Pruning UNL Texts for Summarizing Purposes

Camilla Brandel Martins
Lucia Helena Machado Rino
Departamento de Computação
Universidade Federal de São Carlos
Rodovia Washington Luis (SP-310) km 235, São Carlos, SP, Brazil
camilla@dc.ufscar.br, lucia@dc.ufscar.br

Abstract

This paper presents a summarization model based on the Universal Networking Language (UNL), which is a conceptual language for representing texts sentence by sentence, using semantic binary relations that are claimed to convey all the information of the corresponding sentence in natural language. Our summarization model is based on heuristics for pruning sentences, focusing on UNL binary relations.

1 Introduction

While conventional text summarization (TS) tools focus on Natural Language (NL) texts, the TS model we present deals with UNL texts, i.e., texts encoded in the *Universal Networking Language* (Uchida, 2000). Heuristics for pruning texts are defined based on UNL, which has been developed under the UNL Project. This project architecture (full description available in http://www.unl.ias.unu.edu) is interesting for many reasons. First, we can make use of the UNL Project generic tools: EnCo, a NL-UNL text enconverter, and DeCo (Uchida, 1997), a UNL-NL text deconverter. By plugging into such generic tools specific resources for, e.g., Brazilian Portuguese (BP), one can assess encodings/decodings of any text, having BP either as a source- or as a target-language. In particular, by plugging into a UNL System our TS tool, one can a) summarize texts written in any NL and b) generate both, the full text and its corresponding summary in any of the NLs considered in the UNL Project, provided that the corresponding UNL encodings have been previously produced and that all the decoding resources for the NL under focus have been provided. Such a scenario is depicted in Figure 1; blue boxes show the generic UNL Decoding System and yellow boxes illustrate our TS proposal and the way it will be plugged into the UNL environment.

Second, by summarizing UNL texts, our TS model is independent from any NL. Besides reducing complexity, this renders the system more controllable, due to the limited semantic setting provided by UNL. This leads to the third reason: making TS UNL-based, difficulties concerning either NL interpretation or the generation of texts in any NL are left to those modules, rendering our UNL TS model focused entirely upon the conceptual level. Those difficulties, in turn, have been already handled in the UNL Project itself. So, by the time our TS model is fully implemented, solving that kind of difficulty will benefit from the UNL Project progress.

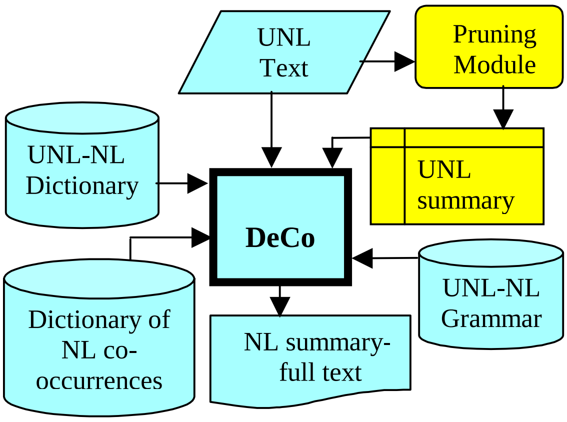

Figure 1: UNL-NL decoding system added by our Pruning Module

Fourth, by being UNL-based, our TS model may be applied to a wide range of Web demands, the most interesting for us being summarizing texts and retrieving information from the Web, provided that the corresponding texts are previously encoded in UNL. Considering the UNL Project goals, making our TS tool available on the Web is surely feasible.

Finally, by making available an extra tool for the UNL System, it will be possible to simultaneously promote summarization and translation of source texts and, thus, to improve communication.

We present in Section 2 the main features of UNL texts and in Section 3 our TS Model, hereafter called UNL Summarizer, or **UNLSumm**. Heuristics for pruning UNL texts are illustrated, followed by a discussion on their validity. So far, results have been produced by using an automatic pruning prototype, while the resulting UNL summaries are still hand-decoded. We especially address original texts in BP and English, aiming at automatically decoding the corresponding UNL texts into BP. Customizing the UNL generic tools to BP has been carried out at NILC, a Brazilian Computational Linguistics Interinstitutional Center (http://nilc.icmc.sc.usp.br). So far, the UNL-BP dictionary and grammar have been successfully applied to decode UNL texts of significant complexity, encoded by different UNL teams. Similarly, our architecture showed in Figure 1, customized to BP, will help us to properly assess UNLSumm results.

