

Universidade de São Paulo - USP
Universidade Federal de São Carlos - UFSCar
Universidade Estadual Paulista - UNESP

Heurísticas de Poda de Sentenças para a Sumarização Automática de Textos UNL



Camilla Brandel Martins
Lucia Helena Machado Rino

NILC-TR-02-01

Março, 2002

Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional
NILC - ICMC-USP, Caixa Postal 668, 13560-970 São Carlos, SP, Brasil



Índice

RESUMO	4
1. INTRODUÇÃO.....	5
2. A UNIVERSAL NETWORKING LANGUAGE	7
3. HEURÍSTICAS DE SUMARIZAÇÃO.....	9
3.1. METODOLOGIA DE DESENVOLVIMENTO	9
3.2. TIPOS DE HEURÍSTICAS	10
3.2.1. <i>Heurísticas para RBs autônomas (Grupo A)</i>	10
3.2.1.1. Caracterização.....	10
3.2.1.2. Descrição	12
3.2.2. <i>Heurísticas para RBs inter-relacionadas (Grupo B)</i>	16
3.2.2.1. Caracterização.....	16
3.2.2.2. Descrição:	16
4. METODOLOGIA DE PODA	24
4.1. PROCESSO GERAL	24
4.1.1. <i>Procedimentos Especiais</i>	24
4.1.2. <i>Exemplo de Poda</i>	26
4.2. PRIORIZANDO HEURÍSTICAS PARA SELEÇÃO.....	27
5 DISCUSSÃO FINAL	29
REFERÊNCIAS BIBLIOGRÁFICAS.....	30
ANEXO 1 – Relação completa das Heurísticas de Sumarização – Versão 1.3.....	32
ANEXO 2 – Relation Labels Considerados Relevantes para a Sumarização Automática de Textos UNL	35
ANEXO 3 – Valor de Prioridade de cada Heurística	37

RESUMO

Este relatório apresenta um conjunto de heurísticas para a sumarização automática de textos codificados na *Universal Networking Language*, UNL. Esta língua é utilizada para representar o significado de textos sentença por sentença. Estas, por sua vez, são compostas por conjuntos de relações semânticas, binárias, que se propõem a representar todas as informações contidas na sentença correspondente escrita em língua natural. Com base em características específicas da UNL, as heurísticas de sumarização procuram e excluem, da sentença, relações binárias que correspondam a informações irrelevantes ou menos proeminentes.

O conjunto de heurísticas compõe o protótipo UNLSumm que, acoplado ao sistema de decodificação UNL-Português do Brasil, permitirá que se sumarie um código UNL, gerando textos condensados e, caso se deseje, gerando também as correspondentes traduções para o Português do Brasil.

1. Introdução

Descrevemos, neste relatório, um conjunto de heurísticas para a sumarização automática de textos UNL, que constitui o engenho de inferência do UNLSumm (Martins, 2000). Textos UNL são textos codificados na interlíngua Universal Networking Language (Uchida, 2000), a qual permite representar o significado de textos (ou permite especificar uma representação semântica de asserções de uma língua natural), sentença por sentença, pretensamente de forma não ambígua. A UNL é parte do Projeto UNL¹, de comunicação interlingual. De um modo geral, a comunicação no Sistema UNL consiste em codificar em UNL um texto em língua natural e decodificar a representação UNL em outra língua natural. Esses processos podem ser realizados, respectivamente, pelas ferramentas EnCo e DeCo (Uchida, 1997), personalizadas para as LNs envolvidas.

Aproveitando o ambiente de decodificação personalizado para o Português do Brasil² (PB)³, a proposta é adicionar, na forma de plug-in, o UNLSumm (*UNL Summarizer*), protótipo que implementa as heurísticas de sumarização, como ilustra a Figura 1 (caixas em azul representam o sistema de decodificação UNL-PB e caixas em amarelo representam os recursos específicos do UNLSumm).

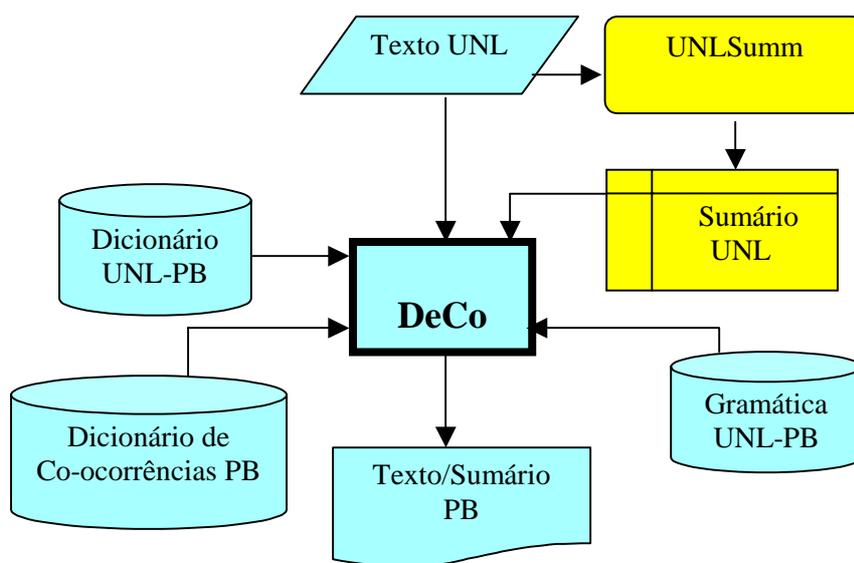


Figura 1: Ambiente de decodificação UNL-Português acrescido do plug-in UNLSumm

O UNLSumm realiza a sumarização automática (SA) pela poda de estruturas UNL, que se dá pela exclusão de conteúdos considerados irrelevantes em textos UNL, com base em características específicas daquelas estruturas. Como resultado do processo de SA, temos um texto UNL condensado, chamado aqui de Sumário UNL, que poderá, então, ser decodificado para qualquer língua natural de interesse, por exemplo, PB (vide Figura 1).

O conjunto de heurísticas, já em fase de testes no protótipo UNLSumm, é formado hoje por 84 heurísticas, divididas em dois grandes grupos, conforme particularidades que

¹Uma descrição completa deste projeto e seus módulos pode ser encontrada em <http://www.unl.ias.unu.edu>

² A customização da gramática e dicionários foi realizada pela equipe do Projeto UNL Brasil, do Núcleo Interinstitucional de Linguística Computacional (NILC).

³ Os termos Português, Português do Brasil e a sigla PB serão utilizados com o mesmo significado neste trabalho.

veremos adiante. No próximo capítulo apresentamos a interlíngua UNL, enfatizando particularidades importantes para a elaboração das heurísticas. No capítulo 3 apresentamos as heurísticas de sumarização. No capítulo 4 apresentamos a metodologia geral de poda e alguns critérios utilizados para a seleção e aplicação das heurísticas descritas. Finalmente, no capítulo 5, discutimos alguns problemas previstos para a aplicação do conjunto de heurísticas e trabalhos futuros.

2. A Universal Networking Language

Como dissemos, a UNL (Uchida, 2000) representa textos sentença por sentença. Estas, por sua vez, são representadas por meio de três componentes principais: conceitos (por exemplo, ‘Pedro’ e ‘correr’), relações entre conceitos (tais como agente, local, objeto) e atributos dos conceitos (por exemplo, tempo passado, conceito definido). Considera-se que os três componentes são suficientes para carregar o significado de uma sentença e são, respectivamente, denominados como: Universal Words (UWs), Relation Labels (RLs) e Attribute Labels (ALs).

As UWs são representadas por termos em inglês. No entanto, o vocabulário UNL, apesar de remeter à língua inglesa, não é equivalente a ela, sendo muito mais abrangente, de forma a representar conceitos de outras línguas naturais inexistentes na língua inglesa (Martins et al, 2000). As UWs podem ser simples, como “book”, “run”, “Pedro” ou complexas, incluindo informações adicionais de diversos tipos, por exemplo informações morfosintáticas (run.@past). Tais informações compõem o grupo de ALs da UNL e são sempre precedidas pelo símbolo ‘@’.

Os RLs assinalam relações semânticas e são expressos por mnemônicos de 3 letras (por exemplo, ‘agt’ para relações de agente, obj’ para objeto)⁴. Existem atualmente 41 RLs cobrindo diferentes funções. Basicamente, um RL é utilizado para relacionar semanticamente dois conceitos. Por exemplo, na sentença “Pedro corre”, o RL ‘agt’ é utilizado para relacionar o agente à ação, gerando assim a seguinte relação binária (RB):

agt(run, Pedro)

Uma sentença UNL, então, será formada por um conjunto de relações binárias da forma RL(UW1,UW2). Os exemplos 1 e 2 mostram, respectivamente, possíveis codificações UNL para as sentenças “Uruguai took the lead versus Argentina through a penalty” [S1] e “Uruguay took the lead versus Argentina through a goal scored by Pablo Dorado“ [S2].

agt(take.@past,Uruguay)	(1)
obj(take.@past,lead.@def)	(2)
ptn(take.@past,Argentina)	(3)
met(take.@past,penalty.@indef)	(4)

Exemplo 1: Representação UNL para S1

Como é possível verificar pelos exemplos, as RBs estão inter-relacionadas pelo compartilhamento de algumas UWs. Por exemplo, no Exemplo 2, as primeiras 4 RBs estão ligadas pela UW “take.@past”, enquanto as RBs 5 e 6 estão relacionadas pela UW “score.@past”. Esse relacionamento é o que irá garantir, na decodificação, a coesão da sentença em LN e, conseqüentemente, sua coerência. Considerando o subgrupo formado pelas RBs 1-4 e o subgrupo formado pelas RBs 5-6, verificamos que a RB 4 faz a ligação entre esses subgrupos, como podemos ver mais claramente na representação gráfica do Exemplo 2 (Figura 2). Ou seja, a remoção dessa única RB “quebraria” a sentença em duas partes, ou seja, restariam dois subgrafos sem conexão e, assim, o DeCo não conseguiria completar a decodificação desse código UNL.

