito que representam o valor do mês ou então a subsequência "--";
- <Dia> corresponde a 2 caracteres de tipo dígito que representam o valor do dia ou então a subsequência "--";
- <Hora> corresponde a 2 caracteres de tipo dígito que representam o valor da hora ou então a subsequência "--";
- <Minuto> corresponde a 2 caracteres de tipo dígito que representam o valor dos minutos ou então a subsequência "--";
- <ESTACAO> corresponde a duas letras capitalizadas correspondente às estações do ano. *IN* para Inverno, *PR* para Primavera, *VE* para Verão e *OU* para Outono. No caso da data absoluta não ser expressa em termos de estação do ano, este campo terá por valor a subsequência "--";
- <limite_aberto> indica se a expressão normalizada de data absoluta introduz um intervalo de tempo com limite anterior ou limite posterior não determinado (em aberto). Os valores respectivos são "A" (no caso de limite anterior em aberto; este caso a expressão temporal corresponde ao limite posterior); no caso de limite posterior em aberto "P"; neste caso, a expressão temporal

corresponde ao limite anterior) e "-" quando a data absoluta não corresponde a um intervalo com um dos limites aberto.

Exemplos:

- *Nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ABSOLUTO" VAL_NORM="+19860103T--E--LM-">**a 3 de Janeiro de 1986**;
- *A Lia nasceu* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ABSOLUTO" VAL_NORM="+1996----T----EPRLM-">**na Primavera de 1996**;
- *A Inês vai à escola* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ABSOLUTO" VAL_NORM="+200709--T---E--LMP">**desde Setembro de 2007**.

NOTA: Se a expressão temporal data absoluta referir uma data anterior ao ano 9999 a.C. ou posterior ao ano 9999 d.C. , então o valor de *VAL_NORM* não é calculado atribuindo-se-lhe convencionalmente os valores:

-99999999T9999E--LM- ou +99999999T9999E--LM-

3.6.2 Atributo *VAL_NORM* para expressões de tipo *HORA*

No caso de expressões do subtipo *HORA*, também é utilizado o formato:

<Era><Ano><Mes><Dia>**T**<Hora><Minuto>**E**<ESTACAO>**LM**<limite_aberto>

que foi empregue para as expressões de subtipo *DATA*, *TEMPO_REF*="ABSOLUTO". Neste caso, no entanto, os campos correspondente a <ERA><Ano><Mes><Dia> correspondem necessariamente à subsequência "+-----" e o campo <ESTACAO> corresponde a "--". O campo <limite_aberto> pode corresponder a "D" ou "E" se a expressão de tipo *HORA* corresponder a uma expressão que introduz um intervalo aberto anterior ou posterior.

Exemplos:

- *A reunião durou* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="+-----T0220E--LM-">**2 horas e 20 minutos**.
- *Está disponível* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="+-----T15--E--LMA">**antes das 3:00 da tarde**.
- *O discurso de* <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="+-----T--50E--LM-">**50 minutos** foi muito maçador.

3.6.3 Atributo *VAL_NORM* para expressões de tipo *DURACAO*

Finalmente, para expressões de tipo *DURACAO*, o valor do atributo *VAL_NORM* corresponde ao valor utilizado para *VAL_DELTA* e exprime uma distância temporal.

Relembra-se aqui o formato deste valor:

A<digitos>*M*<digitos>*S*<digitos>*D*<digitos>*H*<digitos>*M*<digitos>*S*<digitos>

Onde:

- as letras *A*, *M*, *S*, *D*, *H*, *M*, *S* são constantes que devem aparecer nesta ordem e que correspondem respectivamente ao valores de Anos, Meses, Semanas, Dias, Horas, Minutos e Segundos;
- os <digitos> à direita das letras constantes correspondem ao número de Anos, Meses, Semanas, Dias, Horas, Minutos e Segundos que se devem adicionar ou diminuir à data de referência para obter o valor temporal da expressão anotada.

Exemplos:

- *Fiquei* <EM ID=... CATEG="TEMPO" TIPO="DURACAO" VAL_NORM="A0M2S0D0H0M0S0">**dois meses** em Lisboa.

Finalmente, para certas expressões de tipo *DURACAO*, que apresentam uma forma quantificação indefinida, tais como *vários anos*, *durante muitos dias*, *durante séculos*, etc. e na presença de quantificadores como *todo*, *inteiro*, etc. em *todo o inverno*, *o ano todo*, etc. não se especifica o atributo *VAL_NORM*.

- *Fiquei* <EM ID=... CATEG="TEMPO" TIPO="DURACAO" VAL_NORM="">**durante muitos meses** em Lisboa.
- *O urso fica* <EM ID=... CATEG="TEMPO" TIPO="DURACAO" VAL_NORM="">**todo o inverno** na toca.