2 UNL Texts

UNL texts convey independent sentences, which are the most complex units dealt with by the UNL main processing tools. Intersentential automatic processing is not considered in the UNL Project. However problematic this may be for NLP, it is not our aim here to discuss the very original assumptions under the project, which have been thoroughly adopted[1].

Each UNL sentence conveys a set of binary relations (BRs), represented in the format *relation-label(UW1,UW2)* (Uchida, 2000). *Relation labels*, or RLs, signal semantic relations and are expressed by means of mnemonics (e.g., 'agt' for *agent*, 'mod' for *modifier*, 'obj' for *object*, 'ptn' for *partner*, or 'met' for *method*). These refer back to other well-known conceptual languages (e.g., Fillmore, 1968; Jackendoff, 1990; Dorr, 1992), as discussed in (Martins et al., 2000), and currently amount to 41 RLs. They basically convey semantic relationships between concepts corresponding to meaningful sentence components, which are called *Universal Words*, or UWs. These may be single UNL terms, such as "book", "run", or "John", or: a) to express complex, ontological structures that convey refined meanings of a general concept (e.g., "book(icl>room)", 'icl' indicating a hyperonymic relation between "book" and "room", yielding the meaning 'to book a room'); b) to convey morphosyntactic information, when *Attribute Labels*, or ALs, are used. These are added to generic UWs by means of the symbol '@', as in "run.@past" (in this case, to convey tense information).

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

Example 1: S1 UNL encoding

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

Example 2: S2 UNL encoding

UNL addresses only literal meaning, i.e., it is solely based upon the source text surface structures. Examples 1 and 2 below show complex UNL texts[2], respectively corresponding to '*Uruguay took the lead versus Argentina through a penalty.*' **[S1]** and '*Uruguay took the lead versus Argentina through a goal scored by Pablo Dorado.*' **[S2]**.

The intertwined organization of BRs to represent intrasentential dependencies is an important feature in UNLSumm, as we will see in next section.

3 The UNLSumm Model

By addressing only intrasentential dependencies, UNLSumm is constrained in the following ways: a) when summarized, multisentential UNL encodings result in juxtaposed UNL sentences exerts; b) there is no full sentence exclusion, only suppression of some material from each UNL sentence. (a) relates to the classic extraction method of AS (Baxendale, 1958); (b) indicates the main bottleneck of this proposal: we actually summarize limited amounts of information, excluding UNL units or segments that have been considered non-essential[3] in a certain context. Although constraints (a) and (b) may indicate a severe AS system limitation, they render our UNL-based approach interesting, if we consider the original UNL Project assumptions and the fact that we are summarizing *conceptual structures.* In this case, 1) we only exclude BRs that are indeed superfluous and 2) the resulting UNL summary does not present extraction method problems, e.g., failures in thematic progression. Also, enclosing language decisions in the linguistic module and, thus, focusing on the conceptual level, tends to improve UNLSumm efficiency and robustness.

3.1 Defining pruning heuristics

Our pruning heuristics are based on the RLs: in denoting semantic relationships, these may indicate non-essential sentence units. For example, object attributes, signaled by the 'aoj' RL, usually correspond to adjuncts at the surface (e.g., noun+adverbial phrase; noun+adjective in BP, as observed by Sossolote et al., 1997). Since adjuncts may be non-important (cf. Boguraev and Kennedy, 1997), this results in a heuristic that excludes the BR labeled by 'aoj' from the corresponding UNL text.

Before properly creating the heuristics, we specified a trial set of RLs, based on previous works. For example, Rino and Scott (1994) suggest that details and examples (signaled in UNL by the RLs 'aoj', 'mod' and others) are non-essential; Robin (1994) suggests that locative or temporal information (signaled in UNL respectively by RLs 'ppl' or 'plc', and 'tim') may be floating in a sentence and, thus, may be non-essential. We also applied an *ad-hoc* analysis based on our native speakers subjective judgement, adding some other RLs to our trial set. Table 1 shows some of those.

[1] For a discussion on the UNL Project limitations and contributions, see (Martins et al., 1998; 2000).