⁴ Uma descrição de cada RL citado pode ser encontrada no Anexo 2.

agt(take.@past,Uruguay)	(1)
obj(take.@past,lead.@def)	(2)
ptn(take.@past,Argentina)	(3)
met(take.@past,goal)	(4)
obj(score.@past,goal)	(5)
agt(score.@past,'Pablo Dorado')	(6)

Exemplo 2: Representação UNL para S2

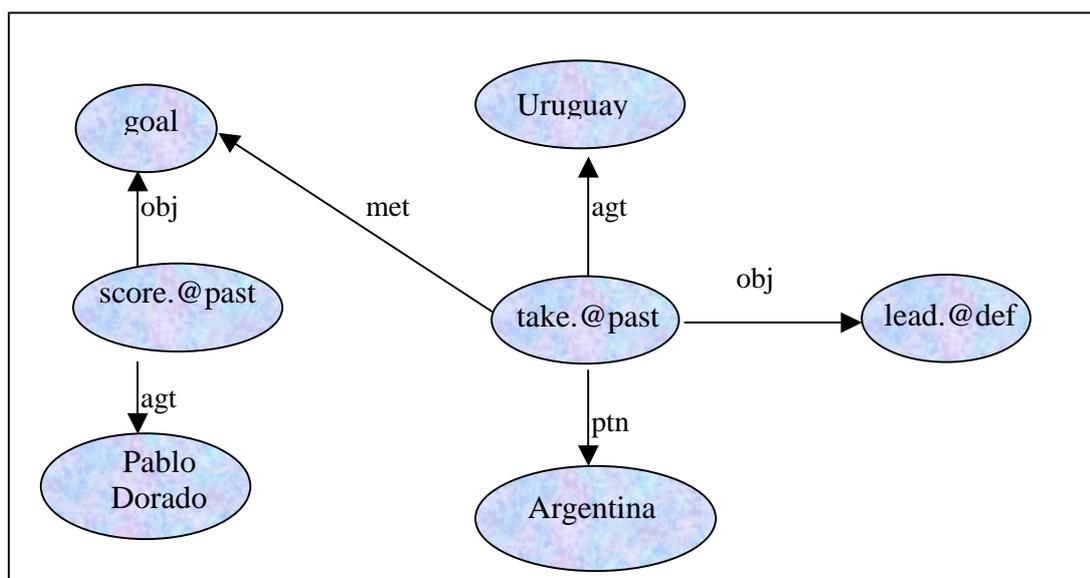


Figura 2: Representação gráfica de S2

Em alguns casos, o significado de uma sentença pode ser melhor representado ao se fazer referência a um conjunto de RBs agrupadas. A UNL permite essa referência por meio de um número determinado ‘escopo’. Por exemplo, a sentença S2 poderia ser representada, utilizando escopo, como no Exemplo 3, onde a RB 4 faz referência ao escopo ‘:01’, formado pelas RBs 5 e 6. Ou seja, isso significa que o método está representado por todo aquele conjunto de relações binárias. Como veremos na próxima seção, essa característica também é importante para o desenvolvimento das heurísticas de poda.

agt(take.@past,Uruguay)	(1)
obj(take.@past,lead.@def)	(2)
ptn(take.@past,Argentina)	(3)
met(take.@past.@pred,:01)	(4)
obj:01(score.@past,goal)	(5)
agt:01(score.@past,'Pablo Dorado')	(6)

Exemplo 3: Representação UNL para S2 usando escopo

3. Heurísticas de Sumarização

As heurísticas de sumarização de textos UNL têm por objetivo identificar informações irrelevantes ou menos proeminentes em uma sentença UNL e realizar sua exclusão, realizando, assim, a ‘poda’ da sentença. Por meio da poda sucessiva de partes das sentenças de um texto, é realizada, então, a sumarização do texto como um todo. Outra função das heurísticas é verificar condições que garantam que a exclusão das informações selecionadas não irá prejudicar a coesão e a coerência da sentença podada.

Para identificar as informações irrelevantes de uma sentença, as heurísticas se baseiam nos RLs, sob a hipótese de que, uma vez que eles representam relações semânticas entre conceitos, eles podem indicar trechos não essenciais da sentença. Assim, as heurísticas são definidas pelo rótulo (RL) das possíveis RBs a serem excluídas.

3.1. Metodologia de desenvolvimento

Para identificar quais RLs indicam uma RB supérflua, diversas atividades foram realizadas, antes de desenvolver propriamente as heurísticas. Primeiro foi realizada a retroprojeção dos RLs em suas funções sintáticas, de modo a detectar componentes que, por sua classificação sintática, pudessem ser considerados irrelevantes ou menos proeminentes em uma sentença.

O mapeamento dos RLs nas manifestações morfossintáticas foi realizado: a) para o Inglês, utilizando os Corpora *Booklet* e *UN Charter*, contendo textos UNL e suas correspondentes versões em inglês e b) para o Português, por Sossolote et al. (1997), que utilizaram textos UNL e traduções para o PB. Verificamos, também, que as manifestações da UNL no Português e no Inglês possuem muitas semelhanças e, portanto, nenhuma distinção será feita para textos UNL codificados a partir de textos em Inglês ou Português⁵.

A segunda atividade foi coletar evidências literárias⁶ que indicassem componentes morfossintáticos possivelmente irrelevantes ou menos proeminentes para, em seguida, comparar com as retroprojeções da UNL em língua natural. Por exemplo, por meio das retroprojeções, verificamos que atributos de um conceito (que pode ser uma pessoa ou objeto, p. ex.) são representados em UNL por meio do RL ‘aoj’, geralmente correspondendo a adjuntos na superfície do texto, como observado por Sossolote et al. (1997). Como conceitos que aparecem na posição de adjuntos podem ser considerados pouco importantes ou menos importantes do que os outros (conforme Boguraev and Kennedy, 1997) em uma sentença, similarmente uma relação binária com o RL ‘aoj’ também pode ser considerada pouco relevante em uma sentença UNL. Nesse caso, criamos uma heurística que exclua de uma sentença UNL RBs que apresentem esse RL.

Além de evidências literárias, análises subjetivas de corpora ajudaram a formar um elenco de RLs chave para a elaboração das heurísticas. Essas análises foram realizadas sobre dois corpora de textos UNL, que chamamos de corpora de treinamento: *Corpus UN Charter* e *Corpus Booklet*, ambos codificados a partir de textos-fonte em Inglês. O primeiro é composto por codificações UNL de 12 capítulos do estatuto da ONU, totalizando 255 sentenças UNL, ou 3182 RBs, no gênero legal e domínio jurídico. O *Corpus Booklet* é composto por 101 sentenças UNL, ou 1230 BRs, sendo de gênero narrativo e domínio aberto. Ambos foram codificados por equipes de diferentes países

⁵ Para os RLs considerados neste artigo, as manifestações morfossintáticas são apresentadas juntamente com as heurísticas, na seção 3.2

⁶As correspondências entre as heurísticas e a literatura utilizada serão apresentadas na seção 3.2.

pertencentes ao Projeto UNL, formadas por falantes nativos das línguas naturais consideradas. Para a análise dos corpora, consideramos também a tradução em PB dos textos UNL.

A análise subjetiva dos corpora de textos UNL, juntamente com a análise da função morfossintática dos RLs e diversas indicações de informações que podem ser consideradas supérfluas encontradas na literatura, levaram à formação de um elenco de 84 heurísticas, divididas em dois grupos principais, Grupo A e Grupo B, por particularidades que veremos na seção seguinte. Em cada grupo, elas são, por sua vez, agrupadas por RL. Para uma melhor referência, as heurísticas são nomeadas HG.S.N, onde G corresponde ao seu grupo principal (A ou B); S indica o subgrupo e N é o número da heurística dentro do subgrupo. Por exemplo, as heurísticas denominadas por HA.1 são relativas ao RL ‘met’ no Grupo A (ver quadro no Anexo 1). Portanto, a primeira heurística desse subgrupo é a HA.1.1, a segunda HA.1.2 e assim por diante.

3.2. Tipos de Heurísticas

O Grupo A é composto por heurísticas que excluem apenas uma relação binária e, logo, temos uma poda simples. O Grupo B inclui heurísticas mais complexas, que simultaneamente excluem um grupo de RBs inter-relacionadas, ou seja, temos uma poda encadeada.

Em ambos os casos, a aplicação de uma heurística ocorre apenas após verificar algumas condições, de forma a garantir a textualidade da sentença (ou texto) UNL resultante, i.e., garantir a sua coesão e coerência. Embora o termo ‘textualidade’ seja mais comumente utilizado para textos multi-sentenciais, utilizamos a mesma terminologia para textos mono-sentenciais, i.e., para sentenças únicas, uma vez que estas são o foco do mecanismo de poda. Além disso, embora a noção de textualidade, para alguns autores, inclua critérios centrados no usuário, como intencionalidade e informatividade (Beaugrande e Dressler, 1981) ou fatores de contextualização e de conexão de ações (Marcuschi, 1983), as verificações realizadas pelas heurísticas dizem respeito, principalmente, à coesão e coerência.

Assim, por exemplo, ao excluir o verbo da sentença “Pedro comeu a uva”, gerando a sentença sumarizada “Pedro a uva”, pode-se considerar que esta não possui textualidade, pois não está clara a conexão (ou coesão) entre Pedro e a uva. Em consequência da perda de coesão perde-se também a coerência (cf. Koch, 1995). Assim, a sentença perdeu a coerência e, conseqüentemente, não transmite nenhum significado. Como veremos adiante, as heurísticas previnem que isso ocorra, verificando o inter-relacionamento entre UWs.

3.2.1. Heurísticas para RBs autônomas (Grupo A)

3.2.1.1. Caracterização

As heurísticas do Grupo A focalizam somente uma relação binária, i.e., identificam aquela RB que pode ser individualmente excluída da sentença UNL. Para o Exemplo 1 (seção 2), i.e., correspondente à sentença “Uruguay took the lead versus Argentina through a penalty” [S1], se aplicarmos a heurística HA.1.1 (Figura 3), a RB 4 será excluída, resultando no Sumário UNL ilustrado no Exemplo 4, correspondente à sentença “Uruguay took the lead versus Argentina”, que é coerente e condensada em relação a S1. A coerência é garantida pela verificação da segunda UW da relação binária (expressa na heurística por

‘b’). Ou seja, se ‘b’ não ocorre em nenhuma outra relação binária, então a heurística pode ser aplicada, garantindo a manutenção da coerência, i.e., da textualidade.

agt(take.@past,Uruguay)	(1)
obj(take.@past,lead.@def)	(2)
ptn(take.@past,Argentina)	(3)

Exemplo 4: Sumário UNL da sentença S1

Excluir met(a,b) da sentença S Se $b \notin$ outras RBs em S e não existe obj(a,x)
--

Figura 3: Heurística HA.1.1

Na descrição que segue, identificamos as seguintes informações para cada heurística:

- Função: aponta a função da heurística em foco;
- Formato da RB: indica a sintaxe da relação binária em foco, detalhando o campo semântico de cada UW envolvida;
- UW relevante: indica qual das UWs é determinante para a aplicação da heurística. Ser determinante significa, aqui, que a UW carrega o foco da relação binária, noção correspondente ao foco de uma sentença, de Vallduvi (1993), i.e., a parte informativa, dominante, em oposição ao *ground* (neste caso, a outra UW), a parte conhecida, esperada ou não-informativa.
- Hipótese: Justificativa da não relevância da RB em foco. Salvo quando especificado, a hipótese é subjetiva.
- Função sintática: indica as possíveis manifestações sintáticas, no Português do Brasil, das relações binárias rotuladas pelo RL em foco, conforme observamos nos corpora de treinamento e em Sossolote et al. (1997). Indicaremos aqui unicamente a função sintática da UW relevante, de modo a deixar claro que o componente em língua natural correspondente à UW a ser excluída da sentença não é essencial para garantir a estrutura textual da sentença.
- Função morfossintática: indica a função morfossintática no Português do Brasil das relações binárias rotuladas por um determinado RL, conforme observamos nos corpora de treinamento e em Sossolote et al. (1997). Novamente, indicaremos a função morfossintática da UW relevante, evidenciando o item léxico que será excluído da sentença.
- Exemplo típico em LN: ilustração de um caso típico⁷, cujo trecho a ser excluído é indicado em negrito.
- Exemplo típico em UNL: ilustração, em UNL, correspondente ao exemplo típico em LN, ressaltando-se em negrito a RB que deverá ser excluída.

