Capítulo 1

Segundo HAREM: Modelo geral, novidades e avaliação

Paula Carvalho, Hugo Gonçalo Oliveira, Diana Santos, Cláudia Freitas e Cristina Mota

Cristina Mota e Diana Santos, editoras, *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*, 2008, Capítulo 1, p. 11–31.

No Segundo HAREM, foi mantida a filosofia subjacente ao Primeiro HAREM, nomeadamente o modelo semântico (Santos, 2007d) e o modelo geral de avaliação (Santos et al., 2007). Contudo, e como seria de esperar, procurou-se corrigir e aperfeiçoar algumas arestas em relação à edição anterior, o que se reflectiu numa caracterização mais precisa e linguisticamente motivada de certas entidades mencionadas (EM), bem como numa avaliação mais justa dos sistemas. Esta segunda edição do HAREM passou também a incluir duas novas tarefas/pistas, designadamente a tarefa de reconhecimento e normalização de expressões temporais e a tarefa de reconhecimento de relações semânticas entre EM, o ReRelEM, a que dedicamos os capítulos 2 e 4, respectivamente, deste livro.

Neste capítulo, discutimos especificamente a pista geral de reconhecimento de entidades mencionadas no Segundo HAREM, a que nos referiremos, daqui em diante, como HAREM clássico. Mais especificamente, na secção 1.1, apresentamos, de forma sucinta, o modelo semântico subjacente ao HAREM. Em 1.2, centramo-nos na proposta de classificação das EM tida em consideração no Segundo HAREM, bem como nas alterações que esta sofreu em relação à proposta de classificação utilizada no Primeiro HAREM. Na secção 1.3, discutimos as melhorias introduzidas no Segundo HAREM, face à primeira edição. Em 1.4, descrevemos o processo de constituição das colecções usadas especificamente no âmbito desta avaliação, nomeadamente a colecção do Segundo HAREM e a respectiva colecção dourada (CD). Fazemos, ainda, uma breve caracterização de ambas as colecções, e enumeramos as principais fases inerentes ao processo de anotação e revisão da CD. Por fim, na secção 1.5, discutimos os resultados obtidos pelos sistemas participantes, nos diferentes tipos de avaliação tidos em conta no Segundo HAREM.

1.1 Filosofia do HAREM

O modelo semântico do HAREM assenta em dois aspectos essenciais, que o distinguem de outros modelos vulgarmente utilizados na avaliação de REM[1]. Esses aspectos prendem-se nomeadamente com (i) a ideia de que identificação e classificação de uma dada expressão como entidade mencionada depende exclusivamente do seu uso em contexto, não estando lexicalmente "presa" a nenhum dos atributos a que possa estar associada noutros recursos linguísticos, por exemplo, dicionários, almanaques, ontologias e com (ii) o facto de ser possível atribuir mais do que uma classificação (categoria, tipo e/ou subtipo) a uma mesma EM (considerando-a portanto vaga entre as várias classificações), se o contexto em que a mesma se encontra não permitir escolher apenas uma delas.

Embora, na maioria das avaliações levadas a cabo neste domínio, a classificação das entidades mencionadas esteja intimamente relacionada com a sua caracterização (semântica) nos recursos lexicais, no HAREM considera-se que essa caracterização só pode ser feita numa situação de uso concreto da língua. Não consideramos, portanto, que uma EM possui, intrinsecamente, um dado significado, que pode eventualmente ganhar diferentes nuances conforme o contexto que essa EM integre. Isso implicaria, entre outras coisas, assumir a existência de "um significado de base" e de "um significado derivado do uso". Como referimos antes, a nossa posição é a de que o significado de qualquer EM é, à partida, quase imprevisível, e só pode ser compreendido através da sua função em contexto. De facto, apesar de poder parecer fazer sentido definir lexicalmente algumas categorias

[1] Para uma análise contrastiva entre o HAREM e outras avaliações realizadas neste domínio, em particular o MUC e o CoNLL, veja-se Santos (2007c), Santos e Cardoso (2007b) e Seco (2007).

semânticas, como é o caso paradigmático de *país*[2], não é obrigatório que exista uma relação de univocidade entre esse conceito e uma única categoria ou conjunto de categorias que considerámos pertinentes no HAREM, nomeadamente, LOCAL e/ou ORGANIZACAO. Por exemplo, *Portugal* pode ser usado para fazer referência a um conjunto variado de sentidos (como ilustrado nos exemplos (1.1) a (1.5)[3]), sem que nenhum deles tenha necessariamente primazia sobre os outros.

(1.1) Regressou então a <EM ID="ub-67792-10" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">**Portugal**</EM>, onde iniciou meteórica carreira na experimentação de novas formas de expressão

(1.2) O acordo político quanto à revisão foi obtido durante a Presidência Alemã, tendo cabido a <EM ID="a46996-5" CATEG="ORGANIZACAO" TIPO="ADMINISTRACAO">**Portugal**</EM> concluir o processo de revisão.

(1.3) Este debate passou completamente ao lado de <EM ID="2-dftre765-" CATEG="PESSOA" TIPO="POVO">**Portugal**</EM>

(1.4) o problema do PSD é começar a ter só um <EM ID="ub-24360-32" CATEG="ABSTRACCAO" TIPO="IDEIA">**Portugal**</EM> ou dois dentro de si

(1.5) <EM ID="x-1G" CATEG="PESSOA" TIPO="GRUPOMEMBRO">**Portugal**</EM> perdeu com a Suíça por 2-0

Mas, se para o exemplo de *Portugal* não é difícil acordar sobre uma definição, a de "país" (a qual, segundo uma certa visão da língua, estaria, pelo menos, associada às "variações" LOCAL e ORGANIZACAO), o mesmo não acontece para EM mais abstractas. Por exemplo, *Big-Bang* tanto pode ser definida como uma "teoria" sobre a criação do universo (exemplo (1.6)) ou como uma "explosão cósmica" (exemplo (1.7)), sendo, respectivamente, classificada como ABSTRACCAO e ACONTECIMENTO.

(1.6) A radiação de origem cósmica, prevista pelo <EM ID="bb1" CATEG="ABSTRACCAO">**Big Bang**</EM> seria descoberta em 1964, quase acidentalmente, por Arno Penzias e Robert Wilson.[4]

(1.7) Esse ponto deve ter sido o começo dos tempos, pelo qual tem início a expansão das galáxias, que os cosmologistas descrevem como uma explosão, ou seja, o <EM ID="bb2" CATEG="ACONTECIMENTO">**Big Bang**</EM>[5]

Diferentemente de outras avaliações de REM, em que se considera que as entidades devem receber uma única classificação, mesmo que arbitrária em última análise, no HAREM propomos que as entidades poderão (e deverão) estar associadas a mais do que uma etiqueta, sempre que o contexto em que essas EM ocorrem não permita seleccionar uma de

[2] Por exemplo, na Wikipédia, *país* é definido como um "território social, política, cultural e geograficamente delimitado" e na Infopédia como um "espaço demarcado por fronteiras geográficas e dotado de soberania própria; estado; nação".
[3] Para mais pormenores sobre o esquema de anotação, veja-se a próxima secção ou o apêndice A.
[4] http://www.if.ufrj.br/teaching/cosmol/expriml.html, em 24 de Outubro de 2008
[5] http://www.coladaweb.com/astronomia/bigbang.htm, em 24 de Outubro de 2008

entre as várias análises possíveis (Santos, 2007d). Trata-se, pois, de preservar aquilo que consideramos uma propriedade essencial da linguagem natural, a vagueza, que não pode ser resolvida nem eliminada, de modo a não se perder informação (Santos, 1997, 2006).

Ilustramos, em seguida, alguns exemplos de EM vagas – extraídas da CD do Segundo HAREM – retomando o caso de *Portugal*.