[2] In any example shown, UWs have been simplified.

[3] Non-essential, irrelevant, or superfluous terms will all be considered interchangeable here.

To improve reliability and complete our RL set, we carried out analyses on two corpora of UNL texts: the UN Charter Corpus, composed of UNL encodings of 12 chapters of the ONU Statute, amounting to 255 UNL sentences, or 3182 BRs; (genre: legal; domain: laws) and the Booklet Corpus (Uchida, 1996), with a total of 101 UNL sentences, or 1230 BRs (genre: narrative; domain: open). Both corpora have been encoded by several UNL teams, which comprise native speakers of the NLs considered in the UNL Project. This feature played an important role during corpora analysis, since encoding diversity brings about several problems that are very often difficult to tackle, as pinpointed by Martins et al. (2000). However, although troublesome, encoding diversity was useful for devising heuristics widely applicable to UNL texts whose encoding originates from any NL. When controversies were found, our heuristics were designed to consider the majority of the cases, for a more robust pruning set.

In pursuing the above, we compared UNL texts with their corresponding source texts, by identifying in these the non-essential segments and, thus, finding the corresponding BRs in the UNL texts. This yielded a second set of RLs. Then, we determined their frequency distribution in order to verify in the full corpus their rate of non-essentiality. For example, for the UN Charter Corpus, Table 1 shows the *irrelevant frequency*, or IR_Freq, for each RL, which has been calculated according to the following expression[4]:

$$\textbf{IR\_Freq} = \frac{\textbf{N\_Irrel x 100}}{\textbf{N\_Occur}}$$

RLs	N_Occur	N_Irrel	IR_Freq
ppl	8	2	25%
tim	16	4	25%
met	11	2	18.18%
mod	894	68	7.6%
pur	54	3	5.56%
aoj	96	5	5.21%

Table 1: Distribution of RLs signaling non-essential information

RLs with high IR-Freq indicate how significant it is to consider them for pruning the corresponding BR. Comparing BRs involving ´tim´ and ´mod´, e.g., pruning based on the former may be preferred, if we are to choose only one way of pruning. Although non-essentiality criteria may be drawn from such a distribution, due to the sparse results derived from the corpora, the main metrics for pruning have been kept on fundamental grounds, as previously specified in our trial set. Table 2 illustrates phrases (through their syntactic functions) that are potential candidates for exclusion, which lead to their corresponding UNL encodings (i.e., semantic roles) and, thus, to the RLs that must be focused upon for pruning. In defining the whole set of RLs, the corresponding BRs surroundings played an important role, as we will shortly see. So far, 58 heuristics have been specified, involving a total of 18 RLs.

3.2 Basic pruning heuristics

The 58 heuristics were divided in two main groups. *Group A* includes heuristics targeting *single pruning*, i.e., the exclusion of independent BRs, one by one. *Group B* includes complex heuristics, targeting *chained pruning*, i.e., the ones that trigger the exclusion of a group of interconnected BRs. S1, for example, can be pruned by applying the Group A H1 (Figure 2), yielding the UNL summary showed in Example 3, whose corresponding decoding may be **'***Uruguay took the lead versus Argentina.'* [**S3**].

Exclude met(a,b) from sentence S
if UW b ∉ others BRs in S

Figure 2: A Group A heuristic (H1)

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

Example 3: A UNL summary for S1 and S2

Applying the Group B heuristics H2 (Figure 3), for example, to S2, results in the same UNL summary, which, in turn, may be again decoded into S3. It is worth stressing that although the summaries, in this example, are the same for different source sentences, they have been generated through diverse pruning strategies. Additionally, H1 would not be applicable to S2, as H2 would not to S1, since their corresponding conditions are not verifiable for each sentence. This makes evident that heuristics application is not interchangeable.

Exclude met(a,b)
obj(c,b)
Agt(c,d) from S
If UWs b,c and d ∉ others BRs in S

Figure 3: A Group B heuristic (H2)

3.3 Single vs. chained pruning

A serious problem that we face in proposing single and/or chained pruning (thus, Group A and Group B heuristics) is how to decide on the heuristics application order, when considering both types of pruning. In principle, Group A ones may always be applied. However, the inter-dependency between BRs could trigger dangling BRs after pruning. This would be the case of the S2 UNL encoding (Example

[4] N_Irrel: number of occurrences of the specific RL in superfluous UNL spans; N_Occur: number of occurrences of the RL in the whole corpus.