A função morfossintática, neste trabalho, indicará tanto a função morfológica de UWs simples, quanto a função morfossintática de UWs compostas ou UWs simples que correspondam, em língua natural, a mais de um termo. Por exemplo, para a sentença “O povo começou a construir uma torre na cidade”, ‘na cidade’ é morfossintaticamente

⁷ Os exemplos foram retirados dos corpora de textos UNL utilizados e traduzidos para o português, para melhor visualização.

classificado como locução adverbial, enquanto seus componentes são classificados como preposição⁸ + substantivo. E, nesta sentença, a função sintática da locução adverbial ‘na cidade’ é de adjunto adverbial.

Como grande parte das construções sintáticas e morfossintáticas do Inglês é comum ao PB, em relação a hipóteses levantadas como determinantes para a exclusão baseadas em funções sintáticas, as baseadas na língua inglesa ou no PB serão utilizadas sem distinção.

3.2.1.2. Descrição

Subgrupo HA.1. - met

- Função: excluir relação binária rotulada por met
- Formato da RB: met(UW1, UW2), onde UW1 = ação, UW2 = método
- UW relevante: UW2
- Hipótese – O modo⁹ pelo qual uma ação é feita é uma informação complementar (Robin, 1994). Ainda, em textos científicos, é comum que o resumo não contenha material de seções nomeadas Métodos (Buxton & Meadows, 1978).
- Função sintática: adjunto adverbial
- Função morfossintática: verbo, advérbio, núcleo do objeto ou núcleo de expressão adverbial
- Exemplo típico em LN: “Uruguai tomou a liderança contra a Argentina **por meio de um pênalti**”.
- Exemplo típico em UNL:

```
agt(take.@past,Uruguay)
obj(take.@past,lead.@def)
ptn(take.@past,Argentina)
met(take.@past.@pred,penalty.@indef)
```

Subgrupo HA.2. - man

- Função: excluir relação binária rotulada por man
- Formato da RB: man(UW1, UW2), onde UW1 = ação, UW2 = modo
- UW relevante: UW2
- Hipótese – A ‘maneira’ particular com que uma ação é feita é uma informação complementar (Hovy, 1988)¹⁰.
- Função sintática: adjunto adverbial
- Função morfossintática: verbo, advérbio ou locução adverbial.
- Exemplo típico em LN: Elas não conseguiriam mais se comunicar **umas com as outras**.
- Exemplo típico em UNL:

```
man(communicate, no longer)
agt(communicate,they)
man(communicate, each other)
```

⁸ As descrições de funções morfossintáticas se referirão explicitamente somente às palavras de classe aberta, i.e., aos substantivos, adjetivos, advérbios e verbos.

⁹ Deve-se ressaltar que, embora tanto ‘met’ quanto ‘man’ descrevam o modo como uma ação é realizada, ‘met’ descreve os passos componentes, procedimentos ou instrumentos pelos quais a ação é realizada, enquanto ‘man’ apenas descreve o evento como um todo, evidenciando características particulares (Uchida et al., 1999).

¹⁰ Hovy considera que informações de modo são complementares e podem ser supérfluas, para a SA, como, por exemplo, em “Quietly, Jim told Sue...” (Hovy, 1988, p. 123). ‘Quietly’, neste caso, expressa o modo como Jim disse algo a Sue e, dependendo do objetivo comunicativo, pode ser suprimido.

Subgrupo HA.3. – plc, lpl, ppl¹¹

- Função: excluir relação binária rotulada por plc, lpl ou ppl
- Formato da RB: RL(UW1,UW2), onde UW1 = ação e UW2=local (físico ou lógico)
- UW relevante: UW2
- Hipótese - Hovy (1988) e Robin (1994) sugerem que o local onde uma ação ocorre é uma informação complementar, que pode ou não ser apresentada, dependendo do objetivo comunicativo e do grau de detalhe que se pretende apresentar.
- Função sintática: adjunto adverbial
- Função morfossintática: substantivo, advérbio ou locução adverbial
- Exemplo típico em LN: O povo começou a construir uma torre **na cidade**.
- Exemplo típico em UNL:

```
ppl(begin, city)
agt(begin, people)
obj(begin, build)
obj(build, tower)
```

Subgrupo HA.4 - aoj

- Função: excluir relação binária rotulada por aoj
- Formato da RB: aoj(UW1,UW2), onde UW1= atributo ou estado e UW2=entidade
- UW relevante: UW1
- Hipótese: Oka e Ueda (2000) atribuem pouca relevância aos modificadores de conceitos (em UNL, representados por ‘aoj’ e ‘mod’, como observados nos corpora UNL analisados). Hutchins (1987), ao exemplificar sua *Deletion rule*, sumariza a sentença “Peter saw a blue ball.” como “Peter saw a ball.”. Em UNL, isso equivale a excluir a RB rotulada por ‘aoj’ ou ‘mod’.
- Função sintática: adjunto adnominal
- Função morfossintática: adjetivo, verbo principal ou verbo de ligação em um predicado nominal¹²
- Exemplo típico em LN: O povo começou a construir uma torre **imensa**.
- Exemplo típico em UNL:

```
agt(begin, people)
obj(begin, build)
obj(build, tower)
aoj(huge, tower)
```

Subgrupo HA.5 – tim

- Função: excluir relação binária rotulada por tim
- Formato da RB: tim(UW1,UW2), onde UW1=ação e UW2=tempo em que a ação ocorre
- UW relevante: UW2
- Hipótese – Hovy (1988) e Robin (1994) sugerem que a informação temporal é uma informação complementar, que pode ou não ser apresentada, dependendo do objetivo comunicativo e do grau de detalhe que se pretende apresentar.
- Função sintática: adjunto adverbial

¹¹ Os RLs ‘lpl’ e ‘ppl’, respectivamente equivalentes a *logical place* e *physical place* foram excluídos da especificação UNL, tendo sido substituídos pelo RL ‘plc’ ou *place*. Como trabalhamos com códigos UNL produzidos a partir de diferentes versões da especificação UNL, decidimos manter todos esses rótulos, considerando-os neste único grupo.

¹² Em Sossolote et al. (1997), um verbo de ligação não aparece como categoria morfossintática de ‘aoj’, mas sim de ‘soj’. No entanto, ‘soj’ não existe na especificação UNL atual, tendo sido subsumida por ‘aoj’.

- Função morfossintática: advérbio ou locução adverbial
- Exemplo típico em LN: Existem 3000 línguas diferentes faladas no mundo **hoje**.
- Exemplo típico em UNL:

```
obj(exist, language)
qua(language, 3000)
aoj(different, language)
obj(speak, language)
ppl(speak, earth.@def)
tim(speak, today)
```

Subgrupo HA.6 - pos

- Função: excluir relação binária rotulada por pos
- Formato da RB: pos(UW1,UW2), onde UW1=objeto ou coisa possuída e UW2=possuidor de UW1
- UW relevante: UW2
- Hipótese : Se o tópico da sentença estiver no objeto e não no possuidor, então essa RB pode ser excluída.
- Função sintática: complemento nominal
- Função morfossintática: substantivo, sintagma preposicional
- Exemplo típico em LN: Diferenças são uma barreira para o fluxo de informação na **nossa** sociedade.
- Exemplo típico em UNL:

```
aoj(barrier, difference.@pl)
ben(barrier, flow.@def)
mod(flow.@def, information)
lpl(flow.@def, society)
pos(society, we)
```

Subgrupo HA.7 - mod

- Função: excluir relação binária rotulada por mod
- Formato da RB: mod(UW1,UW2), onde UW1=ação, objeto ou coisa e UW2=modificador de UW1
- UW relevante: UW2
- Hipótese: Oka e Ueda (2000) atribuem pouca relevância aos modificadores de conceitos (em UNL, representados por ‘aoj’ e ‘mod’).
- Função sintática: adjunto adnominal, adjunto adverbial
- Função morfossintática: diversas classes de palavras
- Exemplo típico em LN: O **único** problema que resta é como superar a barreira lingüística.
- Exemplo típico em UNL:

```
mod(problem.@def, only)
obj(remain, problem.@def)
aoj(overcome, problem.@def)
man(overcome, how)
obj(overcome, barrier.@def)
mod(barrier.@def, language)
```

Subgrupo HA.8 - nam

- Função: excluir relação binária rotulada por nam
- Formato da RB: nam(UW1,UW2), onde UW1=objeto e UW2=nome específico (próprio)
- UW relevante: UW2

- Hipótese: Se o foco da sentença está em outro conceito, o nome do objeto nomeado pela RL 'nam' é de pouca relevância. Ainda, Aone et al. (1999) observaram que nomes próprios geralmente não indicam o tópico de textos jornalísticos, o que também pode indicar uma baixa relevância.
- Função sintática: complemento nominal
- Função morfossintática: substantivo
- Exemplo típico em LN: Os resultados das pesquisas conduzidas pela companhia de pesquisa Americana **Jupiter** revelam que...
- Exemplo típico em UNL:

```
mod(result.@pl.@def, survey:01.@pl)
obj(conduct, survey:01.@pl)
mod(company.@pl, survey:02)
mod(company, US:01)
nam(company.@pl, Jupiter)
agt(conduct, company.pl)
cau(reveal, result.@pl.@def)
```

Subgrupo HA.9 - src

- Função: excluir relação binária rotulada por src
- Formato da RB: src(UW1,UW2), onde UW1=ação, objeto ou coisa e UW2=fonte ou estado inicial
- UW relevante: UW2
- Hipótese: Boguraev and Kennedy (1997) consideram que informações na posição de adjuntos são menos proeminentes. Como informações representadas por meio do RL 'src' geralmente correspondem a adjuntos, essa RB pode ser excluída. Ainda, algumas informações que poderiam ser codificadas utilizando 'src', por exemplo, '*Based on these data*' e '*From these results*' são consideradas não relevantes (Pollock and Zamora, 1975).
- Função sintática: adjunto adverbial, aposto no sujeito
- Função morfossintática: substantivo
- Exemplo típico em LN: o professor Jorge, **da UFSCar**, está envolvido na pesquisa.
- Exemplo típico em UNL:

```
mod(Jorge, professor)
src(Jorge,UFSCar)
agt(envolve,Jorge)
scn(envolve,research)
```

