

Summarization of specialized discourse

The case of medical articles in Spanish

Iria da Cunha, Leo Wanner and Teresa Cabré

In this article, we present the current state of our work on a linguistically-motivated model for automatic summarization of medical articles in Spanish. The model takes into account the results of an empirical study which reveals that, on the one hand, domain-specific summarization criteria can often be derived from the summaries of domain specialists, and, on the other hand, adequate summarization strategies must be multidimensional, i.e., cover various types of linguistic clues. We take into account the textual, lexical, discursive, syntactic and communicative dimensions. This is novel in the field of summarization. The experiments carried out so far indicate that our model is suitable to provide high quality summarizations.

Keywords: summarization, extraction, medical article, linguistic analysis, Spanish, specialized discourse

1. Introduction

Already in 1970, some authors identified the “information overload” faced by modern society (Toffler 1970). This overload dramatically increased since the Internet became our daily companion. It is *de facto* impossible to browse even a reasonably small share of the material with the goal to filter out information that is relevant to us, let alone to read it. Professionals in research intensive domains are especially affected. They experience a real predicament when they need to know the state of the art in their area in order to account for their tasks. On the other hand, they often simply do not have the time for extensive reading. Automatic summarization is likely to be of great use for these professionals since it facilitates the “distillation” of the essence of the information material, allowing thus the reader to assess the relevance of the material and to grasp its meaning in the shortest possible time. However, it is important to note that techniques used

for automatic summarization of general discourse are not necessarily suitable for summarization of specialized discourse. Recent works in the field show that high quality automatic summarization of specialized discourse must take into account the characteristics of the discourse in question (Teufel and Moens 2002; Johnson et al. 2002; Farzindar et al. 2004). These characteristics concern both the way the content is presented and the way material is to be summarized. Journals in certain specialized discourses (as, e.g., in the medical discourse), tend to have detailed guidelines for how to write a summary and what it should contain.

In this article, we address the problem of automatic summarization of medical articles in Spanish from the extraction angle.¹ For medical professionals, articles are an important source of information. Several types of articles are being published: *Original Articles*, *Short Original Articles*, *Revisions*, *Clinical Notes*, etc. We focus on Original Articles and Short Original Articles (henceforth, referred to as “medical articles”).

Given that existing proposals for automatic summarization of medical articles tend to be based on general purpose summarization strategies (cf. Afantenos et al. 2005 for a state of the art study), our goal has been twofold: firstly, to verify whether the summaries of medical articles reveal different characteristics than summaries of general purpose discourse, and if yes, to identify these characteristics; secondly, to develop a summarization model that takes into account the criteria that have been identified before.

To assess the degree of idiosyncrasy of medical article summaries and to evaluate whether the summaries by domain specialists (i.e., medical doctors) are, on the one hand, similar among each other and, on the other hand, significantly divergent from the summaries by specialists in general discourse analysis and writing (i.e., linguists), we undertook an empirical study. The results of the study reveal that, indeed, adequate summaries of medical articles show major idiosyncrasies which must be taken into account during automatic summarization.

In the next stage of our work, we identified the criteria that need to be taken into account in summarization oriented towards medical articles (cf. also da Cunha and Wanner 2005 for a short preliminary presentation). In general, we found that the summarization model must

- Take into account all domain-specific summarization criteria (those that are available explicitly in terms of author guidelines and those that can be derived studying human written summaries);
- Consider the whole range of linguistic dimensions reflected in an article: textual, discursive, communicative,² lexical, and syntactic.

In the course of the final stage, we developed a multidimensional summarization model that integrates all of the above criteria — in contrast to most of both gen-

eral purpose and domain-specific models which focus on one or two linguistic dimensions only. The evaluation of the experiments carried out so far shows that automatic summaries produced by the application of our model are comparable with the author written summaries. We are confident that the design principles underlying our model can be ported to other domains.

The remainder of the article is structured as follows. In the next section (Section 2), we review the state of affairs in automatic summarization in general discourse, in specialized discourse in general, and in the medical domain in particular. Section 3 analyzes the predefined text structure of a medical article. Section 4 describes first the empirical study that we carried out to verify the distinctiveness of medical article summaries and to identify the characteristics that have to be taken into account, and draws then the conclusions for our model. In Section 5, we present a description of our model. In Section 6, the experiment that validates the model is described. Section 7, finally, recapitulates the essentials of the article and suggests some lines of future work.