RLs	Semantic roles	Syntactic functions	UNL Binary Relation	NL sentence
aoj	aoj(Attribute,Concept)	Complement	aoj(recent,breakthrough)	Recent breakthrough lead to advances.
met	met(Action,Method)	Adjunct	met(solve,algorithm)	He solved the problem using an algorithm
mod	mod(Concept,Modifier)	Complement	mod(language,native)	English is the native language in Australia
ppl	plc(Action,Place)	Adjunct	plc(write,home)	She studies at home
pur	pur(Action,Purpose)	Adjunct	pur(go,eat)	She went home to eat
tim	tim(Action,Time)	Adjunct	tim(study,tonight)	She will study tonight

Table 2 – RLs and corresponding syntactic realizations indicating non-essential information

2), if we allowed the application of H1, resulting in a UNL summary composed of BRs 1-3, and 5 and 6, these being responsible for dangling phrases after decoding. This is due to two main reasons, namely: a) Group A and Group B do not convey heuristics whose premises focus upon distinct RLs; b) often some RLs trigger the exclusion of more than one BR, as it is illustrated by H2. The problem is thus to identify when chained pruning must be given priority upon single pruning. If we consider a UNL graph representation[5], this corresponds to looking for *independent* subgraphs, i.e., those that are linked to the main one only by means of a single node, or a specific UW. 'take.@past' in Example 2 is a good illustration of such a case: it links the UNL graph involving 1-3 BRs with the one involving 4-6 BRs by means of the RL 'met'. Since BRs 5 and 6 do not convey any UW that appears in the main graph (1-3), UNL spans 4-6 may be excluded, safely guaranteeing that UNL summary will be coherent.

Besides deciding on single vs. chained pruning, a priority strategy between heuristics applicable to a unique sentence has to be pursued, given that multiple choices are also feasible. For example, we could apply even the less significant heuristics related to the RLs shown in Table 1 for one sentence, however wasteful such a strategy might be. To prevent such an effort, we pursued a better strategy, as discussed below.

3.4 Precision of Heuristics

Priority-based UNL texts pruning allows to look for the most reliable heuristics in a certain pruning time step. This is important because some RLs encode different syntactic functions. The RL 'mod' (see Table 2), e.g., may represent a non-essential adjective or a restrictive clause whose suppression could imply a serious meaning problem. To overcome the above problems, our model for heuristics application has been based on precision measures. For both corpora used in modeling, the precision of each heuristic is calculated as (Krenn and Samuelsson, 1997):

$$\text{Precision(H)} = \frac{\text{Sat\_Num}}{\text{Total\_Num}}$$

where **Sat_Num**: number of applications of H leading to satisfactory results;
Unsat_Num = number of applications of H leading to unsatisfactory results;
Total_Num: Satisf-Num + Unsat-Num

Precision of a heuristic H is, thus, a measure to determine the probability of successfully applying H to a UNL sentence and obtaining a reasonable summary. By reasonable we mean a summary that preserves the main idea, or the *gist*, of the source UNL text and guarantees coherence, or *textuality* (cf. Rino, 1996).

Heur.	Key RLs	Total-Num	Sat-Num	Precision	Priority value
H6	man	20	18	0,9	1
H12	aoj	17	13	0,76	2
H16	tim	9	9	1	1
H28	mod	3	2	0,67	3

Table 3: Examples of heuristics priorities

Precision	Priority Value
85%-100%	1
75%-84%	2
65%-74%	3
50%-64%	4
31%-49%	5

Table 4: Heuristics priority intervals

Table 3 illustrates some of the priority values used in UNLSumm, which are drawn on precision intervals, as shown in Table 4[6]. Such intervals have so far been determined in an *ad-hoc* way, but comprehensive experiments will be carried out in the future, to corroborate such a priority-based strategy.

4 Applying pruning heuristics

To illustrate how UNLSumm works, let us take the UNL text *"Fall Cushioning"*, given below in its

[5] See (Uchida, 1997) for details on UNL graphs.

[6] Precision rates below 30% have not been considered.