Subgrupo HA.10 - qua

- Função: excluir relação binária rotulada por qua
- Formato da RB: qua(UW1,UW2), onde UW1=objeto ou ação e UW2=quantidade
- UW relevante: UW2
- Hipótese: Boguraev and Kennedy (1997) consideram que informações na posição de adjuntos são menos proeminentes. Portanto, como informações representadas por meio do RL 'qua' geralmente correspondem a adjuntos, essa RB pode ser excluída.
- Função sintática: adjunto adnominal
- Função morfossintática: numeral
- Exemplo típico em LN: Ocorreu nos últimos **dois** anos.
- Exemplo típico em UNL:

```
tim(occur,year)
mod(year,last)
qua(year,two)
```

Subgrupo HA.11 - cnt

- Função: excluir relação binária rotulada por cnt
- Formato da RB: cnt(UW1,UW2), onde UW1=objeto ou ação e UW2=conceito equivalente a UW1
- UW relevante: UW2
- Hipótese: Um conceito equivalente, ou paráfrase, é um material redundante que pode ser excluído em um sumário (Brown and Day, 1983).
- Função sintática: adjunto adnominal, aposto
- Função morfossintática: substantivo
- Exemplo típico em LN: A Universal Networking Language (UNL) é uma linguagem eletrônica para computadores.
- Exemplo típico em UNL:

```
cnt(Universal Networking Language, UNL.@parenthesis)
aoj(language, Universal Networking Language)
mod(language, electronic)
pur(language, computer.@pl)
```

3.2.2. Heurísticas para RBs inter-relacionadas (Grupo B)

3.2.2.1. Caracterização

As heurísticas do Grupo B são um pouco mais complexas do que as do Grupo A, uma vez que simultaneamente excluem um grupo de relações binárias. O ponto principal é encontrar o RL que “aciona” o processo de poda, indicando as RBs subsequentes (que também serão excluídas). Por exemplo, no Exemplo 2 (transcrito abaixo) não seria possível aplicar a heurística HA.1.1, pois a segunda UW (no caso, ‘goal’) está presente em outra RB. Como podemos verificar por esse exemplo, há casos em que a poda encadeada é essencial. Como veremos adiante, HB.1.3 (Anexo 1) poderia ser aplicada neste caso.

agt(take.@past,Uruguay)	(1)
obj(take.@past,lead.@def)	(2)
ptn(take.@past,Argentina)	(3)
met(take.@past,goal)	(4)
obj(score.@past,goal)	(5)
agt(score.@past,'Pablo Dorado')	(6)

Exemplo 2: Representação UNL para S2

A seguir, são descritos todos os subgrupos de heurísticas do Grupo B, seguindo o mesmo formato da seção anterior. Adicionalmente, a UW relevante, neste caso, é também a que aciona o processo de poda encadeada, levando à exclusão de outras RBs. É importante notar que, como tais heurísticas são sempre disparadas por um único RL, a maioria delas terá sua descrição coincidente com as do Grupo A.

3.2.2.2. Descrição:

Subgrupo HB.1 - met

- Função: excluir relação binária rotulada por met + RBs relacionadas a UW2
- Formato da RB: met(UW1, UW2), onde UW1 = ação, UW2 = método
- UW relevante: UW2

- Hipótese – O modo pelo qual uma ação é feita é uma informação complementar (Robin, 1994). Ainda, em textos científicos, é comum que o resumo não contenha material de seções nomeadas Métodos (Buxton & Meadows, 1978).
- Função sintática: adjunto adverbial
- Função morfossintática: verbo, advérbio, núcleo do objeto ou núcleo de expressão adverbial
- Exemplo típico em LN: Recentemente o conceito de tradução automática encontrou aplicação prática **por meio de produtos de tradução automática.**
- Exemplo típico em UNL:

```

tim(find.@present, recently)
agt(find.@present, concept.@def)
mod(concept.@def, machine translation:01)
obj(find.@present, application)
aoj(practical, application)
met(find.@present, through)
obj(through, product.@pl)
mod(product.@pl, machine translation:02)

```

Subgrupo HB.2 - man

- Função: excluir relação binária rotulada por man + RBs relacionadas a UW2
- Formato da RB: man(UW1, UW2), onde UW1 = ação, UW2 = modo
- UW relevante: UW2
- Hipótese – A ‘maneira’ particular com que uma ação é feita é uma informação complementar (Hovy, 1988).
- Função sintática: adjunto adverbial
- Função morfossintática: verbo, advérbio ou locução adverbial.
- Exemplo típico em LN: Há cerca de três mil línguas diferentes faladas na Terra hoje, **dependendo de como você as conta.**
- Exemplo típico em UNL:

```

...
obj(exist.@present, language.@pl.@indef)
mod(3000, as many as)
qua(language.@pl.@indef, 3000)
aoj(different, language.@pl.@indef)
obj(speak, language.@pl.@indef)
ppl(speak, earth.@def)
tim(speak, today)
man(exist.@present, depend)
obj(depend, count.@present)
man(count.@present, how)
agt(count.@present, you)
obj(count.@present, they)

```

Subgrupo HB.3 – lpl, ppl e plc

- Função: excluir relação binária rotulada por lpl + RBs relacionadas a UW2
- Formato da RB: RL(UW1,UW2), onde UW1 = ação e UW2=local (físico ou lógico)
- UW relevante: UW2
- Hipótese - Hovy (1988) e Robin (1994) sugerem que o local onde uma ação ocorre é uma informação complementar, que pode ou não ser apresentada dependendo do objetivo comunicativo e do grau de detalhe que se pretende apresentar.
- Função sintática: adjunto adverbial
- Função morfossintática: substantivo, advérbio ou locução adverbial

- Exemplo 1 em LN: As diferentes culturas nos legaram a Terra que nós conhecemos hoje, cheia de diversidade e abundância **em tantos aspectos**.
- Exemplo 1 em UNL:

```

...
obj(know, earth)
agt(know, we:02)
tim(know, today)
obj(fill, earth)
mat(fill, :01)
and:01(abundance, diversity:03)
lpl(:01, aspect(icl>phase).@pl)
qua(aspect(icl>phase).@pl, many(equ>a plenty of))
man(many(equ>a plenty of), so(equ>very,icl>degree))

```

Subgrupo HB.4 – aoj

- Função: excluir relação binária rotulada por aoj + RBs relacionadas a UW1
- Formato da RB: aoj(UW1,UW2), onde UW1= atributo ou estado e UW2=entidade
- UW relevante: UW1
- Hipótese: Oka e Ueda (2000) atribuem pouca relevância às relações entre substantivo e modificador (em UNL, representados por aoj e mod). Hutchins (1987), ao exemplificar sua *Deletion rule*, sumariza a sentença “Peter saw a blue ball.” como “Peter saw a ball.”. Em UNL, isso equivale a excluir a RB rotulada por aoj ou mod.
- Função sintática: adjunto adnominal
- Função morfossintática: adjetivo, verbo principal ou verbo de ligação em um predicado nominal¹³
- Exemplo típico em LN: O povo começou a construir uma torre imensa, **que parecia alcançar os céus**.
- Exemplo típico em UNL:

```

agt(begin, people)
obj(begin, build)
agt(build, people)
obj(build, tower)
aoj(huge, tower)
aoj(seem, tower)
obj(seem, reach)
obj(reach, tower)
gol(reach, heaven)

```

Subgrupo HB.5 - tim

- Função: excluir relação binária rotulada por tim + RBs relacionadas a UW2
- Formato da RB: tim(UW1,UW2), onde UW1=ação e UW2=tempo em que a ação ocorre
- UW relevante: UW2
- Hipótese – Hovy (1988) e Robin (1994) sugerem que a informação temporal é uma informação complementar, que pode ou não ser apresentada dependendo do objetivo comunicativo e do grau de detalhe que se pretende apresentar.
- Função sintática: adjunto adverbial
- Função morfossintática: advérbio ou locução adverbial

¹³ Em Sossolote et al. (1997), um verbo de ligação não aparece como categoria morfossintática de ‘aoj’, mas sim de ‘soj’. No entanto, como ‘soj’ não existe na especificação UNL atual e tendo sido subsumida por ‘aoj’, incluímos tal categoria morfossintática como uma das manifestações de ‘aoj’.

- Exemplo típico em LN: Providenciar visitas periódicas aos territórios tutelados **em épocas fixadas de acordo com a autoridade administradora.**
- Exemplo típico em UNL:

```

...
obj(provide for,visit.@pl)
obj(visit.@pl,territory(icl>area).@pl.@def)
mod(visit.@pl,periodic)
mod(territory.@pl.@def,trust)
tim(provide for,time)
obj(agree.@present,time)
ptn(agree.@present,authority.@def)
agt(administer,authority.@def)

```

Subgrupo HB.7 – mod

- Função: excluir relação binária rotulada por mod + RBs relacionadas a UW2
- Formato da RB: mod(UW1,UW2), onde UW1=ação, objeto ou coisa e UW2=modificador de UW1
- UW relevante: UW2
- Hipótese: Oka e Ueda (2000) atribuem pouca relevância aos modificadores de conceitos (em UNL, representados por ‘aoj’ e ‘mod’). Para o Grupo B, consideramos que o modificador é complexo, sendo expresso por mais de uma RB.
- Função sintática: adjunto adnominal, adjunto adverbial
- Função morfossintática: diversas classes de palavras
- Exemplo típico em LN: Um elemento chave no mútuo entendimento internacional é a troca de informação, **particularmente informação sobre a sociedade de um país.**
- Exemplo típico em UNL:

```

...
aoj(exchange.@def.@present, element.@indef)
obj(exchange.@def.@present, information:01)
mod(information:02, particularly)
mod(information:01, information:02)
mod(information:02, society.@indef)
pos(society.@indef, country)

```

Subgrupo HB.9 - src

- Função: excluir relação binária rotulada por src + RBs relacionadas a UW2
- Formato da RB: src(UW1,UW2), onde UW1=ação, objeto ou coisa e UW2=fonte ou estado inicial
- UW relevante: UW2
- Hipótese: Boguraev and Kennedy (1997) consideram que informações na posição de adjuntos são menos proeminentes. Como informações representadas por meio do RL ‘src’ geralmente correspondem a adjuntos, essa RB pode ser excluída. Ainda, algumas informações que poderiam ser codificadas utilizando ‘src’, por exemplo, ‘*Based on these data*’ e ‘*From these results*’ são consideradas não relevantes (Pollock and Zamora, 1975).
- Função sintática: adjunto adverbial, aposto no sujeito
- Função morfossintática: substantivo
- Exemplo típico em LN: Estas métricas não são fáceis de se estabelecer pois qualidade é um termo genérico e dependendo do ponto de vista **de cada pessoa**, qualidade tem um significado diferente.
- Exemplo típico em UNL:

...

```

obj(depending,point of view)
src(point of view,people)
mod(people,each)
seq(depending,:02)
pos:02(meaning.@indef,quality:02)
mod:02(meaning.@indef,different)

```

Subgrupo HB.12 - scn

- Função: excluir relação binária rotulada por scn + RBs relacionadas a UW2
- Formato da RB: scn(UW1,UW2), onde UW1=ação e UW2=cenário de ocorrência da ação
- UW relevante: UW2
- Hipótese: Boguraev and Kennedy (1997) consideram que informações na posição de adjuntos são menos proeminentes. Ainda, informações normalmente codificadas utilizando 'scn', por exemplo, '*In our work*', '*In this study*', são consideradas pouco relevantes (Pollock e Zamora, 1975)¹⁴.
- Função sintática: adjunto adnominal, oração adjetiva
- Função morfossintática: substantivo
- Exemplo típico em LN: O *Administrative Services* provê suporte às atividades acadêmicas do instituto **em serviços gerais, de viagem ...**
- Exemplo típico em UNL:

```

agt(provide,Administrative Services)
obj(provide,support)
mod(support,activity.@def.@pl)
mod(activity.@def.@pl,academic)
mod(activity.@def.@pl,institute.@def)
scn(support,service.@pl)
mod(service.@pl,:01)
and:01(general,travel)
...