(1.8) Pela mão do ministro Freitas do Amaral, e sem necessidade alguma, <EM ID="a66435-10" CATEG="ORGANIZACAO|PESSOA" TIPO="ADMINISTRACAO|POVO">**Portugal**</EM> foi enxovalhado, coberto de vergonha e de cobardia, por um dos mais tristes textos políticos que já alguém escreveu.

(1.9) Mais de 32 mil pessoas poderiam morrer se uma pandemia de gripe humana de origem aviária atingisse <EM ID="ub-28874-3" CATEG="PESSOA|LOCAL" TIPO="POVO|HUMANO" SUBTIPO="|PAIS">**Portugal**</EM>

(1.10) Os dois reinos católicos, <EM ID="a66435-5" CATEG="PESSOA|ORGANIZACAO" TIPO="GRUPOIND|ADMINISTRACAO">**Portugal**</EM> e Espanha, partiram à conquista do mundo e tornaram-se Impérios marítimos do <EM ID="aa66435-54" CATEG="LOCAL|LOCAL" TIPO="FISICO|HUMANO" SUBTIPO="REGIAO|DIVISAO">**Novo Mundo**</EM>

Em (1.8), *Portugal* tanto pode referir o governo (ORGANIZACAO ADMINISTRACAO) como o povo (PESSOA POVO) português; em (1.9), a vagueza observa-se entre esta última análise (a de PESSOA POVO) e a de LOCAL; por fim, no exemplo (1.10), *Portugal*, tanto pode referir o governo português como um grupo indeterminado de pessoas individuais que não possuem um nome convencional (PESSOA GRUPOIND). A vagueza não se observa simplesmente ao nível da categoria (CATEG) das EM; em muitos casos, esta propriedade estabelece-se a um nível de subcategorização mais fino das entidades, nomeadamente no que respeita aos tipos e subtipos envolvidos. Por exemplo, em (1.10), *Novo Mundo*, que no contexto em questão faz menção a um LOCAL, pode representar tanto um local da geografia física (LOCAL FISICO REGIAO) como da geografia humana (LOCAL HUMANO DIVISAO).

Não queremos dar, contudo, a ideia (completamente errada) de que esta situação se passa sobretudo no caso dos nomes de países ou cidades, embora este seja um exemplo tão discutido na literatura que é incontornável não o referir (veja-se, a propósito, a vasta literatura citada em Santos (2007d)). Apresentamos, em seguida, outros casos, completamente distintos dos anteriormente ilustrados, em que, uma vez mais, conceitos complexos se desdobram em sentidos múltiplos, no texto.

(1.11) O carácter diferente da <EM ID="H2-dftre765-41" CATEG="ABSTRACCAO|ACONTECIMENTO" TIPO="IDEIA|EFEMERIDE">**Reforma Inglesa**</EM> deve-se ao facto de ter sido promovida inicialmente pelas necessidades políticas de Henrique VIII.

(1.12) Assim aceitam os dois sacramentos do <EM ID="H2-dftre765-122" CATEG="ABSTRACCAO|OBRA" TIPO="IDEIA|PLANO">**Evangelho**</EM>: o Santo Batismo, através do qual a pessoa é feita membro da Igreja de Cristo.

No exemplo (1.11), tanto podemos entender *Reforma Inglesa* como um ACONTECIMENTO ou como uma ABSTRACCAO, mais especificamente uma IDEIA, e nenhuma das interpretações

exclui a outra. O mesmo se passa em relação a *Evangelho*, no exemplo (1.12), que pode corresponder quer a uma ABSTRACCAO quer a uma OBRA.

Naturalmente, a existência de vagueza entre várias interpretações depende do número de interpretações que o modelo semântico reputa como relevantes. Quanto mais diferenças finas de sentido quisermos reconhecer e anotar, maior será a possibilidade de não nos virmos obrigados a decidir por uma única interpretação, ou, por outras palavras, maior será a probabilidade de as EM serem consideradas vagas.

Esta questão não é meramente teórica e corresponde a uma fatia significativa dos casos que tivemos de anotar. Para um resumo quantitativo, veja-se a tabela 1.1, mais à frente, em que apresentamos a quantificação dos casos de vagueza presentes na CD, isto é, as EM em que não foi possível atribuir uma única classificação.

1.2 Esquema de anotação no Segundo HAREM

Nesta secção, procuramos, por um lado, fazer uma breve descrição do formato das etiquetas utilizado no Segundo HAREM, e, por outro, apresentar a proposta de classificação adoptada na anotação das EM, apontando as principais diferenças entre esta proposta e a que foi utilizada no Primeiro HAREM. Para informações mais detalhadas, sugerimos a consulta das directivas, no apêndice A.

1.2.1 Sintaxe das anotações

A anotação no Segundo HAREM foi feita de acordo com o formato XML. No que se refere às EM, todas as etiquetas começam com <EM ID="xxx"> e acabam com </EM>. O único atributo obrigatório é o identificador (ID), que, para facilidade de processamento, restringimos a uma combinação de apenas letras não acentuadas (maiúsculas ou minúsculas), algarismos, e os caracteres "-" e "_". Contrariamente ao que acontecia no Primeiro HAREM, cuja sintaxe de anotação das EM obrigava à explicitação da respectiva categoria (a qual incluía a etiqueta de abertura e de fecho da EM, por exemplo <PESSOA> e </PESSOA>), no Segundo HAREM a sintaxe das anotações é mais flexível, combinando numa mesma caracterização de saída (i) apenas a identificação (ii) a identificação e classificação de categorias, (iii) a identificação e classificação de categorias e tipos, (iv) a identificação e classificação de categorias, tipos e subtipos e (v) a identificação e categorias, tipos, subtipos e outros atributos previstos na classificação das EM (em concreto, os atributos previstos na classificação de expressões temporais ou na identificação de relações entre EM), sendo todas estas classificações opcionais.

Nos casos em que existem diferentes possibilidades de segmentação de uma dada sequência no texto, as diferentes análises alternativas associadas a essa sequência encontram-se compreendidas entre as etiquetas <ALT> e </ALT>, estando separadas entre si pelo símbolo "|"[6]; as diferentes EM identificadas no âmbito dessas análises recebem cada uma delas um ID distinto (cf. exemplo (1.13)).[7]

(1.13) aproximava a <ALT> <EM ID="2-dftre765-10" CATEG="ABSTRACCAO" TIPO="DISCIPLINA">**Igreja de Inglaterra**</EM> | <EM ID="2-dftre765-106-a" CATEG="ABSTRACCAO"

[6] Neste caso, o "|" não faz parte da linguagem XML, é uma representação própria do HAREM.

[7] Embora a notação sejam muito parecida com a do MUC-7 (Chinchor, 1998), chamamos a atenção para que nem o sentido de ALT nem o uso do símbolo "|" correspondem ao desta última avaliação conjunta.

TIPO="DISCIPLINA">**Igreja**</EM> de <EM ID="2-dftre765-1" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">**Inglaterra**</EM> </ALT> do calvinismo.

O mesmo símbolo é também utilizado para separar as diferentes possibilidades de análise associadas a uma mesma EM, uma EM vaga (como ilustrado nos exemplos (1.8)-(1.10), anteriormente apresentados).

1.2.2 Classificação das EM

O conjunto de etiquetas usado no Segundo HAREM não é significativamente distinto do usado no Primeiro HAREM (cf. figura 1.1). O número de categorias nas duas avaliações é idêntico: dez categorias, as quais permaneceram intactas em relação à sua designação, excepto no que respeita a VARIADO, que foi substituída por OUTRO. Estas categorias pareceram-nos, pois, as mais pertinentes no âmbito de uma avaliação de REM em português, mas não rejeitamos a possibilidade de outras o poderem ser também, nomeadamente tendo em conta os interesses específicos de cada participante. Nesta perspectiva, a categoria, tipo ou subtipo OUTRO serve precisamente para dar conta de outras possibilidades de classificação das EM que não estejam contempladas no elenco de categorias (e/ou respectivos tipos e/ou subtipos) que definimos.