English version[7]. Its full corresponding UNL text contains 100 BRs, which are given as input to UNLSumm. By adopting the above priority strategy, 13 heuristics have been applied: 4 Group A and 9 Group B ones. The resulting UNL summary conveys 59 BRs, representing a compression rate of 41%. An example UNL-English decoding of such a UNL summary is presented as the *"Fall Cushioning summary"* below. This has been produced by excluding those phrases of each source sentence that correspond to the pruned BRs. So far, decoding of UNL summaries has been handmade. We could not even produce the *"Fall Cushioning summary"* automatically, since we do not have the English language resources to make DeCo operational (see Figure 1). Alternatively, we could apply our BP environment to decode the very same UNL summary onto a BP text. Automation of such a process will be shortly pursued, as described in Section 1.

"Fall Cushioning" source text:
[S1] Helicopters are very convenient for dropping freight by parachute, but this system has its problems. [S2] Somehow the landing impact has to be cushioned to give a soft landing. [S3] The movement to be absorbed depends on the weight and the speed at which the charge falls.
[S4] Unfortunately most normal spring systems bounce the load as it lands, sometimes turning it over. [S5] To avoid this, Bertin, developer of the aerotrain, has come up with an air-cushion system which assures a safe and soft landing. [S6] It comprises a platform on which the freight is loaded with, underneath, a series of balloons supported by air cushions. [S7] These are fed from compressed air cylinders equipped with an altimeter valve which opens when the load is just over six feet from the ground. [S8] The platform then becomes a hovercraft, with the balloons reducing the deceleration as it touches down.
[S9] Trials have been carried out with freight dropping at rates from 19 feet to 42 feet per second. [S10] The charge weighed about one and half tons, but the system can handle up to eight tons. [S11] At low altitudes freight can be dropped without a parachute.

"Fall Cushioning" summary:
[S1] Helicopters are convenient for dropping freight by parachute. [S2] The landing impact has to be cushioned. [S3] The movement to be absorbed depends on the weight and the speed.

[S4] Most normal spring systems bounce the load, sometimes turning it over. [S5] Bertin has come up with an air-cushion system which assures a safe and soft landing. [S6] It comprises a platform on which the freight is loaded with a series of balloons supported by air cushions. [S7] These are fed from compressed air cylinders. [S8] The platform then becomes a hovercraft, with the balloons reducing the deceleration.
[S9] Trials have been carried out with freight-dropping at rates from 19 feet to 42 feet per second. [S10] The charge weighed about one and half tons. [S11] At low altitudes freight can be dropped without a parachute.

5 Evaluation and results

We applied our UNL pruning heuristics to two extra corpora of UNL texts: the UNU Corpus (source texts in English), composed of bureaucratic texts encoded by diverse UNL teams, and the THESES Corpus, composed of postgraduate-level Computer Science monographs and articles (source texts in BP), which have been manually encoded by ourselves. The former amounts to 23 UNL sentences, corresponding to 294 BRs; the latter, to 62 UNL sentences, or 632 BRs. For assessment, the full repository of heuristics has been indiscriminately applied to both corpora, with no distinction of genre or source language.

So far, pruning has been fully automated on a prototype basis. For both testing corpora, our pruning heuristics have been repeatedly applied to every UNL sentence. In all, only 44 heuristics have been chosen, 23 from Group A and 21 from Group B. Average compression rates of 44,58% and 40,9% were observed in the UNL summaries of UNU and THESES Corpora, respectively. The corresponding UNL summaries were hand-decoded onto English and BP, according to their respective source NLs. In comparing the resulting summaries with their corresponding source texts, we could certify that they fulfill the main formerly posed AS constraints, namely: gist preservation and textuality, as defined by Rino (1996). Also, our heuristics apply similarly to both corpora, no matter how they have been encoded. It is worth noting that NL-UNL troublesome encodings have been observed throughout the development of the UNL Project, and this is a serious issue that we do not focus upon in our UNLSumm proposal.

On the one hand, results obtained so far are not representative, given that compression rates are relatively low and the UNLSumm assessment has not been profound or comprehensive enough. On the other hand, by dealing solely with intrasentential relationships, we should not expect summaries to be as good as they would be if generated by another kind of automatic summarizer. In spite of such features, UNLSumm can still be useful, if we consider the UNL scenario.