4 Resumo das principais modificações

A principais modificações feitas em relação à proposta do Mini-HAREM são as seguintes:

- O tipo *CÍCLICO*, definido no guia de anotação do Mini-HAREM, desaparece;
- Expressões como *Natal* e *Páscoa* são consideradas como *TIPO="DATA"* mesmo que sejam cíclicas, na medida em que integram uma entidade que poderá ser ancorada num calendário;
- Propõe-se também não anotar entidades mencionadas que correspondam a períodos implícitos tais como os exemplificados na página 10 do guia de anotação:

Depois da IBM fui trabalhar para a Sun

Depois de trabalhar na IBM em 1993, fui trabalhar para a Sun

Estes exemplos contêm uma elipse e uma anotação como a que é proposta no guia de anotação faz com que, no primeiro exemplo, *IBM* seja considerada como uma entidade *TEMPO* enquanto que, no segundo, seja considerada como uma organização, o que parece pouco coerente e não corresponde aos objectivos gerais da tarefa de REM (ou pelo menos desta subtarefa de reconhecimento de entidades temporais).

5 Alguns exemplos de anotação

Nesta secção ilustram-se com alguns exemplos comentados a proposta de anotação das entidades *TEMPO* apresentada neste documento. Apresentam-se alguns exemplos que parecem colocar algumas dificuldades.

```
Aconteceu <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND"
SUBTIPO="DATA" TEMPO_REF="ABSOLUTO" VAL_NORM="-
99999999T9999E99LM-">na era dos dinossauros</EM>
```

A expressão *na era dos dinossauros* deve ser considerada como uma entidade a anotar (critério 1 e 2-6 da caracterização. Trata-se de uma expressão de subtipo *DATA* cuja granularidade é de centenas de bilhões de anos. (a era dos dinossauros começou há cerca de 240 bilhões de anos e eles povoaram a terra cerca de 165 bilhões de anos). Como se ultrapassa o ano -9999 o valor de *VAL_NORM* corresponde a "-99999999T9999E99LM-"> (ver nota no ponto 3.6.1)

```
<EM ID=... CATEG="TEMPO" TIPO="FREQUENCIA">De vez em quando</EM>
vou passear na montanha.
```

Trata-se de uma expressão de tipo frequência (responde à pergunta *com que frequência?*). que deverá ser anotada, em virtude do critério definatório 2-9 do ponto 3.1.1.1)

```
Fui ver o meu pai <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND"
SUBTIPO="DATA" TEMPO_REF="TEXTUAL">na semana seguinte ao Natal
passado</EM>
```

Trata-se de um exemplo um pouco artificial mas que vai ilustrar como considerar a referência em caso de expressões temporais encaixadas.

Primeiro, no que diz respeito à delimitação da expressão temporal tem de se considerar a totalidade da expressão (ver ponto 3.1.1.2). Trata-se obviamente de uma expressão de data referencial. Para determinar qual é a referência, considera-se o primeiro nível de encaixe da expressão (i.e. *na semana seguinte a X* (sendo *X* a referência, mesmo que esta referência também seja ela própria de tipo referencial).

Vimo-nos dois dias depois do Natal.

```
<ALT <EM ID="1" CATEG="TEMPO" TIPO="TEMPO_CALEND"
SUBTIPO="DATA" TEMPO_REF="TEXTUAL" SENTIDO="POSTERIOR"
VAL_DELTA="A0M0S0D2H0M0S0">dois dias depois do Natal</EM>
```

|

```
<EM ID="2" CATEG="TEMPO" TIPO="DURACAO" VAL_NORM="A0M0S0D2H0M0S0">dois dias</EM>
```

```
<EM ID="3" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL" SENTIDO="POSTERIOR">depois do Natal</EM>  
</ALT>
```

No exemplo acima, tem-se uma ambiguidade que permite duas interpretações:

Vimo-nos no dia 27 de Dezembro

Vimo-nos durante dois dias depois do dia 25 de Dezembro

Esta ambiguidade está representada através do elemento *ALT* que introduz as várias alternativas, e do elemento '|', que as separa.

```
<EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="TEXTUAL">Naquela semana</EM> vimo-nos <EM ID=... CATEG="TEMPO" TIPO="DURACAO" VAL_NORM="A0M0S0D3H0M0S0">três manhãs</EM>.
```

O exemplo acima tem semelhanças com uma das interpretações do exemplo anterior. Neste caso há dois elementos de categoria *TEMPO* que deverão ser anotados. Um é do subtipo *DATA* relativa com referência textual, e com atributo *SENTIDO* e *VAL_DELTA* indefinidos. O segundo é do tipo *DURACAO* (responde à pergunta *quanto tempo?*).

Nota-se que esta representação não distingue partes de dias como *manhã* ou *tarde*. O valor normalizado de *três manhãs* vai corresponder aqui ao valor normalizado de *três dias*.

```
Vou a Madrid <EM ID=... CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA" TEMPO_REF="ENUNCIACAO" SENTIDO="POSTERIOR" VAL_DELTA="A1M0S0D0H0M0S0">para o ano que vem</EM>
```

Este exemplo permite lembrar qual a delimitação duma expressão temporal de tipo *PREP+unidade_medida_temporal+que+verbo vir* (ou *verbo passar*).