```

Subgrupo HB.13 – rsn (cau)

- Função: excluir relação binária rotulada por rsn + RBs relacionadas a UW2
- Formato da RB: RL(UW1,UW2)
- UW relevante: UW2
- Hipótese – Hovy (1988), Marcu (1999) e Rino and Scott (1994) sugerem que relações de causa podem, em alguns casos, ser utilizadas para a sumarização. Em UNL, relações de causa são codificadas por meio do RL 'rsn'.
- Função sintática: adjunto adverbial
- Função morfossintática: substantivo ou núcleo do adjunto adverbial
- Exemplo típico em LN: **Com as recentes inovações em termos de transporte e tecnologias de comunicação de informações**, a demanda por comunicação internacional está crescendo.
- Exemplo típico em UNL:

```

obj(rise,demand)
mod(communication,information)
mod(demand,communication)
aoj(international,communication)
cau(rise,breakthrough)
mod(breakthrough, recent)

```

¹⁴ Vale lembrar que, neste caso, apenas os trechos '*In our work*' e '*In this study*' serão excluídos, preservando a informação que eles introduzem que, possivelmente, terá grande importância.

```

lpl(breakthrough, technology)
mod(means(icl>method), transportation)
and(technology, means)
mod(technology, communication)

```

Subgrupo HB.14 - pur

- Função: excluir relação binária rotulada por pur + RBs relacionadas a UW2
- Formato da RB: src(UW1,UW2), onde UW1=ação, objeto ou coisa e UW2=fonte ou estado inicial
- UW relevante: UW2
- Hipótese: Boguraev and Kennedy (1997) consideram que informações na posição de adjuntos são menos proeminentes. Como informações representadas pelo RL ‘pur’ podem corresponder a adjuntos, serão consideradas complementares. Ainda, Marcu (1998) verificou que sujeitos humanos consideram a informação com a função de propósito pouco importante. Como esse tipo de informação é representada por meio do RL ‘pur’, essa RB pode ser excluída.
- Função sintática: adjunto adverbial, objeto indireto
- Função morfossintática: substantivo, núcleo do adjunto adverbial
- Exemplo típico em LN: O ILO foi estabelecido em 1919 para promover justiça social **para pessoas trabalhadoras em todos os lugares.**
- Exemplo típico em UNL:

```

...
obj(promote, justice)
mod(justice, social)
pur(justice, people)
mod(people, working)
plc(people, everywhere)

```

Subgrupo HB.15 – dur

- Função: excluir relação binária rotulada por dur + RBs relacionadas a UW2
- Formato da RB: dur(UW1,UW2), onde UW1=ação ou evento e UW2=duração de UW1
- UW relevante: UW2
- Hipótese –Boguraev and Kennedy (1997) consideram que informações na posição de adjuntos são menos proeminentes. Como informações representados por meio do RL ‘dur’ geralmente remetem a adjuntos, essa RB pode ser excluída.
- Função sintática: adjunto adverbial, adjunto adnominal
- Função morfossintática: substantivo, oração preposicional
- Exemplo típico em LN: **Nos últimos anos**, a Internet se tornou uma presença dramática, forçando mudanças radicais no modo como a informação flui.
- Exemplo típico em UNL:

```

dur(become.@present, year.@pl.@def)
qua(year.@pl.@def, a couple of)
mod(year.@pl.@def, past)
obj(become.@present, Internet.@def)
gol(become.@present, presence.@indef)
aoj(dramatic, presence.@indef)
man(become.@present, force)
obj(force, change.@pl)
aoj(radical, change.@pl)
lpl(change.@pl, way.@def)
man(flow.@present, way.@def)
obj(flow.@present, information)

```

Subgrupo HB.16 - con

- Função: excluir relação binária rotulada por con + RBs relacionadas a UW2
- Formato da RB: con(UW1,UW2), onde UW1=ação ou evento e UW2=condição paraUW1 ocorrer
- UW relevante: UW2
- Hipótese – Boguraev and Kennedy (1997) consideram que informações na posição de adjuntos são menos proeminentes. Como informações representadas utilizando o RL ‘con’ geralmente correspondem a adjuntos, essa RB contém informação pouco proeminente e, portanto, pode ser excluída.
- Função sintática: adjunto adverbial.
- Função morfossintática: verbo, núcleo de oração subordinada condicional
- Exemplo típico em LN: O Conselho de Segurança deverá realizar reuniões periódicas, nas quais cada membro, **se assim o desejar**, será representado por um membro do governo ou por outro representante especialmente designado.
- Exemplo típico em UNL:

```
agt(hold.@present,Security Council.@def)
obj(hold.@present,meeting.@pl)
mod(meeting.@pl,periodic)
obj(represent,member:01.@pl)
mod(member:01.@pl,each)
lpl(represent,meeting.@pl)
con(represent,d Desire.@present)
man(d Desire.@present,if)
mod(member:01.@pl,it:01)
agt(represent.@present,representative)
or(representative,member:02.@indef)
man(d Desire.@present,so)
agt(d Desire.@present,it:02)
mod(member:02.@indef,government.@def)
...
```

Subgrupo HB.17 - agt

- Função: excluir relação binária rotulada por agt + RBs relacionadas a UW1
- Formato da RB: agt(UW1,UW2), onde UW1=ação e UW2=agente
- UW relevante: UW2
- Hipótese – Quando o sujeito aparece duas vezes na sentença UNL, sendo uma principal, outra secundária, esta última pode ser excluída. Geralmente, isso acontece em orações subordinadas.
- Função sintática: sujeito
- Função morfossintática: substantivo, pronome
- Exemplo típico em LN: Se você não teve a chance de aprender inglês, **ou se você esteve evitando a tarefa**, então você está automaticamente excluído do acesso a informação high-tech.
- Exemplo típico em UNL:

```
agt:01(have.@present.@not, you)
obj:01(have.@present.@not, chance.@def)
mod:01(chance.@def, learn)
obj:01(learn, English)
agt:01(avoid.@present, you)
obj:01(avoid.@present, undertake.@def)
or:01(avoid.@present,have.@present.@not)
con(exclude.@present, S01)
man(exclude.@present.@pred.@entry, then)
```

```
obj(exclude.@present.@pred.@entry, you)
man(exclude.@present.@pred.@entry, automatically)
src(exclude.@present.@pred.@entry, access)
obj(access, information)
aoj(high-tech, information)
```

Subgrupo HB.18 - ben

- Função: excluir relação binária rotulada por ben + RBs relacionadas a UW2
- Formato da RB: ben(UW1,UW2), onde UW1=ação ou objeto e UW2=beneficiário ou receptor de UW1
- UW relevante: UW2
- Hipótese: O beneficiário ou vítima de um evento (ou estado) está numa posição menos proeminente do que o evento ou estado.
- Função sintática: complemento nominal, aposto ou objeto indireto
- Função morfossintática: substantivo
- Exemplo típico em LN: Acessar a Internet se tornou tão fácil **para os lares em geral**, que se tornou comum para as pessoas usar a Internet diariamente.
- Exemplo típico em UNL:

```
obj(access, Internet:01)
obj(become:01, access)
man(easy, so)
gol(become:01, easy)
ben(easy, household.@pl)
aoj(general, household.@pl)
...
```

4. Metodologia de poda

4.1. Processo Geral

O processo geral de aplicação das heurísticas é composto pelos seguintes passos:

1. Busca: Procura-se as heurísticas aplicáveis à sentença UNL, agrupando-as em um conjunto Hp (Heurísticas possíveis), percorrendo-se a sentença RB por RB. Para cada RB, todas as heurísticas do subgrupo relacionado ao seu RL são verificadas.
2. Ordenação: As heurísticas do conjunto Hp são ordenadas conforme as prioridades previamente determinadas pelo usuário (como veremos adiante na seção 4.2).
3. Aplicação: A primeira heurística de Hp é aplicada (salvo por restrição de limite, como veremos adiante; nesse caso, o processo é encerrado) e, em seguida, excluída da lista. É gerada, assim a sentença sumarizada S'.
4. Atualização: Verifica-se novamente todas as heurísticas encontradas no passo 1, para ver se nenhuma ficou invalidada após o passo 3. Em caso afirmativo, ela é retirada da lista.
5. Re-poda: Se o número máximo de heurísticas estabelecido pelo usuário não foi alcançado, volta-se ao passo 3. Caso contrário, o processo é encerrado.

Para garantir a aplicação apenas de heurísticas válidas, após o passo 3, todas as heurísticas de Hp são novamente verificadas para verificar se ainda são válidas (passo 4). Isso é necessário, pois, em alguns casos, a aplicação de uma heurística pode modificar a sentença UNL de forma a invalidar a aplicação de outra.

O protótipo permite também ao usuário customizar a poda. Ele pode: a) determinar um número máximo de heurísticas a ser aplicado por sentenças, controlando assim o tamanho do sumário e b) dar prioridade ao Grupo A ou B de heurísticas. Em relação ao item a), observamos que quanto mais heurísticas são aplicadas, pior é a qualidade do sumário. Assim, o usuário poderá ter um certo controle sobre a relação tamanho do sumário/qualidade, conforme seus objetivos. Em relação ao item b), como verificamos em estudos preliminares, a prioridade de um ou outro grupo também terá influência na qualidade e taxa de compressão do sumário.

Além dos cuidados já citados, para garantir a coesão da sentença sumarizada, alguns procedimentos especiais precisam também ser adotados durante o processo de poda.

4.1.1. Procedimentos Especiais

Esses procedimentos dizem respeito, principalmente, ao escopo (como visto no Cap.2) utilizado para agrupar RBs. Como eles são posicionados, em uma RB, no lugar de uma UW, são necessários alguns cuidados para garantir a aplicação correta das heurísticas. São tomados também cuidados em relação ao número de RBs excluídas. Há, portanto, quatro tipos de casos especiais de processamento. Os 3 primeiros dizem respeito ao escopo e são verificados no passo 1, enquanto o último caso é verificado no passo 3, antes da aplicação da heurística.