2. Automatic summarization

Although the summarization of specialized discourse received a certain attention already since the very beginnings of the automatic summarization research (e.g., Luhn 1958), the focus in the field over the years has been on general discourse. Thus, it is not surprising that strategies developed for general discourse have often also been applied to specialized discourse and, in particular, also to summarization of medical articles. In this section, we consider only extraction-oriented works.³

2.1 Summarization of general discourse

Extraction-oriented strategies of summarization may be classified in terms of surface-oriented strategies, medium-level strategies and deep strategies (Radev et al. 2002). Many of them are “mono-dimensional” in that they draw upon one specific type of criteria to identify summary relevant text passages.

Surface-oriented strategies tend not to use any (or use only rudimentary) linguistic information of the textual material to be summarized. Well-known are techniques that use positional or cue phrase criteria. The positional criteria specify where — at the beginning of the first or last section, following specific section headings, etc. — text chunks that are relevant to the summary are located (see, e.g., Brandow et al. 1995; Lin and Hovy 1997). Cue phrases such as “It is important to note that ...”, “To conclude ...” etc. are supposed to introduce summary relevant text chunks (Edmundson 1969).

Of similar level of abstraction are techniques that use statistical criteria of varying complexity: simple frequency of terms (Edmunson 1969), *Bayesian* models (Kupiek et al. 1995), *Maximal Marginal Relevance* (Goldstein et al. 1999), etc.

Medium-level strategies make use of certain linguistic, mostly lexical or surface-syntactic, information. *Lexical chains* (sequences of words between which lexico-semantic relations hold) received particular attention (Barzilay and Elhadad 1997; Silber and McCoy 2000).

Deep strategies draw upon more profound linguistic information such as discourse structure. Discourse structure (modelled, first of all, in terms of the *Rhetorical Structure Theory* (RST), Mann and Thompson 1988) proved to be of special importance since it ensures the coherence of the summary if the latter is obtained by cutting off selected branches of the discourse tree (Ono et al. 1994; Marcu 2000).

2.2 Summarization of specialized discourse

It is well-known that summarization must be considered in the light of the text genre to be summarized. For instance, news summaries must contain the global event settings and the most relevant details; novel summaries must contain the general outline of the plots, without revealing details and the outcome, etc. As Kaplan et al. (1994), who research abstracts in the area of applied linguistics, point out, summaries must be differentiated with respect to their purpose, targeted audience and extension. The same argumentation is adopted by several other authors. Thus, Teufel and Moens (2002) argue that the summary of a scientific article must primarily capture the novelty of the work (given that the purpose of a scientific article is the presentation of new scientific research), while a general purpose summary is usually supposed to present a short overview of the content of the material in question. Ciapuscio (1998) argues along similar lines that the summary of a scientific submission is written with the goal to convince the reviewers that the submission is novel and should thus be accepted for presentation or publication.

Despite this cognizance, the main focus of the research on automatic summarization in the past has been on general discourse — with a few exceptions such as the early experiments by Luhn (1958), who used term frequency to determine summary relevant sentences in technical material, and Pollock and Zamora (1975)'s work, which presents an extraction algorithm for automatic summarization of texts on chemistry.

Starting from the nineties, an increasing number of researchers focused on the automatic summarization of specialized discourse (cf., among others, Paice 1990; Riloff 1993; Lehman 1995; McKeown and Radev 1995; Abracos and Lopes 1997; Saggion and Lapalme 2000) — although often using the same strategies as for the summarization of general discourse. Especially summarization of news had been

an issue. For instance, Abracos and Lopes (1997) use statistics to select the most important content of newspaper articles. McKeown and Radev (1995) do multidocument abstracting of news statements on specific themes (such as terrorist attacks) using knowledge templates known from Information Extraction.

Some recent works on specialized discourse in fact do address the specifics of the genre in question. Thus, as already mentioned above, Teufel and Moens (2002) suggest a summarization strategy of scientific articles (more precisely, contributions to conferences on computational linguistics) that is based on the rhetorical status of each declaration included in the article in question. The material to be included in the summary is selected taking into account that the summary of a scientific article must highlight the novelty of the described work and contrast it to the state of the art. Teufel and Moens present an algorithm based on annotated training material that first classifies the material to be summarized in terms of seven rhetorical categories and chooses then the content to be included in the summary in accordance with these categories.