As categorias ACONTECIMENTO, VALOR e COISA não sofreram quaisquer alterações, exceptuando-se a inclusão do tipo OUTRO, que passou a ser um tipo possível de qualquer categoria.

Pelo contrário, as categorias LOCAL e TEMPO foram as que sofreram alterações mais substanciais, tendo sido alterados e/ou rebaptizados a maioria dos tipos anteriormente previstos. Além disso, estas categorias passaram ainda a incluir subtipos.

A categoria TEMPO encontra-se detalhadamente descrita no capítulo 2, pelo que não nos ocuparemos dela aqui.

No que respeita a LOCAL, deixámos de considerar o tipo CORREIO como uma EM, preferindo a marcação separada de ruas, estados e países dentro de moradas. Além disso, a informação abrangida, no Primeiro HAREM, pela etiqueta LOCAL ALARGADO passou a ser considerada como informação adicional em relação aos tipos ADMINISTRATIVO ou GEOGRAFICO (agora rebaptizados de HUMANO ou FISICO).

Deste modo, criou-se uma tripartição da categoria LOCAL em FISICO, HUMANO e VIRTUAL, em que FISICO substitui o anterior termo GEOGRAFICO, e HUMANO o anterior termo ADMINISTRATIVO.

Além da categoria TEMPO, esta foi a única categoria em que os participantes demonstraram interesse numa classificação mais fina em subtipos. A definição destes subtipos resultou de uma discussão entre os participantes especificamente interessados nesta categoria e a organização, reflectindo, assim, a soma das várias sensibilidades, experiências e opiniões das duas partes envolvidas.

A categoria PESSOA passou a incluir um novo tipo, que designámos como POVO, para dar conta de casos em que uma dada entidade, geralmente associada a um determinado local, é usada para referir a população desse local. Este conceito não era integralmente captado por nenhum dos tipos contemplados nas anteriores directivas.

A categoria ORGANIZACAO deixou de incluir o tipo SUB, que, na verdade, correspondia a uma subespecificação (ou se quisermos, subtipo) dos tipos ADMINISTRACAO, INSTITUICAO ou EMPRESA. Estes três tipos, já presentes no Primeiro HAREM, foram mantidos, e usados quer para a instituição (ou empresa, etc.) completa quer para uma subparte dela.

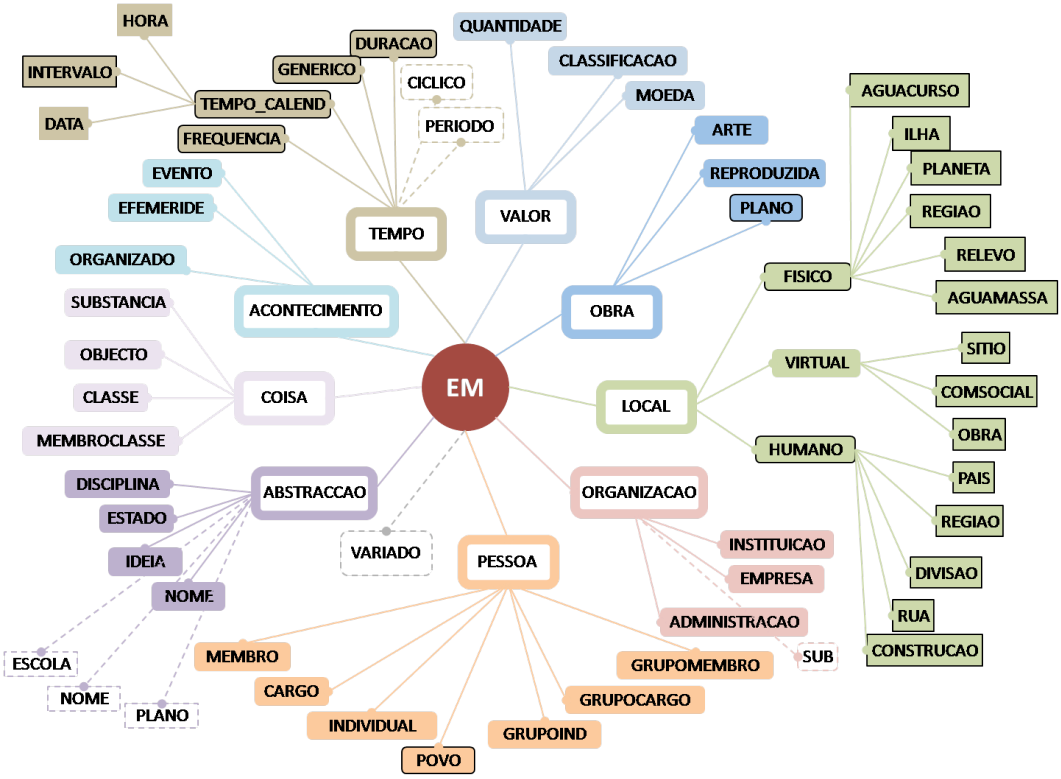


Figura 1.1: Árvore de categorias no Segundo HAREM: as categorias, tipos e subtipos representados nas caixas com contorno sólido preto só existem no Segundo HAREM; as categorias, tipos e subtipos representados nas caixas com contorno pontilhado só existem no Primeiro HAREM

A categoria OBRA passou a incluir o tipo PLANO (que anteriormente correspondia a um tipo da categoria ABSTRACCAO), deixando de parte o tipo PUBLICACAO, que, tal como CORREIO, correspondia a uma estrutura complexa, que preferimos não contemplar como EM.

A categoria ABSTRACCAO foi consideravelmente simplificada, retendo apenas os tipos DISCIPLINA, ESTADO, IDEIA e NOME. Por um lado, foram retirados desta categoria os tipos MARCA (convertido para a categoria COISA de tipo CLASSE ou IDEIA) e PLANO (transferido para categoria OBRA de tipo PLANO). Por outro lado, os tipos DISCIPLINA, ESCOLA e OBRA passaram a ser todos eles representados por um único tipo, DISCIPLINA.

Cada uma das categorias, tipos e subtipos referidos encontram-se ilustrados no apêndice E.

1.3 Melhorias no Segundo HAREM

A repetição de qualquer evento, neste caso, um evento de avaliação, não pode/deve corresponder, na nossa perspectiva, a uma mera cópia do evento anterior, sobretudo se considerarmos que há espaço para introdução de melhorias. É assim que entendemos o Se-

gundo HAREM: uma avaliação que tenta reter os aspectos positivos do Primeiro HAREM, mas que, naturalmente, procura melhorar os aspectos menos positivos, alguns dos quais previamente identificados aquando da realização do balanço do Primeiro HAREM (Santos e Cardoso, 2007b). Nas próximas subsecções, abordaremos as principais melhorias, em nosso entender, introduzidas especificamente no HAREM clássico.

1.3.1 Delimitação e classificação das EM

Ainda que, na maior parte dos casos, os critérios para a identificação e classificação de EM propostos no Primeiro HAREM tenham sido aplicados com sucesso ao reconhecimento de entidades mencionadas em português, considerámos que, em casos pontuais, a definição operacional de EM deveria ser ligeiramente modificada, de modo a ter uma classificação mais coerente e precisa, a qual pudesse, ao mesmo tempo, servir adequadamente os propósitos das aplicações em extracção e/ou recuperação de informação.

Neste sentido, as EM estruturalmente complexas, como moradas (anterior `LOCAL CORREIO`) e referências bibliográficas (anterior `OBRA PUBLICACAO`), embora relevantes num contexto de extracção de informação, deixaram de ser consideradas no Segundo HAREM, dada a dificuldade em motivar a sua identificação como entidades, numa tarefa de REM. De facto, neste contexto, parece-nos mais adequado privilegiar a análise autónoma das EM que constituem estas sequências, do que as sequências em si mesmo.