1) UW escopo no Grupo A

As heurísticas do Grupo A excluem apenas uma RB. Portanto, se a UW relevante verificada pela heurística for uma referência a escopo (p. ex., ':01'), a heurística não deve ser (e não será) aplicada, pois a informação relativa àquela RL está representada por meio do grupo de heurísticas que fazem parte do escopo.

Por exemplo, na sentença “Uruguay took the lead versus Argentina through goal scored by Pablo Dorado” [S2] codificada em UNL usando escopo (Exemplo 3, transcrito abaixo), a heurística HA.1.1 poderia ser inapropriadamente aplicada à RB 4, se a referência ao escopo ‘:01’ fosse confundida com uma UW, já que essa referência não ocorre em outras RBs dessa sentença.

agt(take.@past,Uruguay)	(1)
obj(take.@past,lead.@def)	(2)
ptn(take.@past,Argentina)	(3)
met(take.@past.@pred,:01)	(4)
obj:01(score.@past,goal)	(5)
agt:01(score.@past,'Pablo Dorado')	(6)

Exemplo 3: Representação UNL para S2 usando escopo

2) RB pertencente a escopo no Grupo A

No Grupo A, se a RB a ser excluída for a única com um determinado escopo, a heurística não será aplicada. Isso é feito para impedir que alguma RB faça referência a um escopo inexistente. Por exemplo, na sentença UNL (Exemplo 5) correspondente à sentença “Todos os membros das Nações Unidas devem promover a paz no mundo inteiro” [S3], a RB 6 não poderá ser excluída, caso contrário a RB 5 fará referência a uma RB inexistente.

mod(member, all)	(1)
mod(member.@pl,United Nations.@def)	(2)
agt(further,member)	(3)
obj(further,peace)	(4)
plc(peace, :01)	(5)
mod:01(world,whole)	(6)

Exemplo 5: Representação UNL para S3

3) UW escopo no grupo B

Nas heurísticas do grupo B, se, dentro do subgrupo de RBs a serem excluídas, houver alguma referência a escopo, todas as RBs pertencentes a esse escopo serão também excluídas. Por exemplo, na codificação UNL correspondente à sentença “Todos os membros das Nações Unidas devem promover a paz, ajudando membros das Nações Unidas e outros povos” [S4] (Exemplo 6), aplicando a heurística HB.2.3, a RB 5 mais todas RBs com escopo ‘:01’ devem ser excluídas. Como a RB 6, que será uma das excluídas, faz referência ao escopo ‘:02’, a RB 7 também será excluída.

mod(member.@pl:01, all)	(1)
mod(member.@pl:01,United Nations.@def)	(2)
agt(further,member)	(3)
obj(further,peace)	(4)
man(further, :01)	(5)
obj:01(help,:02)	(6)
and:02(member.@pl:02,people)	(7)
mod:01(member.@pl:02, United Nations)	(8)
mod:01(people,other)	(9)

Exemplo 6: Representação UNL para S4

4) Limite de aplicação

Em heurísticas do Grupo B, dependendo da codificação da sentença UNL, a poda encadeada pode levar à exclusão de todas as RBs, ou seja, de toda a sentença. Para que isso não ocorra, a heurística só será aplicada se, após sua exclusão, ainda restarem pelo menos 4 RBs. De acordo com nossos dados, esse número de RBs é o mínimo utilizado para representar sentenças curtas. Conseqüentemente, sentenças originalmente formadas por 4 RBs não serão sumarizadas.

4.1.2. Exemplo de Poda

Para melhor ilustrar o processo de poda, vamos apresentar o processo de sumarização da sentença “Como diversas são as áreas que se beneficiam com a utilização da Internet e alguns atributos relevantes dependem do domínio da aplicação, a área educacional foi focalizada neste trabalho.” [S5], retirada de um texto científico, cuja representação UNL é apresentada no Exemplo 5. Vamos considerar que as heurísticas do Grupo B serão priorizadas e que não há limite para o número de heurísticas aplicadas.

[S]	
rsn(:04,:01)	(1)
and:01(:02,:03)	(2)
qua:02(area.@pl,many)	(3)
ben:02(benefit.@entry,area.@pl)	(4)
src:02(benefit.@entry,Internet.@def)	(5)
agt:03(depend,attribute.@pl)	(6)
mod:03(attribute.@pl,relevant)	(7)
qua:03(attribute.@pl,some)	(8)
obj:03(depend,domain.@def)	(9)
mod:03(domain.@def,application)	(10)
obj:04(focalize.@past,area.@def)	(11)
mod:04(area.@def,educational)	(12)
plc:04(focalize.@past,work)	(13)
mod:04(work,this)	(14)
[/S]	

Exemplo 7: Representação UNL para S5

- Passo 1: Percorrendo-se a sentença UNL, as heurísticas aplicáveis vão sendo adicionadas ao conjunto Hp, gerando-se o conjunto $H_p = \{1) HB.13.3; 2) HA.10.1; 3) HB.18.1; 4) HA.9.1; 5) HA.7.10; 6) HA.7.12; 7) HA.10.1; 8) HA.7.7; 9) HA.7.7; 10) HB.3.5\}$ ¹⁵.
- Passo 2: Assumindo que as heurísticas do grupo B serão priorizadas, Hp é reordenado e passa a ser $\{1)HB.13.3; 2)HB.18.1; 3)HB.3.5; 4)HA.10.1; 5)HA.9.1; 6) HA.7.10; 7) HA.7.12; 8) HA.10.1; 9) HA.7.7; 10) HA.7.7\}$.
- Passo 3: A primeira heurística de Hp, isto é, HB.6.3, é aplicada, gerando a sentença UNL sumarizada S4' (Exemplo 6). Retirando essa heurística, Hp passa a ser $\{1)HB.18.1; 2)HB.3.5; 3)HA.10.1; 4)HA.9.1; 5) HA.7.10; 6) HA.7.12; 7) HA.10.1; 8) HA.7.7; 9) HA.7.7\}$.

¹⁵ Vide definição de todas as heurísticas no Anexo 1.

[S]	
obj:04(focalize.@past,area.@def)	(11)
mod:04(area.@def,educational)	(12)
plc:04(focalize.@past,work)	(13)
mod:04(work,this)	(14)
[/S]	

Exemplo 8: Representação UNL para S5' (sumário de S5)

- Passo 4: Tomando-se a primeira heurística de Hp, HB.18.1, verificamos que ela não é mais válida, pois se refere à exclusão das RBs 3 e 4, que já foram excluídas. Portanto, essa heurística é retirada de Hp. De forma semelhante, todas as demais heurísticas são verificadas, gerando-se assim o conjunto Hp = {1) HB.3.5; 2) HA.7.7}.
- Passo 5: assumindo que não há número máximo de aplicação de heurísticas, volta-se ao passo 3.
- Passo 3: Verificamos que só restam 4 RBs, número que foi estabelecido como mínimo para uma sentença. Portanto, mais nenhuma heurística pode ser aplicada e o processo é encerrado.

A sentença sumarizada, portanto, é “A área educacional foi focalizada neste trabalho” [S5'] (Exemplo 8).

4.2. Priorizando Heurísticas para Seleção

Vimos que, no passo 2, as heurísticas podem ser selecionadas conforme o Grupo, A ou B, que se queira priorizar. No entanto, quando existem mais de uma heurística do mesmo grupo aplicáveis a uma mesma sentença, é necessário algum método para decidir qual delas será aplicada primeiro.

Para estabelecer uma medida de prioridade que auxilie essa seleção, conduzimos um estudo para verificar o desempenho de cada heurística. Aquelas que apresentaram o melhor desempenho receberam maiores prioridades. Aquelas com desempenho ruim receberam baixas prioridades.

Utilizamos nesse estudo os dois corpora de treinamento (*UN Charter e Booklet*), para os quais aplicamos, sentença por sentença, todas as heurísticas possíveis, uma a uma. Ou seja, a sentença era sumarizada por meio de uma única heurística e, em seguida, avaliada em relação à textualidade e manutenção da idéia principal, conforme a Tabela 1, recebendo um *score* de 1 a 6.

Assim, cada heurística recebeu um *score*, conforme a avaliação do sumário que produziu. Após a aplicação nos dois corpora, foram calculadas as médias dos *scores* de cada heurística, nos quais foram baseados os valores de prioridade.

Textualidade	Informação Principal	Score
Mantida	Preservada	6
Prejudicada	Preservada	5
Mantida	Preservada parcialmente	4
Prejudicada	Preservada parcialmente	3
Mantida	Não preservada	2
Prejudicada	Não preservada	1

Tabela 1: Scores dos sumários em relação à textualidade e preservação da idéia principal

Estabelecemos que seriam consideradas aceitáveis as heurísticas cujos sumários produzidos por sua aplicação tivessem obtido *score* acima de 3.5 (média entre 1 e 6). Assim, heurísticas com *score* entre 1 e 3.5 receberam o menor valor de prioridade, ou seja, 6 (ver Tabela 2). Os *scores* restantes foram igualmente distribuídos entre os valores de prioridade restantes.

Algumas heurísticas não tiveram nenhuma ocorrência nos dois corpora, i.e., não foram aplicadas nenhuma vez. Nesse caso, os valores de prioridade foram calculados em função dos valores das outras heurísticas do mesmo subgrupo. Por exemplo, a heurística HA.3.3 (Anexo 3), relativa ao RL ‘plc’, não foi aplicada nenhuma vez. Isso se deve ao fato de que naqueles corpora utilizava-se os RLs ‘ppl’ e ‘lpl’, que foram mais tarde excluídos do conjunto de RLs da UNL, tendo sido substituídos por ‘plc’. Portanto, o valor de prioridade para HA.3.3 foi calculado em função dos *scores* obtidos pelas heurísticas correspondentes a ‘ppl’ e ‘lpl’ (Figura 4), recebendo, de acordo com a Tabela 2, o valor 4.

Score HA.3.1 (lpl) = 5.2	Score HA.3.2 (ppl) = 3.77
Score HA.3.3 (plc) = $\frac{5.2 + 3.77}{2} = 4.49$	

Figura 4: Cálculo de prioridade para HA.3.3

No Anexo 3, as heurísticas com o símbolo * tiveram sua prioridade calculada dessa forma. Nos subgrupos onde nenhuma heurística foi aplicada, as heurísticas receberam temporariamente o valor de prioridade 3. Futuramente, testes adicionais deverão ser realizados para reajustar a distribuição de prioridades de aplicação.

Score	Valor de Prioridade
5,6 – 6	1
5,1 – 5,5	2
4,6 – 5	3
4,1 – 4,5	4
3,6 – 4	5
1 – 3,5	6

Tabela 2: Intervalos de Prioridade de Heurísticas

5. Discussão Final

Este relatório apresentou um conjunto de heurísticas para a sumarização automática de textos UNL, incluindo a metodologia de especificação, sua descrição completa e o modo como elas são aplicadas no protótipo UNLSumm.