Farzindar et al. (2004) address the problem of automatic summarization of specialized juridical documentation. Their goal is to help experts of this area to determine the key ideas of a judgment in order to find other documents that could be relevant. Farzindar et al. use the text structure to automatically construct a coherent summary.⁴ More precisely, they construct a summary in four phases: (1) text structure segmentation that identifies the five topics of the document structure: Decision Data, Introduction, Context, Juridical Reasoning and Conclusion; (2) elimination of summary irrelevant information (as, e.g., citations of legal articles); (3) selection of relevant chunks from the four last sections using specific linguistic markers and textual position criteria for each of these topics; and, finally, (4) aggregation of the selected chunks applying a length limit.

2.3 Automatic summarization of medical documentation

As pointed out by Afantenos et al. (2005), automatic summarization of medical material became recently a very prominent research topic. Both extraction and abstraction paradigms are being followed. Consider Kan et al. (2001), Damianos et al. (2002), Lenci et al. (2002), Johnson et al. (2002) for proposals within the extracting paradigm and Gaizauskas et al. (2001), and Kan (2003) for proposals within the abstracting paradigm.

As already in the case of proposals presented in the previous section, most of the proposed techniques are general purpose summarization techniques adapted to medical articles. Typical criteria drawn upon by extraction techniques are *surface* criteria:⁵ *text structure*, *cue phrases*, *sentence positioning* and *named entities*. Some techniques combine surface-guided extraction with clustering algorithms

used to identify key features for semantically related articles. The summary is then built with sentences containing these key features (e.g., Johnson et al. 2002). None of these techniques uses, to our knowledge, more elaborate linguistic criteria such as syntactic dependency relations, communicative structure or discourse structure.

3. Medical articles in Spanish: Their structure and content distribution

Before we set out to study how medical articles in Spanish are most adequately summarized, let us analyze the default structure of a Spanish medical article and the distribution of the content across the article. In our analysis, we follow the guidelines of *Medicina Clínica*, which is the most widely read Spanish medical journal.

As already mentioned in the Introduction, we focus on two types of articles: Original Article and Short Original Article. *Medicina Clínica* requests authors of both types of articles to adopt the following text structure: *Title, Spanish Abstract, Spanish Keywords, English Abstract, English Keywords, Introduction, Patients and methods, Results, Discussion, Acknowledgments* (optional) and *Bibliography*. The core structure is thus the so-called “IMRD structure” (Swales 1990),⁶ which is sup-

Introduction: must be brief, providing only the information necessary for the comprehension of the following sections. It must not contain tables or figures, but it must include clearly stated goal(s) of the work. If the authors claim to publish a previously unnoticed observation, they must outline in the Introduction the method of the bibliographic search for the state of affairs, the used keywords, the years of coverage and the date of the last search.

Patients and methods: must contain the location where the experiment or the research has been carried out, the duration of the experiment, characteristics of studied series, used criterion for selection and the techniques used. Sufficient details must be provided such that a specific experiment can be repeated on the basis of this information. Statistical measures applied in the experiment must be described in detail.

Results: must contain the presentation of the findings obtained by the method without that any interpretation of these findings is added; if adequate, the findings must be complemented by tables and figures.

Discussion: must contain the opinion of the authors on the findings, and in particular: 1) meaning and practical application of the results; 2) considerations of a possible inconsistency of the methodology and reasons for which results are valid; 3) relation with similar publications and comparison between areas of agreement and disagreement, and 4) indications and guidelines for future research. The discussion must not contain any revision of the state of the art (as in the Introduction). Results of the work must not be repeated either.

Summary: must contain the essence of all four IMRD sections.

Figure 1. Author guidelines for articles in *Medicina Clínica*

posed to reflect the logical order common to scientific argumentation. For the composition of each section within the core structure, guidelines are given; cf. Figure 1 above.

4. How should a summary of a medical article in Spanish look like?

The guidelines in Figure 1 give a general idea of how a summary of a medical article should look like. Based on this idea, an abstract-oriented strategy similar to the one suggested by McKeown and Radev (1995) could be drafted. However, an abstract-oriented strategy would require deep linguistic analysis techniques, which are costly to develop and rather unreliable. Therefore, we opted for an extraction-oriented strategy.