Numa outra perspectiva, mas tendo igualmente em linha de conta a própria noção de unidade lexical e semântica das EM, deixámos de fragmentar palavras ou expressões (compostas) cujos constituintes não obedeciam ao critério formal (das maiúsculas) previamente definido no HAREM para a identificação das EM. Concluímos que, nuns casos, as palavras ou expressões que anteriormente haviam sido classificadas como EM não o eram de facto (caso de *de Belém* para identificar *pastel de Belém* como EM, que agora não foi assim considerado) e que, noutros casos, toda a expressão deveria ser identificada como EM, desde que os elementos grafados em minúsculas integrassem a lista das minúsculas permitidas (cf. apêndice A, secção A.6), a qual foi criada para o efeito no âmbito desta avaliação (caso de *doença* em *doença de Chagas*).

Um outro caso em que decidimos refinar a identificação das EM está directamente relacionado com a representação de intervalos de valores e/ou especificação mais fina desses valores. Em particular, passámos a considerar intervalos de valores, tais como *entre 3 e 4%* ou *de 5 a 10 kg*, como uma única EM, e não duas como acontecia no Primeiro HAREM. Os quantificadores ou modificadores que permitem precisar o valor da entidade, como acontece em *cerca de 200 gramas*, *menos de 10%* ou *aproximadamente 15 euros*, também passaram a ser incluídos no âmbito da EM.

1.3.2 Representação sistemática das análises alternativas

No Primeiro HAREM, demos conta da possibilidade de uma dada sequência poder ser segmentada de forma distinta, nomeadamente nos casos em que essa sequência corresponde a uma EM estruturalmente ambígua, como ilustrado em (1.13), ou, numa outra perspectiva, quando não há certeza de que a sequência em análise corresponda efectivamente a uma entidade mencionada, explicitando-se, assim, a possibilidade de a mesma ser, ou não, identificada como EM, como ilustrado em (1.14).

(1.14) Portugal e Espanha, partiram à conquista do mundo e tornaram-se <ALT> <EM ID="a66435-5" CATEG="OUTRO">**Impérios**</EM> | Impérios </ALT> marítimos;

No Segundo HAREM, a etiqueta ALT passsou ainda a ser utilizada para representar, de forma sistemática, a estrutura interna das entidades constituídas por outras EM, como é o caso da EM que apresentamos em (1.15).

(1.15) <ALT> <EM ID="a55968-47" CATEG="PESSOA" TIPO="CARGO">**presidente da Câmara de Nova Iorque**</EM> | presidente da <EM ID="a55968-" CATEG="ORGANIZACAO" TIPO="ADMINISTRACAO">**Câmara de Nova Iorque**</EM> | presidente da <EM ID="a55968-475a" CATEG="ORGANIZACAO" TIPO="ADMINISTRACAO">**Câmara**</EM> de <EM ID="a55968-47" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="DIVISAO">**Nova Iorque**</EM> </ALT>

Este procedimento pode ser, de certo modo, encarado como uma forma de representar o encaixe de EM, uma situação não contemplada no Primeiro HAREM e que pode ter interesse a vários níveis. Por exemplo, além de permitir uma análise mais fina sobre o próprio mecanismo de composição de certas EM, possibilita a identificação de EM que, de outro modo, não seriam analisadas. Tendo em conta que uma das indicações fornecidas nas directivas do Primeiro HAREM apontava no sentido de marcar preferencialmente a EM mais longa (Cardoso e Santos, 2007), a identificação de, por exemplo, *Câmara de Nova Iorque* no exemplo acima não seria considerada. Isto poderia trazer inconvenientes, por exemplo, aos participantes que estivessem interessados em reconhecer especificamente organizações.

Apresentamos, no apêndice D, a lista de regras criadas para o efeito. Estas regras gerais foram, em alguns casos, refinadas, em função das propriedades lexicais e/ou semânticas dos constituintes de certas EM. Por exemplo, a regra PESSOA de LOCAL não deve ser empregue nos casos em que o indivíduo, referido pelo seu título nobiliário (que marcámos como CARGO) corresponde a uma das seguintes palavras: *conde*, *duque* e *marquês*. Esta opção deve-se ao facto de termos considerado como demasiado remota, e daí pouco pertinente, a relação que se estabelece entre a menção ao título e ao nome do local (caso de *Conde de Ourém*, *Duque de Bragança* e *Marquês de Pombal*). Não segmentámos também em constituintes menores as expressões classificadas como OBRA, se estas estiverem delimitadas por aspas ou plicas. Além disso, também não considerámos possível a segmentação de locais do tipo *Mosteiro dos Jerónimos*, no sentido em que se considera que é esta EM (CONSTRUCAO) que está na base da denominação de um dos seus constituintes, *Jerónimos* (DIVISAO), e não o contrário.[8]

1.4 Recursos

No Segundo HAREM, foram desenvolvidos e disponibilizados vários recursos, tanto para treino como para a avaliação propriamente dita dos sistemas. Para treino, disponibili-

[8] Para os nossos leitores não familiarizados com a história de Lisboa, convém talvez referir que o Mosteiro dos Jerónimos foi assim baptizado devido ao facto de este ter sido habitado pelos Jerónimos, os frades pertencentes à ordem de São Jerónimo, após ter sido erigido no século XVI. Actualmente, *Jerónimos* é usado (pelo menos, pelos lisboetas) para designar tanto o mosteiro como a zona onde este se encontra. Temos pois um caso em que o LOCAL vago *Jerónimos* provém do local (construção) *Mosteiro dos Jerónimos*, não sendo, por isso, parafraseável por "Mosteiro que se situa nos Jerónimos" (contrariamente, ao caso da *Torre de Pisa*, que é parafraseável por "Torre que se situa em Pisa").

zámos diferentes colecções anotadas de acordo com as directivas do HAREM clássico e da pista do TEMPO, assim como um *Exemplário* (cf. apêndice E), isto é, um conjunto de exemplos com EM ilustrativas de cada uma das categorias, tipos e subtipos previstos nas directivas do HAREM clássico (cf. apêndice A).

Para efectuar a própria avaliação, criámos a colecção do Segundo HAREM – a colecção que todos os sistemas tiveram de anotar – e a colecção dourada, um subconjunto da colecção do Segundo HAREM em que foi feita a anotação humana de tudo o que pretendíamos avaliar. Em seguida, descrevemos estes recursos com mais pormenor.

1.4.1 Constituição das colecções do Segundo HAREM

A colecção do Segundo HAREM é constituída por 1040 documentos (15737 parágrafos, 670610 palavras), entre os quais se encontram, como referimos antes, os documentos seleccionados para a colecção dourada. A colecção dourada é constituída por 129 documentos (correspondendo a 2274 parágrafos perfazendo 147991 palavras), representando cerca de 12% dos documentos que compõem a colecção do Segundo HAREM.

Os documentos da colecção do Segundo HAREM foram seleccionados tendo essencialmente em consideração os seguintes requisitos: (i) o português de Portugal e o do Brasil deveriam estar equitativamente representados na colecção, (ii) os documentos deveriam contemplar diferentes géneros e registos textuais, e (iii) a colecção deveria incluir algum material utilizado no Primeiro HAREM (nomeadamente, de forma a permitir comparar o desempenho dos sistemas nesses documentos) e noutras avaliações, como é o caso da colecção CHAVE (Santos e Rocha, 2005), a qual tem vindo a ser usada na avaliação de sistemas de respostas automáticas a perguntas (QA@CLEF (Giampiccolo et al., 2008)) e de recolha de informação geográfica. Neste último caso, os textos foram escolhidos com base na penúltima edição do GeoCLEF: para cada um dos 25 tópicos do GeoCLEF 2007 (Mandl et al., 2008), foram incluídos todos os documentos classificados como relevantes e dez documentos classificados como irrelevantes. Tal permitirá, no futuro, estudar, por exemplo, a influência e a relevância de REM na recuperação de informação geográfica.