A metodologia de especificação compreendeu três tarefas principais: 1) a retroprojeção dos RLs nas manifestações morfossintáticas para o Português do Brasil e para o Inglês, 2) a coleta de evidências literárias indicando possíveis construções irrelevantes e 3) a análise subjetiva de corpora de textos UNL, de modo a determinar empiricamente RLs que indicassem informações pouco relevantes.

Confrontando os resultados gerados nessas três tarefas, foi desenvolvido o elenco de 84 heurísticas, sendo 39 do Grupo A e 45 do Grupo B. Vale notar que as heurísticas propostas nem sempre serão suficientes para garantir apenas a exclusão de informações irrelevantes, pois o contexto de ocorrência ou a diversidade de manifestações de um único RL exercem influência na relevância de uma RB. Assim, a exclusão indiscriminada de RBs de textos UNL, somente pela aplicação de alguma heurística, ainda pode causar desvios de significado. Esse é o caso de ‘aoj’, que pode ser utilizado para representar um adjetivo irrelevante, mas também pode relacionar um sujeito e um verbo que não indica ação, como na sentença ‘I know Mary’¹⁶[S6], codificada como no Exemplo 6.

aoj(know, I) obj(know, Mary)

Exemplo 9: Representação UNL para S6

Assim, estudos deverão ser realizados para verificar a consistência e eficácia da aplicação de algumas heurísticas em casos como esses.

¹⁶ Exemplo extraído da Especificação UNL (Uchida, 2000).

Referências Bibliográficas

- Aone, C., Gorlinsky, J., Larsen, B. and Okurowski, M. E. (1999). A Trainable Summarizer with Knowledge Acquired from Robust NLP Techniques. In: I. Mani and Maybury (eds.) *Advances in Automatic Summarization*, pp. 71-80. The MIT Press. London, England.
- Beaugrande, R. De and Dressler, W. V. (1981). *Introduction to Text Linguistics*. Longman. London.
- Boguraev, B. and Kennedy, C. (1997). Saliency-based Content Characterisation of Text Documents. In I. Mani and M. Maybury (eds) *Proceedings of Intelligent Scalable Text Summarization ACL Workshop*, pp.2-9. Madrid, Spain.
- Brown, A. L. and Day, J. D. (1983). Macrorules for summarizing texts: the development of expertise. *Journal of Verbal Learning and Verbal Behavior* 22, pp. 1-14.
- Buxton, A. B. and Meadows, A. J. (1978). Categorization of the information in experimental papers and their author abstracts. *Journal of Research in Communication Studies* 1:161-182.
- Hovy, E. (1988). *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey.
- Hutchins, John. *Summarization: Some problems and Methods*. In: Jones. *Meaning: The frontier of informatics*. Cambridge. London, pp. 151-173, 1987.
- Koch, I. G. V. (1995). Aquisição da Escrita e Textualidade. *Cadernos de Estudos Lingüísticos* 29, pp. 107-117. Campinas, SP.
- Marcu, D. (1998). To build text summaries of high quality, nuclearity is not sufficient. In: *Intelligent Text Summarization*. Papers from the 1998 AAAI Spring Symposium. Technical Report SS-98-06.
- Marcu, D. (1999). Discourse trees are good indicators of importance of text. In: I. Mani and M. Maybury (eds.) *Advances in Automatic Text Summarization*. The MIT Press.
- Marcuschi, L. A. (1983). *Linguística de Texto: o que é e como se faz*. *Série Debates 1*. Universidade Federal de Pernambuco.
- Martins, C. B. (2000). *A Sumarização de Textos UNL*. Projeto de Mestrado. PPG-CC, DC-UFSCar (Bolsa CAPES/Mestrado). São Carlos, SP.
- Martins, R.T.; Rino, L.H.M.; Nunes, M.G.V.; Montilha, G.; Oliveira Jr., O.N. (2000). An Interlingua Aiming at Communication on the Web: How Language-Independent Can it Be? In: *Proceedings of the Workshop on Applied Interlinguas: Practical Applications of Interlingual Approaches to NLP*, pp. 24-33. NAACL-ANLP 2000 Workshop, April. Seattle, Washington, USA.

- Oka, M. and Ueda, Y. (2000). Evaluation of Phrase-Representation Summarization based on Information Retrieval Task. In: *Proceedings of the ANLP/NAACL Automatic Summarization Workshop*, pp. 59-68. Seattle, Washington.
- Pollock, J. J. and Zamora, A. (1975). Automatic Abstracting Research at Chemical Abstracts Service. *Journal of Chemical Information and Computer Sciences* 15(4), pp. 226-232.
- Rino, L. H. M. and Scott, D. (1994). *Automatic Generation of Draft Summaries: Heuristics for Content Selection*. ITRI-94-8 Technical Report. University of Brighton. Brighton, UK.
- Robin, J. (1994) *Revision-based generation of natural language summaries providing historical background: corpus-based analysis, design, implementation and evaluation*. Ph.D. Thesis. CUCS-034-94, Columbia University, Computer Science Department, New York, USA.
- Sossolote, C. R. C., Zavaglia, C., Rino, L. H. M. e Nunes, M. G. V. (1997). *As Manifestações Morfossintáticas da Linguagem UNL no Português do Brasil*. Tech. Report NILC-TR-97-2 (Relatórios Técnicos do ICMSC, n. 36). São Carlos, SP, Brazil.
- Uchida, H. (1997). *DeConverter Specification, Version 1.0*. Tech. Rep. UNL-TR1997-010, UNU/IAS/UNL Center, Tokyo, Japan.
- Uchida, H., Zhu, M. and Della Senta, T. (1999). *The UNL, a Gift for a Millennium*. Institute of Advanced Studies. The United Nations University. Tokyo, Japan (também disponível em <http://www.unl.ias.unu.edu>).
- Uchida, H. (2000). *Universal Networking Language: An Electronic Language for Communication, Understanding and Collaboration*. UNL Center, IAS/UNU, Tokyo, Japan.
- Vallduvi, E. (1993). *Information Packaging: A survey*. Research Paper HCRC/RP-44. HCRC Publications. University of Edinburgh. Scotland, UK.

ANEXO 1
Relação completa das Heurísticas de Sumarização – Versão 1.3

RL	GRUPO A	GRUPO B
met	HA.1.1. Excluir met(a,b) - se b \notin outras RBs e não existe obj(a,x).	HB.1.1. Excluir met(a,b) obj(b, c) mod(c, d) – se b, c,d \notin RBs fora do subgrupo. HB.1.2. Excluir met(a,b) obj(a,c) aoj(d,b) – se b,c e d \notin RBs fora do subgrupo. HB.1.3. Excluir met(x,b) obj(b,a) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.1.4. Excluir met(x, b) mod(b,c) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.1.5. Excluir met(a,b) aoj(c,b) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.1.6. Excluir met(a,b) + {RBs com escopo ‘b’} – se eh_escopo(b). HB.1.7. Excluir met(a,b) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo.
man	HA.2.1. Excluir man(a,b) - se b \notin outras RBs e existe aoj(a,x) HA.2.2. Excluir man(a,b) – se b \notin outras RBs e existe obj(a,x) HA.2.3. Excluir man(a,b) – se b \notin outras RBs e existe mod(x,a) HA.2.4. Excluir man(a,b) - se b \notin outras RBs e existe ins(a,x) HA.2.5. Excluir man(a,b) – se b \notin outras RBs e existe agt(a,x) HA.2.6. Excluir man(a,b) – se b \notin outras RBs e existe gol(a,x) HA.2.7. Excluir man(a,b) – se b \notin outras RBs e existe mod(a,x) HA.2.8. Excluir man(a,b) – se b \notin outras RBs e existe bem(a,x)	HB.2.1. Excluir man(a,b) obj(b,c) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.2.2. Excluir man(a,b) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.2.3. Excluir man(a,b) + RBs com escopo ‘b’ – se eh_escopo(b).
lpl	HA.3.1. Excluir lpl(a,b) – se b \notin outras RBs.	HB.3.1. Excluir lpl(a,b) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.3.2. Excluir lpl(a,b) + RBs com escopo ‘b’ – se eh_escopo(b).
ppl	HA.3.2. Excluir ppl(a,b) – se b \notin outras RBs.	HB.3.3. Excluir ppl(a,b) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.3.4. Excluir ppl(a,b) + RBs com escopo ‘b’ – se eh_escopo(b).
plc	HA.3.3. Excluir plc(a,b) – se b \notin outras RBs.	HB.3.5. Excluir plc(a,b) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo. HB.3.6. Excluir plc(a,b) + RBs com escopo ‘b’ – se eh_escopo(b).
aoj	HA.4.1. Excluir aoj(a,b) – se a \notin outras RBs e existe obj(x,b) HA.4.2. Excluir ben(a,b) – se a \notin outras RBs e existe ben(x,b) HA.4.3. Excluir mod(a,b) – se a \notin outras RBs e existe mod(b,x).	HB.4.1. Excluir aoj(a,b) obj(a,c) – se b, c \notin RBs fora do subgrupo. HB.4.2. Excluir aoj(a,b) obj(a,c) – se existe agt(d,b) e obj(d,e) HB.4.3. Excluir aoj(a,b) + {RBs \in subgrupo S1} – se UWs \in S1 \notin RBs fora do subgrupo.
tim	HA.5.1. Excluir tim(a,b) – se b \notin outras RBs e	HB.5.1. Excluir tim(a,b) + {RBs \in subgrupo