Trata-se de uma expressão temporal relativa, cuja referência é o tempo da enunciação. A distância temporal entre a referência e a expressão corresponde a 1 ano.

6 Referências bibliográficas

[Cardoso & Santos 06] Cardoso, N., Santos D.: *Directivas e categorias para identificação semântica na colecção dourada do HAREM*, versão de 29 de Março de 2006.

[Hagège & Tannier 07] Hagège C., Tannier X.: *XRCE-T: A Temporal XIP Module for TempEval Campaign*. TempEval Temporal Relation Identification. SemEval Workshop, ACL, Budapest, June 2007.

[Reichenbach 47] Reichenbach H.: *Elements of Symbolic Logic*. 1947.

[Saurí et al. 06] Saurí, R., Littman, J., Knippen, B., Gaizauskas, R., Setzer, A., Pustejovsky, J.: *TimeML Annotation Guidelines (2006)* www.timeml.org/site.

[Verhagen et al. 07] Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., Pustejovsky, J.: *SemEval-2007 – Task 15: TempEval Temporal Relation Identification*. SemEval workshop in ACL (2007).

7 Agradecimentos

Agradece-se à Diana Santos e à Cláudia Freitas pela releitura da versão inicial e pelas muitas sugestões que permitiram melhorar esta proposta.

ALGUMAS PRECISÕES SOBRE O TEMPO

O estatuto da vírgula nas expressões de tipo TEMPO

A vírgula, poderá ou não ser um separador nas expressões temporais.

Caso 1) a vírgula faz parte integrante da expressão temporal

Por exemplo:

Quarta-feira, 13 de Fevereiro de 2008
Fevereiro 13, 2008

Nestes exemplos a vírgula está integrada dentro da expressão:

```
<EM CATEG=TEMPO TIPO=TEMPO_CALEND SUBTIPO=DATA  
TEMPO_REF=ABSOLUTO VAL_NORM="+20080213T----E--LM-"> Quarta-feira, 13  
de Fevereiro de 2008 </EM>
```

```
<EM CATEG=TEMPO TIPO=TEMPO_CALEND SUBTIPO=DATA  
TEMPO_REF=ABSOLUTO VAL_NORM="+20080213T----E--LM-">Fevereiro 13,  
2008</EM>
```

Caso 2) assinala a coordenação de expressões temporais.

Por exemplo:

Nos dias 25, 26 e 27 de Janeiro

Neste exemplo, há 3 expressões temporais a considerar:

```
<EM CATEG=TEMPO TIPO=TEMPO_CALEND SUBTIPO=DATA  
TEMPO_REF=ABSOLUTO VAL_NORM="+----0125T----E--LM-">Nos dias  
25</EM>, <EM CATEG=TEMPO TIPO=TEMPO_CALEND SUBTIPO=DATA  
TEMPO_REF=ABSOLUTO VAL_NORM="+----0126T----E--LM-">26</EM> e  
<EM CATEG=TEMPO TIPO=TEMPO_CALEND SUBTIPO=DATA  
TEMPO_REF=ABSOLUTO VAL_NORM="+----0127T----E--LM-">27 de  
Janeiro</EM>.
```

Nota-se que, na normalização, « *Nos dias 25* » e « *26* » foram normalizados como « *25 de Janeiro* » e « *26 de Janeiro* » respectivamente.

Fracções de unidades de medidas temporais na normalização

Propomos, para a normalização de fracções de unidades de medidas temporais, considerar as seguintes equivalências:

1 milénio = 1000 anos
1 século = 100 anos
1 ano = 12 meses
1 mês = 30 dias (aproximação)
1 quinzena = 14 dias
1 semana = 7 dias
1 dia = 24 horas
1 h = 60 minutos
1 minuto = 60 segundos (utilizado nas durações)

Assim, se encontramos uma expressão como: “*Durante meio ano*” esta expressão deverá ser normalizada da mesma forma que “durante 6 meses”, isto é
VAL_NORM=“A0M6S0D0H0M0S0”

Propomos que, na normalização das expressões temporais, só se levem em conta valores inteiros.

Assim, quando a divisão da unidade de medida temporal convertida tem um resto diferente de 0, consideramos apenas o valor inteiro ignorando o resto da divisão.

Isto vai acontecer nas fracções de segundos (para durações) e nas fracções de minutos para horas. Poderá também acontecer em casos como “*durante 2/3 da semana*” (se este caso se produzir) onde a divisão $(2*7)/3$ tem um resultado de 4, ignorando o resto de 0.666.

Por exemplo:

Durante 1/5 do ano

terá o valor de VAL_NORM=“A0M2S0D0H0M0S0” que corresponde a $12/5 \sim 2$ meses

Durante 2/3 do século

terá o valor de VAL_NORM=“A66M0S0D0H0M0S0” que corresponde a $(2*100)/3 \sim 66$ anos