A cada documento da colecção foram associadas diversas informações que caracterizam o documento. Entre outras propriedades, destacamos: variante de português, género e nome da fonte. A distribuição dos valores dessas propriedades na colecção do Segundo HAREM, bem como em cada uma das colecções douradas, encontra-se no apêndice H.

1.4.2 Processo de anotação da CD

A colecção dourada, como referimos anteriormente, constitui um subconjunto da colecção do Segundo HAREM, com base na qual os sistemas são avaliados. Numa primeira fase, o processo de anotação da CD foi cruzado, isto é, duas anotadoras anotaram o mesmo conjunto de textos. Esse processo foi levado a cabo com a ajuda da ferramenta Etiquet(H)AREM (ver apêndice F para informações mais detalhadas sobre esta ferramenta). As anotações foram posteriormente confrontadas/comparadas, recorrendo a um programa que apresentava as diferenças, com base na saída do programa `Alinhador` (capítulo 5). As diferenças encontradas por este programa foram então reanalisadas e discutidas pelas anotadoras (e, em alguns casos, por toda a organização), de forma a chegar a uma anotação consensual. Numa fase posterior, em que as directivas já se encontravam afinadas, os textos da CD passaram a ser alternadamente anotados por cada uma das

anotadoras. Casos problemáticos ou duvidosos eram expostos a (e discutidos por) toda a organização, de modo a tentar encontrar uma solução de anotação em que, pelo menos, a maioria estivesse de acordo.

Depois de anotada toda a CD, procedemos à sua revisão, a qual foi realizada em três fases distintas, mas complementares: numa primeira fase, levámos a cabo uma revisão sequencial dos documentos de toda a CD; seguidamente, efectuámos uma revisão fina e exaustiva das EM por categoria (tendo sempre, naturalmente, em conta o contexto em que estavam integradas), revisão essa levada a cabo por três pessoas[9]; finalmente, revimos especificamente os casos das EM compreendidas entre as etiquetas <ALT> e </ALT>.

Já após a apresentação dos resultados oficiais, mas antes da disponibilização dos recursos finais do Segundo HAREM, fizemos uma última revisão de todas as entidades espúrias nas participações dos sistemas, de modo a garantir, por um lado, que não tínhamos problemas que pudessem prejudicar indevidamente os sistemas, e, por outro, a disponibilizar um recurso final o mais correcto possível. Essa revisão foi feita por quatro pessoas (cada qual revendo um quarto dos quase 10 mil casos espúrios). Os casos problemáticos foram discutidos por toda a equipa, e aqueles que classificámos como erro foram alterados na CD de modo a produzir o recurso que reputamos de final[10].

O processo de anotação e revisão da CD levou à identificação de 7836 entidades mencionadas, distribuídas pelas diversas categorias, de acordo com o gráfico da figura 1.3(b). Observa-se que a categoria mais frequente na CD é a categoria PESSOA, seguida das categorias LOCAL, TEMPO e ORGANIZACAO, com proporções de 27,11%, 18,15%, 15,21% e 14,02%, respectivamente. De referir que, no Primeiro HAREM, a categoria com maior representatividade na CD do Primeiro HAREM é a categoria LOCAL (24,6%), seguida, respectivamente, de PESSOA (21,0%) e ORGANIZACAO (17,8%), como indicado na figura 1.2. Tendo em consideração que a análise do TEMPO mudou radicalmente de uma edição para a outra, a proporção de EM reconhecidas nas duas edições de avaliação (apenas 9,0%, no Primeiro HAREM) não é naturalmente comparável.

No que diz respeito à vagueza, se tivermos apenas em conta a categoria, 535 entidades são vagas (6,38% dos casos). No entanto, observa-se que 633 EM da CD correspondem a EM vagas quanto a pelo menos um dos atributos CATEG, TIPO ou SUBTIPO (cerca de 8% dos casos). Ao nível da categoria, foram identificadas 52 classes de vagueza, encontrando-se na tabela 1.4 todas as classes que ocorrem mais de duas vezes[11] e na figura 1.4 a distribuição das categorias vagas. Na sua grande maioria (91,8% dos casos), a vagueza estabelece-se entre duas categorias. Os três casos mais frequentes foram: LOCAL|ORGANIZACAO (23,18% das entidades vagas), ORGANIZACAO|PESSOA (14,02%) e ABSTRACCAO|PESSOA (10,66%).

Relativamente às análises alternativas de identificação, observa-se que 372 sequências podem ser segmentadas de duas formas distintas, registando-se que apenas 11 sequências se encontram associadas a três possibilidades alternativas de segmentação. Das 7836 entidades existentes na CD, 1022 encontram-se dentro de um ALT (cerca de 13,8%).

Os casos acordados por maioria, e não por unanimidade (122 casos), foram devidamente identificados na CD, através da notação 2/3, que foi guardada no campo COMENT (um atributo opcional previsto na sintaxe de anotação das EM). A tabela 1.1 ilustra os casos de discordância registados. Nos casos em que não foi possível encontrar uma classificação

[9] E que, por essa razão, permitiu a marcação dos casos de decisão por maioria como 2/3.

[10] De referir, no entanto, que os resultados oficiais do Segundo HAREM se baseiam na CD que divulgámos no momento próprio, e, portanto, as mudanças referidas não influenciam a avaliação.

[11] Embora a tabela não mostre, verifica-se também vagueza entre 4 e 5 categorias.

Tabela 1.1: Distribuição de categorias, e discordância na anotação: D2/3 - Número de vezes em que a decisão de anotação não foi unânime; % - percentagem de entidades dessa categoria em que a decisão não foi unânime; DT: Número de vezes em que não houve acordo quanto à categoria

Categoria	Quant.	D2/3	%	DT
PESSOA	2036	13	0,64	2
LOCAL	1311	15	1,14	-
TEMPO	1189	35	2,94	-
ORGANIZACAO	961	16	1,66	2
OBRA	449	5	1,11	5
VALOR	353	-	-	-
COISA	308	5	1,62	1
ACONTECIMENTO	300	-	-	-
ABSTRACCAO	286	2	0,7	-
LOCAL|ORGANIZACAO	124	2	1,61	-
OUTRO	79	4	5,06	-
ORGANIZACAO|PESSOA	75	2	2,67	1
ABSTRACCAO|PESSOA	57	2	3,51	-
LOCAL|OBRA	33	1	3,03	-
ABSTRACCAO|ORGANIZACAO	31	4	12,9	-
EM	29	-	-	-
COISA|OBRA	24	1	4,17	-
LOCAL|PESSOA	14	-	-	-
COISA|LOCAL	14	7	50	-
OBRA|ORGANIZACAO	12	-	-	-
ACONTECIMENTO|LOCAL	11	-	-	1
ABSTRACCAO|LOCAL	11	-	-	-
ACONTECIMENTO|OUTRO	10	-	-	-
LOCAL|ORGANIZACAO|PESSOA	9	-	-	-
ACONTECIMENTO|OBRA	9	1	11,11	-
ABSTRACCAO|ACONTECIMENTO	9	-	-	-
COISA|ORGANIZACAO	8	-	-	-
ABSTRACCAO|COISA	6	-	-	-
ACONTECIMENTO|PESSOA	6	-	-	1
COISA|PESSOA	6	-	-	-
ABSTRACCAO|ACONTECIMENTO|ORGANIZACAO	6	-	-	-
ABSTRACCAO|ORGANIZACAO|PESSOA	4	2	50	-
COISA|OUTRO	4	1	25	-
TEMPO|VALOR	4	-	-	-
ACONTECIMENTO|ORGANIZACAO	3	-	-	-
LOCAL|OUTRO	3	-	-	1
ABSTRACCAO|OBRA	3	-	-	-
OBRA|PESSOA	3	-	-	-
OBRA|OUTRO	2	1	50	-
ABSTRACCAO|OUTRO	2	1	50	-
ABSTRACCAO|LOCAL|PESSOA	2	1	50	-
Outros casos de vagueza que ocorrem 2 vezes	16	-	-	-
Outros casos de vagueza que ocorrem 1 vez	14	-	-	-