	<p>existe man(a,x) HA.5.2. Excluir tim(a,b) – se b ∉ outras RBs e existe agt(a,x) HA.5.3. Excluir tim(a,b) – se b ∉ outras RBs e existe obj(a,x) HA.5.4. Excluir tim(a,b) – se b ∉ outras RBs e existe mod(a,x) HA.5.5. Excluir tim(a,b) – se b ∉ outras RBs e existe aoj(a,x) HA.5.6. Excluir tim(a,b) – se b ∉ outras RBs</p>	<p>S1} – se UWs ∈ S1 ∉ RBs fora do subgrupo. HB.5.2. Excluir tim(a,b) + RBs com escopo 'b' – se eh_escopo(b).</p>
pos	HA.6.1. Excluir pos(a,b) – se b ∉ outras RBs.	-
mod	<p>HA.7.1. Excluir mod(a,b) – se b ∉ outras RBs e existe and(a,x)) ou and(x,a). HA.7.2. Excluir mod(a,b) – se b ∉ outras RBs e existe aoj(a,x). HA.7.3. Excluir mod(a,b) – se b ∉ outras RBs e existe ben(x,a). HA.7.4. Excluir mod(a,b) – se b ∉ outras RBs e existe lpl(x, a). HA.7.5. Excluir mod(a,b) – se b ∉ outras RBs e existe mod(x,a). HA.7.6. Excluir mod(a,b) – se b ∉ outras RBs e existe mod(a,x) HA.7.7. Excluir mod(a,b) – se b ∉ outras RBs e existe obj(x,a). HA.7.8. Excluir mod(a,b) – se b ∉ outras RBs e existe obj(a,x) HA.7.9. Excluir mod(a, b) – se b ∉ outras RBs e existe pof(a, x) HA.7.10. Excluir mod(a,b) - se b ∉ outras RBs e existe qua(a,x). HA.7.11. Excluir mod(a,b) – se b ∉ outras RBs e existe aoj(x,a) HA.7.12. Excluir mod(a,b) – se ∉ outras RBs e existe agt(x,a) HA.7.13. Excluir mod(a,b) – se b ∉ outras RBs e existe or(x,a) ou or(a,x)</p>	<p>HB.7.1. Excluir mod(a,b) aoj(c,b) – se b,c ∉ RBs fora do subgrupo. HB.7.2. Excluir mod(a,b) and(b,c) - se b,c ∉ RBs fora do subgrupo. HB.7.3. Excluir mod(a,b) + RBs com escopo 'b' – se eh_escopo(b). HB.7.4. Excluir mod(a,b) + {RBs ∈ subgrupo S1} – se UWs ∈ S1 ∉ RBs fora do subgrupo.</p>
nam	HA.8.1. Excluir nam(a,b) – se b ∉ outras RBs.	-
src	HA.9.1. Excluir src(a,b) – se b ∉ outras RBs e existe obj(x,a).	<p>HB.9.1. Excluir src(a,b) + {RBs ∈ subgrupo S1} – se UWs ∈ S1 ∉ RBs fora do subgrupo. HB.9.2. Excluir src(a,b) tim:b(c,d) – se c,d ∉ RBs fora do subgrupo. HB.9.3. Excluir src(a,b) + RBs com escopo b – se eh_escopo(b)</p>
qua	HA.10.1. Excluir qua(a,b) – se b ∉ outras RBs.	-
cnt	HA.11.1. Excluir cnt(a,b) – se b ∉ outras RBs.	-
scn		<p>HB.12.1. Excluir scn(a,b) + {RBs ∈ subgrupo S1} – se UWs ∈ S1 ∉ RBs fora do subgrupo. HB.12.2. Excluir scn(a,b) + RBs com escopo b – se eh_escopo(b).</p>
rsn	-	<p>HB.13.1. Excluir rsn(a,b) + {RBs ∈ subgrupo S1} – se UWs ∈ S1 ∉ RBs fora do subgrupo. HB.13.3. Excluir rsn(a,b) + RBs com escopo b – se eh_escopo(b)</p>

cau	-	<p>HB.13.2. Excluir $\text{cau}(a,b) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p> <p>HB.13.4. Excluir $\text{cau}(a,b) + \text{RBs}$ com escopo b – se $\text{eh_escopo}(b)$.</p>
pur	-	<p>HB.14.1. Excluir $\text{pur}(a,b) \text{ obj}(b,c) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p> <p>HB.14.2. Excluir $\text{pur}(a,b) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p> <p>HB.14.3. Excluir $\text{pur}(a,b) + \text{RBs}$ com escopo b – se $\text{eh_escopo}(b)$.</p>
dur	-	<p>HB.15.1. Excluir $\text{dur}(a,b) \text{ mod}(b,c) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p> <p>HB.15.2. Excluir $\text{dur}(a,b) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p> <p>HB.15.3. Excluir $\text{dur}(a,b) + \text{RBs}$ com escopo b – se $\text{eh_escopo}(b)$.</p>
con	-	<p>HB.16.1. Excluir $\text{con}(a,b) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p> <p>HB.16.2. Excluir $\text{con}(a,b) + \text{RBs}$ com escopo b – se $\text{eh_escopo}(b)$.</p>
agt	-	<p>HB.17.1. Excluir $\text{agt}(a,b) \text{ obj}(a,c) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se existe $\text{agt}(x,b)$ e $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p>
ben	-	<p>HB.18.1. Excluir $\text{ben}(a,b) + \{\text{RBs} \in \text{subgrupo } S1\}$ – se $\text{UWs} \in S1 \notin \text{RBs}$ fora do subgrupo.</p> <p>HB.18.2. Excluir $\text{ben}(a,b) + \text{RBs}$ com escopo b – se $\text{eh_escopo}(b)$.</p>

ANEXO 2

Relation Labels Considerados Relevantes para a Sumarização Automática de Textos UNL

Apresentamos abaixo uma breve descrição do significado de cada Relation Label utilizado neste projeto. As definições foram generalizadas aqui para os principais usos, sendo a especificação UNL bem mais abrangente. Para a especificação original, refira-se a (Uchida, 2000).

- **met** (método)
Sintaxe: met(ação, evento)
Descreve o modo utilizado para a realização de uma ação ou evento.
- **man** (modo ou maneira)
Sintaxe: man(ação, modo)
Descreve um modo específico como uma ação foi realizada ou característica de um evento.
- **plc** (local)
Sintaxe: plc(ação, local)
Descreve o lugar onde uma ação ocorre, seja físico ou lógico.
- **lpl** (lugar lógico)
Sintaxe: lpl(ação, lugar)
Descreve o cenário, ou seja, lugar lógico (não real no sentido físico) onde uma ação ocorre.
- **ppl** (lugar físico)
Sintaxe: ppl(ação, lugar)
Descreve o lugar físico onde uma ação ocorre.
- **aoj** (atributo de um elemento¹⁷)
Sintaxe: aoj(atributo, elemento)
Descreve o estado ou o atributo de um elemento
- **tim** (tempo)
Sintaxe: tim(ação, tempo)
Descreve uma informação temporal sobre a ação.
- **pos** (possuidor)
Sintaxe: pos(objeto, elemento)
Descreve o elemento possuidor de um objeto ou lugar.
- **mod** (modificador)
Sintaxe: mod(elemento, modificador)

¹⁷ Utilizaremos o termo ‘elemento’ como tendo o significado de pessoa, objeto ou qualquer ser ou coisa que, em uma sentença, apareça como volitiva.

Descreve um modificador (seja atributo, adjetivo ou outro) que restringe o elemento

- **nam** (nome)
Sintaxe: nam(elemento,nome)
Descreve o nome próprio do elemento
- **src** (estado inicial)
Sintaxe: src(elemento,estado inicial)
Descreve o estado inicial de elemento em um evento ou estado
- **scn** (cenário)
Sintaxe: scn(evento,cenário)
Descreve um cenário virtual ou lógico (ou seja, não físico) no qual um evento ocorre.
- **qua** (quantidade)
Sintaxe: qua(elemento,quantidade)
Descreve a quantidade do elemento, que pode também ser uma unidade de medida.
- **rsn** (razão)
Sintaxe: rsn(evento,razão)
Descreve a razão de um evento acontecer ou um estado existir
- **pur** (propósito)
Sintaxe: pur(ação,propósito)
Descreve o propósito ou objetivo de uma ação ou propósito ou objetivo de algo existir.
- **dur** (duração)
Sintaxe: dur(evento,duração)
Descreve o período de tempo durante o qual um evento ocorre ou um estado existe
- **con** (condição)
Sintaxe: con(evento,condição)
Descreve um evento ou estado que condiciona ou o evento em foco.
- **agt** (agente)
Sintaxe: agt(ação,agente)
Descreve o agente que inicia uma ação
- **ben** (beneficiário)
Sintaxe: ben(ação,beneficiário)
Descreve o beneficiário ou vítima de um evento ou estado

ANEXO 3

Valor de Prioridade de cada Heurística

Os valores “Média Scores” apresentados abaixo foram obtidos por meio da aplicação de cada heurística às sentenças de dois corpora de textos UNL, UN Charter e Corpus Booklet. Os sumários gerando por sua aplicação foram então avaliados de acordo com a sua textualidade e manutenção da idéia principal, recebendo um *score* de 1 a 6. Em seguida, cada heurística recebeu um valor Prioridade, de acordo com a média aritmética simples de todos os *scores* obtidos. Os valores com o símbolo * se referem a heurísticas que não foram aplicadas nenhuma vez e, portanto, tiveram sua prioridade calculada de acordo com os valores das heurísticas do mesmo subgrupo (vide seção 4.2 para mais detalhes).

Heurística	Média Scores	Prioridade
HA.1.1	5.8	1
HA.2.1	5.82	1
HA.2.2	5.56	2
HA.2.3	6	1
HA.2.4	5.62*	1
HA.2.5	5.33	2
HA.2.6	5	3
HA.2.7	5.62*	1
HA.2.8	6	1
HA.3.1	5.2	2
HA.3.2	3.77	5
HA.3.3	4.49*	3
HA.4.1	5.54	2
HA.4.2	5.5	2
HA.4.3	5.48	2
HA.5.1	5.88	1
HA.5.2	6	1
HA.5.3	6	1
HA.5.4	5.96*	1
HA.5.5	5.96*	1
HA.5.6	5.96*	1
HA.6.1	5.53	2
HA.7.1	5.35	2
HA.7.2	4.78	3
HA.7.3	5.4	2
HA.7.4	5.13	2
HA.7.5	5.91	1
HA.7.6	5.35	2
HA.7.7	5.21	2
HA.7.8	6	1
HA.7.9	5	3
HA.7.10	5.2	2
HA.7.11	5.1	2
HA.7.12	5.53	2
HA.7.13	5.29	2

Heurística	Média Scores	Prioridade
HA.8.1	4	5
HA.9.1	6	1
HA.10.1	4.48	4
HA.11.1	-	3
HB.1.1	6	1
HB.1.2	5,16*	2
HB.1.3	5.5	2
HB.1.4	3.93	5
HB.1.5	6	1
HB.1.6	6	1
HB.1.7	3.5	6
HB.2.1	5.48	2
HB.2.2	5.5	2
HB.2.3	5.49*	2
HB.3.1	4.96	3
HB.3.2	3.4	6
HB.3.3	4.83	3
HB.3.4	4.4*	4
HB.3.5	4.4*	4
HB.3.6	4.4*	4
HB.4.1	4	5
HB.4.2	3.82*	5
HB.4.3	3.63	5
HB.5.1	4.73	3
HB.5.2	4.73*	3
HB.7.1	4.33	4
HB.7.2	4.7*	3
HB.7.3	5.04	3
HB.7.4	4.72	3
HB.9.1	4.88	3
HB.9.2	4.44*	4
HB.9.3	4	5
HB.12.1	-	3
HB.12.2	-	3
HB.13.1	4.25*	4

Heurística	Média Scores	Prioridade
HB.13.2	4.25	4
HB.13.3	1*	6
HB.13.4	1	6
HB.14.1	4.88	3
HB.14.2	5.24	2
HB.14.3	5	3
HB.15.1	5.33	2
HB.15.2	3	6
HB.15.3	4.17*	4
HB.16.1	4.56	3
HB.16.2	3.5	6
HB.17.1	-	3
HB.18.1	4.83	3
HB.18.2	6	1