6 Conclusions

We presented in this paper the UNLSumm model, aiming at pruning UNL texts by means of heuristics that focus upon non-essential UNL binary relations.

[7] Extracted from (Hoey, 1983, p. 68) and UNL hand-encoded. Text units have been segmented for referring purposes and are correspondingly delimited in the UNL text shown. Paragraphing has been kept unchanged.

A set of Relation Labels has been delineated, which signal the BRs of interest, on the basis of syntactic and semantic features. Although we have pursued statistical procedures to verify how non-essentiality could be used to define the heuristics, the results were inconclusive. However, they may be used, in the future, in more comprehensive evaluations.

We are about to plug UNLSumm into the generic UNL System, according to the architecture shown in Figure 1. In doing so, besides automatically pruning UNL texts, we will be able to use DeCo to automatically produce both, the source text and its corresponding summaries in Brazilian Portuguese.

After generating in such a way a considerable number of summaries, we will be capable of carrying out more systematic evaluations, to assess both our set of heuristics and the quality of the proposed summaries. For example, snap judgements (White et al., 2000) may be used in order to assess both gist preservation and textuality at the surface level; conten-based measures (Donaway et al., 2000) may be used for an evaluation that does not rely in human judges; UNL summaries may be similarly compared to UNL source texts by UNL experts; and also the classical precision and recall measures may be obtained, e.g., as in Oka and Ueda (2000). Only then it will be possible to consider other, more robust and comprehensive, types of assessment.

Acknowledgments

This work is supported by CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

References

Baxendale, P. B. (1958). Machine-made index for technical literature – an experiment. *IBM Journal of Research and Development*, 2, pp. 354-361.

Boguraev, B. and Kennedy, C. (1997). Salience-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.

Donaway, R. L., Drummey, K. W. and Mather, L. A. (2000). A Comparison of Rankings Produced by Summarization Evaluation Measures. In: *Proceedings of the ANLP/NAACL Automatic Summarization Workshop*, pp. 69-78. Seattle, Washington, USA.

Dorr, B. J. (1992). The Use of Lexical Semantics in Interlingual Machine Translation. *Journal of Machine Translation*, 7:3, pp. 135-193.

Fillmore, C. (1968). The case for case. In Bach, E. and Harms, R.T. (orgs.), *Universals in linguistic theory*, pp. 1-88. Rinehard and Winston, New York.

Hoey, M. (1983). *On the Surface of Discourse.* George Allen & Unwin (Publishers).

Jackendoff, R. (1990). *Semantic Structures*. MIT Press, Cambridge, MA.

Krenn, B., and Samuelsson, C. (1997). *The Linguist's Guide to Statistics. Don't Panic.* A compendium for a course in Statistical Approaches in Computational Linguistics. Saarland University, Germany. Available *at http://www.coli.uni-sb.de/~krenn/edu.html*

Martins, R.T.; Rino, L.H.M.; Nunes, M.G.V.; Oliveira Jr., O.N. (1998). Can the syntactic realization be detached from the syntactic analysis during generation of natural language sentences? In*: III Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR' 98)*. Porto Alegre – RS, Brazil.

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, USA.

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.

Rino, L. H. M. (1996). *Modelagem de Discurso para o Tratamento da Concisão e Preservação da Idéia Central na Geração de Textos*. Tese de Doutorado. IFSC-Usp. São Carlos – SP, Brazil.

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. (1996). *UNL: Universal Networking Language - An Electronic Language for Communication, Understanding, and Collaboration*. UNU/IAS/UNL Center. Tokyo, Japan.

Uchida, H. (1997). *DeConverter Specification, Version 1.0.* Tech. Rep. UNL-TR1997-010, UNU/IAS/UNL Center, Tokyo, Japan.

Uchida, H. (2000). *Universal Networking Language: An Electronic Language for Communication, Understanding and Collaboration*. UNL Center, IAS/UNU, Tokyo, Japan (available on the site www.unl.ias.unu.edu).

White, J., Doyon, J. and Talbott, S. (2000). Task tolerance of MT output in Integrated Text Processes. *Proceedings of the Embedded MT Systems Workshop*, pp. 9-16 (NAACL-ANLP 2000 Workshop). Seattle, WA, USA.