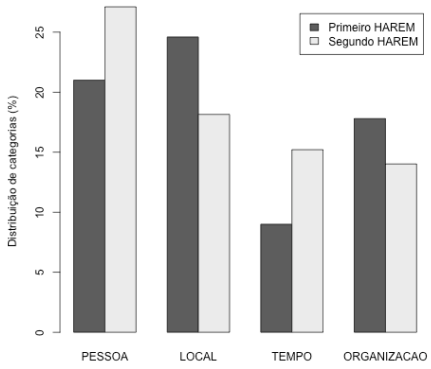
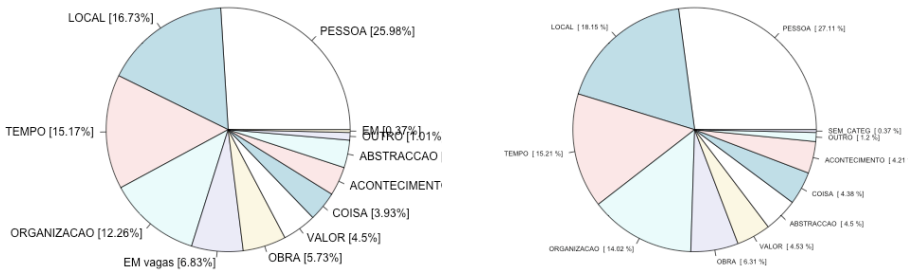


Figura 1.2: Distribuição das categorias mais frequentes na CD do HAREM em comparação com as mesmas categorias na CD do Primeiro HAREM



(a) A combinação de categorias de uma entidade vaga conta com uma única categoria, não contribuindo para cada categoria individualmente

(b) Para esta contabilização, cada categoria de uma entidade vaga contribuiu com $1/n$, sendo n número de categorias vagas dessa entidade

Figura 1.3: Distribuição de categorias na CD

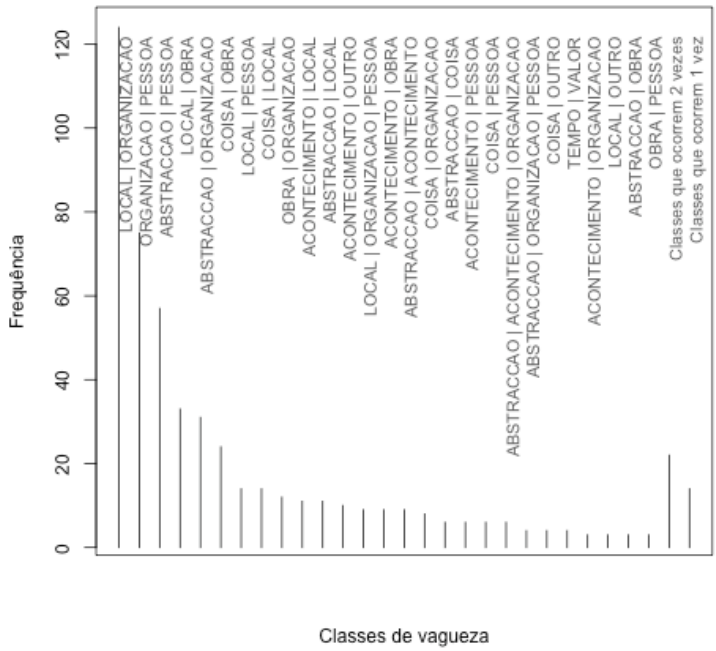


Figura 1.4: Distribuição das categorias vagas

consensual ou maioritária, optámos por omitir as EM em ambas as colecções (fazendo uso das etiquetas <OMITIDO> e </OMITIDO>), de modo a que as mesmas deixassem de ser alvo de avaliação (101 casos). De referir ainda que, em alguns casos, a discussão mostrou que as diferentes análises de interpretação em discordância eram possíveis, o que significa que todas elas passaram a ser representadas na CD, tirando partido dos mecanismos já anteriormente descritos para representação de EM vagas ou de EM que podem fazer parte de análises alternativas (em termos de segmentação).

1.5 Resultados da avaliação

Uma característica que consideramos inovadora e essencial no modelo de avaliação do HAREM diz respeito à flexibilidade oferecida aos sistemas em termos de participação e avaliação. Em concreto, os sistemas têm a possibilidade de escolher as categorias, tipos, subtipos ou outros atributos que pretendem etiquetar e ver avaliados, em função do interesse, pertinência ou adequação que essas anotações possam ter no âmbito de outras aplicações desenvolvidas ou a desenvolver por parte dos participantes, e que dependem directa ou indirectamente dessas informações. A cada conjunto diferente de categorias a que os participantes se propuseram ser avaliados (que aprofundaremos mais adiante), demos o nome de **cenário selectivo de participação**.

Tabela 1.2: Sistemas participantes no HAREM clássico e dados de participação

Sistema	N. corridas	Cenário	ALT
CaGE2	4	Selectivo 2	-
DobrEM	1	PESSOA	-
PorTexTO	4	TEMPO	-
Priberam	1	Total	-
R3M	2	Selectivo 3	-
REMBRANDT	3	Total	Sim
REMMA	3	Selectivo 4	Sim
SEI-Geo	4	Selectivo 5	-
SeRELeP	1	Total só Id	-
XIP-L2F/XEROX	4	Selectivo 6	-

Além disso, no Segundo HAREM implementámos outro tipo de cenários, os **cenários selectivos de avaliação**, que permitem a avaliação num subconjunto de categorias e tipos que não necessariamente o proposto pelo sistema.

A avaliação em cenários selectivos permite, entre outros aspectos, comparar o desempenho dos diferentes sistemas com base em cada uma das categorias que se propuseram reconhecer, assim como noutros conjuntos de categorias que possam fazer sentido.

Dito de outro modo, a avaliação levada a cabo no HAREM não se cinge a avaliar sistemas no âmbito de uma tarefa geral de REM, mas também, e fundamentalmente, a analisar mais detalhadamente o comportamento dos sistemas em tarefas mais específicas, previamente definidas pelos participantes, no âmbito da tarefa geral proposta pela organização. Deste modo, torna-se igualmente possível comparar os sistemas em cenários diferentes do cenário para o qual foram desenvolvidos.

Assim, todos os sistemas foram avaliados no cenário total e em cada um dos cenários selectivos de participação descritos na tabela 1.2. Além disso, todos os sistemas foram avaliados por categoria, o que corresponde a fazer a avaliação utilizando um cenário selectivo constituído apenas por cada uma dessas categorias. Em qualquer dos cenários referidos, os sistemas foram avaliados com avaliação estrita e relaxada de ALT (cf. capítulo 5).

O modelo e programas de avaliação do Segundo HAREM encontram-se descritos em detalhe no capítulo 5. Nesta secção, apenas apresentamos os sistemas participantes no HAREM clássico e os resultados de desempenho das corridas enviadas por esses sistemas.

1.5.1 Sistemas participantes

A tabela 1.2 mostra os dez sistemas participantes (que em conjunto enviaram 27 corridas[12]) e outros dados referentes à forma de participação. Por exemplo, se fez apenas identificação ou também classificação, e quais os cenários em que concorreu[13]. Como ilustra o quadro, os participantes envolveram-se de formas muito distintas na tarefa de reconhecimento de entidades mencionadas, uma situação que pode ter sido motivada pelo facto de o HAREM permitir a avaliação por cenários selectivos.

[12] Cada participante podia enviar no máximo quatro corridas.

[13] Ou seja, na terminologia técnica do HAREM, o cenário selectivo de participação de cada corrida (ver capítulo 5).

Tabela 1.3: Cenários de participação: I - apenas EM; C - classificação usando todos os atributos; CAT - apenas CATEG; CAT/T - sem SUBTIPO; F+H - LOCAL cujo TIPO seja FISICO e HUMANO

Cenário	PES	ORG	LOC	OBR	ACO	ABS	COI	TEM	VAL
PESSOA	I								
TEMPO								C	
Selectivo 2	CAT	CAT	F + H					CAT	
Selectivo 3	I	I	I	I	I	I	I		
Selectivo 4	C	C	CAT/T	C	C	C	C	CAT/T	C
Selectivo 5			F + H						
Selectivo 6	C	C	C	C	C			C	C
Total	C	C	C	C	C	C	C	C	C
Total só Id	I	I	I	I	I	I	I	I	I

1.5.2 Resultados

Apesar da diversidade da participação, a tarefa alvo em avaliação é o reconhecimento de entidades mencionadas. Como tal, começamos por analisar o desempenho dos sistemas no reconhecimento de todas as entidades existentes na CD, em termos de medida F, precisão e abrangência, no cenário total com avaliação estrita de ALT (figura 1.5[14]).

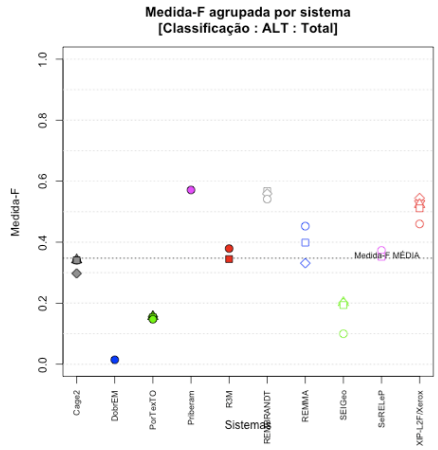
Note-se que não levámos a cabo, por enquanto, nenhum estudo estatístico dos resultados, como será referido no capítulo 6, e por isso a análise apresentada aqui será apenas uma primeira análise, bastante superficial.

O sistema da Priberam (cf. capítulo 9) foi o sistema com melhor medida F (0,5711), tendo ficado, no entanto, muito próximo do segundo melhor sistema, o REMBRANDT (cf. capítulo 11), cuja melhor corrida obteve 0,5674. Estes dois sistemas juntamente com o XIP-L2F/Xerox foram os únicos a obter valores de medida F superiores a 0,5.

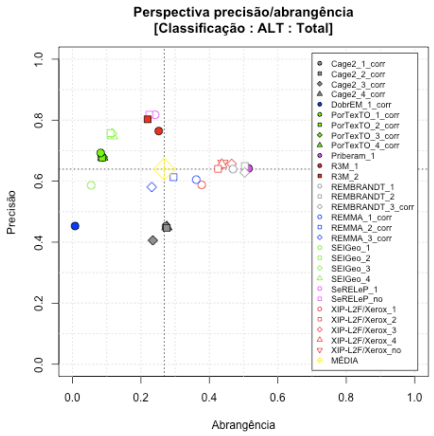
Relativamente às restantes corridas, apenas uma, enviada pelo REMMA (cf. capítulo 12) teve uma medida F superior a 0,4. De referir, no entanto, que isso tem naturalmente a ver com o facto de os cenários de participação dos restantes sistemas incluirem menos categorias (como é o caso do CaGE2 (cf. capítulo 7)) ou menos subtipos (caso do REMMA) e de alguns desses sistemas (caso do R3M (cf. capítulo 10) e do SeRELeP (cf. capítulo 14)) só terem feito identificação de entidades.

Uma explicação que se impõe em relação à interpretação dos resultados prende-se com justificar por que razão, na avaliação da classificação, sistemas que fizeram unicamente identificação têm valores de medida F próximos dos valores de sistemas que fizeram classificação. Compare-se, por exemplo, o desempenho dos sistemas R3M e SeRELeP, que fizeram apenas identificação, com o do sistema REMMA, que também fez classificação. Ao observarmos o gráfico que representa os resultados da avaliação da identificação (figura 1.5(c)), verificamos que os sistemas R3M e SeRELeP se encontram entre os melhores, o que não acontece com o sistema REMMA, que tem claramente um pior desempenho na identificação, o que também se reflecte na avaliação da classificação. Assim, podemos desde já afirmar que ainda estamos insatisfeitos com o peso atribuído à identificação, que acaba por penalizar indevidamente sistemas que fazem classificação – veja-se o capítulo 6 para mais discussão sobre este assunto.

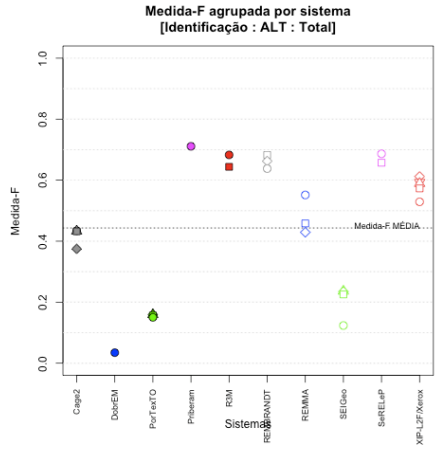
[14] Os valores correspondentes a esta figura e seguintes encontram-se no apêndice I (e no sítio do HAREM).



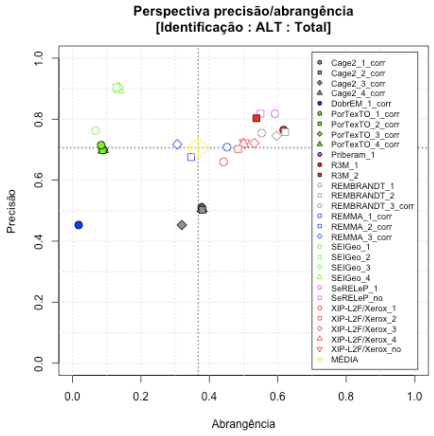
(a) Medida F na classificação



(b) Precisão e abrangência na classificação

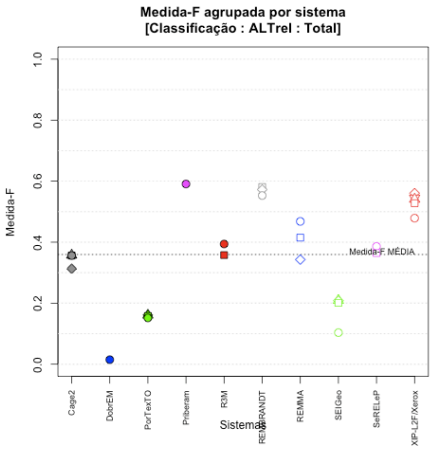


(c) Medida F na identificação

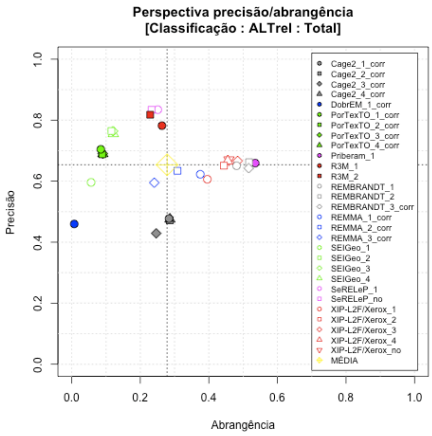


(d) Precisão e abrangência na identificação

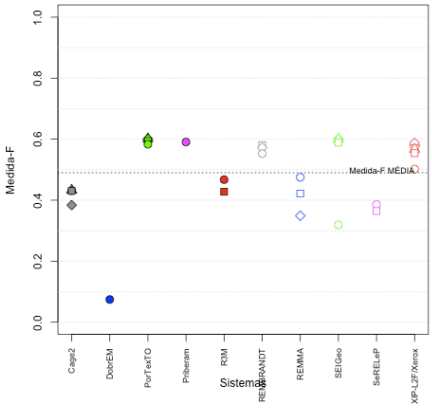
Figura 1.5: Avaliação no cenário total com avaliação estrita de `ALT`



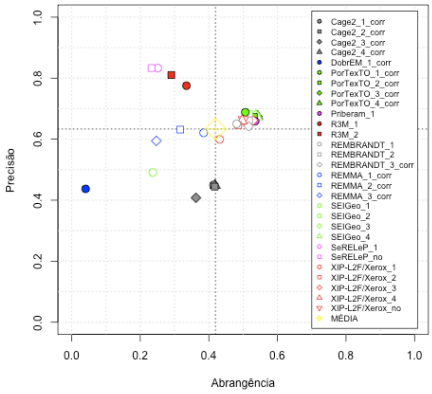
(a) Medida F no cenário total



(b) Precisão e abrangência no cenário total



(c) Medida F no cenário selectivo



(d) Precisão e abrangência no cenário selectivo

Figura 1.6: Classificação com avaliação relaxada de ALT

Relativamente ao desempenho na classificação com avaliação relaxada de ALT, vemos, na figura 1.6(a), que a medida F melhora ligeiramente para todos os sistemas. Em particular, os melhores sistemas, o sistema da Priberam e a melhor corrida do REMBRANDT, obtêm 0,5908 e 0,5808, respectivamente, aumentando um pouco mais a diferença de desempenho entre os dois sistemas. Esse aumento deve-se ao facto de apenas o sistema REMBRANDT ter utilizado ALT nas suas corridas.

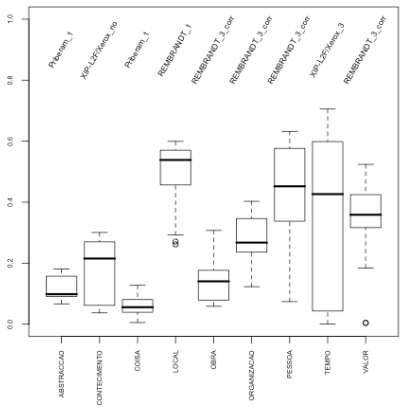
Analisemos agora o desempenho dos sistemas nos seus cenários selectivos, também tendo em conta a avaliação relaxada de ALT (já que apenas o sistema REMBRANDT e REMMA fizeram marcação de análises alternativas). Enquanto os gráficos anteriores ilustravam uma situação em que os sistemas estavam a ser todos avaliados no mesmo cenário, o cenário total, o que naturalmente desfavorece os sistemas que não participaram em todas as categorias, a figura 1.6(c) compara os sistemas tendo em consideração os respectivos cenários selectivos de participação.

Como seria de esperar, os sistemas que têm cenários de participação coincidentes com o cenário total, como seja o REMBRANDT e o da Priberam, não sofreram quaisquer alterações. Quanto aos restantes sistemas, vemos claramente melhores valores de medida F, sobretudo no caso de sistemas como o PorTexTO (cf. capítulo 8) e o SEI-Geo (cf. capítulo 13), que tentaram reconhecer apenas uma categoria, respectivamente TEMPO e LOCAL. Isto significa que, em relação ao objectivo que se propuseram alcançar, obtiveram um desempenho equiparável ao de outros sistemas que tinham objectivos mais ambiciosos. Ou, por outras palavras, estes sistemas podem ter reconhecido apenas uma categoria, mas, em termos relativos, foram tão bons a executar esse reconhecimento como os sistemas que tentaram reconhecer várias categorias.

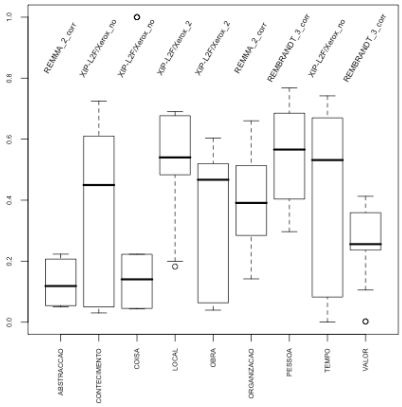
Com isto, não estamos a dizer que, no caso das categorias TEMPO e LOCAL, os sistemas PorTexTO e SEI-Geo, respectivamente, foram os melhores a reconhecer entidades com essa categoria. De facto, não o foram, como se pode ver na figura 1.7, que apresenta os melhores sistemas em cada uma das categorias. No caso da categoria TEMPO, o melhor sistema foi o XIP-L2F/Xerox (corrida 3), com 0,7054, que foi também o melhor sistema a reconhecer entidades ACONTECIMENTO; quanto à categoria LOCAL, o melhor sistema foi o sistema REMBRANDT (corrida 1), com 0,5993, que também foi, aliás, o melhor sistema, embora com uma corrida diferente, a reconhecer as restantes categorias, excepto ABSTRACCAO e COISA. Nestes últimos casos, o melhor sistema foi o da Priberam.

Se pensarmos que o melhor desempenho no reconhecimento de uma categoria traduz a facilidade no reconhecimento dessa categoria, podemos concluir que a entidade mais fácil de identificar é TEMPO, pois foi aquela onde foi obtido o melhor desempenho, imediatamente seguida de PESSOA e LOCAL. Nesta linha de interpretação, entidades como ABSTRACCAO e COISA seriam as mais difíceis de reconhecer, o que de certo modo faz algum sentido, na medida que se tratam de entidades mais abstractas ou, noutra perspectiva, mais abrangentes, e, por isso, mais difíceis de modelar.

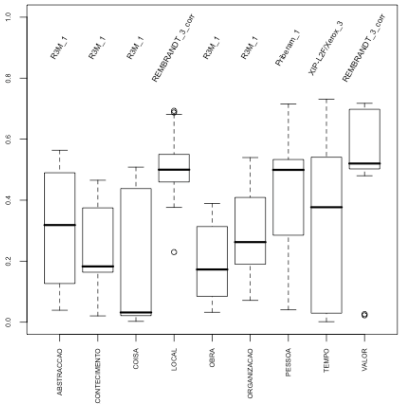
Parece-nos, no entanto, que a categoria onde houve de facto mais sucesso foi LOCAL. Algo que não é completamente surpreendente, uma vez que os autores de três sistemas participantes se dedicam a reconhecimento geográfico. Note-se, por exemplo, que a grande maioria das corridas obteve valores de medida F acima de 0,5, e que o pior sistema tem melhor desempenho na categoria LOCAL do que a maioria dos sistemas noutras categorias, sendo mesmo o melhor desempenho entre os piores das várias categorias. Esta situação contrasta com o desempenho na categoria TEMPO, onde se observa que a maioria dos sistemas está abaixo de 0,5 e onde se verifica uma maior dispersão dos valores, apesar



(a) Medida F



(b) Precisão



(c) Abrangência

Figura 1.7: Resumo de estatísticas da avaliação por categorias com avaliação estrita de ALT: máximo, mínimo, mediana, primeiro e terceiro quartis.

do melhor sistema ter obtido acima de 0,7.

Resta referir que estamos conscientes de que esta análise é bastante superficial e que, antes de tecer quaisquer conclusões definitivas sobre o que é fácil ou difícil, é também necessário fazer uma análise sistemática e aprofundada dos textos anotados, que passa, nomeadamente, pelo estudo das discordâncias de interpretação de certas entidades. Remetemos, pois, o leitor para o capítulo 6 para mais considerações sobre estas questões.