For the design of a well-motivated extraction-oriented strategy for summarization of specialized discourse (here, Spanish medical articles), two related questions (of different levels of abstraction) seem central:

1. Do the summaries of Spanish medical articles follow the same lines as the summaries of general discourse?
2. If yes, can any of the existing strategies be applied to our task? If not, what are the relevant genre-specific summarization criteria?

Although we expected that the first question is to be answered in the negative, we needed this expectation to be confirmed by empirical figures. For this purpose, we carried out an experiment that indeed fully confirmed our expectation. In the course of this section, we first describe this experiment and then assess the outcomes of the experiment with respect to the construction of our summarization model (and, thus, with respect to the second question from above).

4.1 Experiment in summarizing medical articles

If the summarization of medical articles requires the consideration of the genre, this should be reflected within the summaries written by the authors or other professionals of the field — in contrast to the summaries of non-professionals or specialists of general discourse as, e.g., linguists. In other words, the summaries of professionals should be similar among each other and different from the summaries of non-professionals.

This is the basic idea underlying our experiment (for an initial outline of the experiment, see da Cunha and Llopis, in print).

4.1.1 *General Setting of the Experiment*

The experiment involved the participation of six medical doctors and six linguists. All 12 subjects were asked to provide summaries of 20 Spanish articles.⁷ For each article, we had an author written summary as reference.

The 12 subjects were given the 20 medical articles (without the abstracts of the authors) and asked to mark in each article the content they considered essential for inclusion into the summary. To control the length of the resulting summaries, the subjects were asked to observe a limit of the lines of the text they underline (20 lines were suggested). Figure 2 presents the summarization instructions.

Underline in each text the content chunks (sentences, phrases or fragments of text) that you consider *essential* for inclusion in the summary (do not consider titles and subtitles).
 For a summary of an Original Article, do **not underline more than** (approx.) 20 lines in total.
 For a summary of a Short Original Article, do **not underline more than** (approx.) 15 lines in total.

Figure 2. Instructions for carrying out the summaries in the experiment

The content chunks marked by any of the subjects in accordance with the instructions given in Figure 2 have been represented in a content selection table, assigning to each chunk a variable; in total 624 chunks have been identified. This table served as basis for the assessment of the similarity, respectively difference, between the summaries obtained from the subjects as well as the authors; cf. Table 1 for a fragment that illustrates the composition of the table. The first column contains the different content chunks. In the second column, each chunk is assigned a variable used for further processing. The remaining columns contain information on whether the corresponding subject has chosen the chunk in question for inclusion in the summary or not ('1' stands for inclusion, '0' for omission). 'A' stands for the author, 'M' for one of the medical doctors and 'L' for one of the linguists. The indices associated to 'M' and 'L' identify the corresponding medical doctors and linguists, respectively.⁸

4.1.2 *Quantitative Assessment of the Experiment Outcome*

To quantify the degree of similarity, respectively difference, between the summaries, we used the techniques of *Multidimensional Scaling* and *Unsupervised Clustering* as implemented in *Statgraphics*, which is a widely available off-the-shelf statistical program.

Multidimensional Scaling (MS) is a dimension reduction technique which maps a d -dimensional space onto a 2-dimensional space, attempting to find the most similar representation of the original d -dimensional cloud of points in a plane. In our case, d is the number of variables we work with (i.e., $d=624$). The procedure consists, in fact, of the representation of the studied items in the two

Table 1. A fragment of the table with content chunks chosen by at least one subject

content	V	A	M1	M2	M3	M4	M5	M6	L1	L2	L3	L4	L5	L6
Entre el 20 y el 80% de las visitas a los servicios de urgencias hospitalarios son inapropiadas.	X1-1	1	1	1	1	1	1	1	1	1	1	1	1	0
Se evaluaron las visitas a un servicio de urgencias hospitalarias mediante un Protocolo de Adecuación de Urgencias Hospitalarias, previamente validado.	X2-1	1	1	1	1	1	1	1	1	1	1	1	1	0
El 37,9% de las visitas fueron inapropiadas	X3-1	1	1	1	1	1	1	1	0	0	1	1	0	1
y más frecuentes en la población pediátrica.	X4-1	1	1	1	1	1	1	1	1	1	1	1	1	1
Los pacientes enviados por un médico, con traumatismos o proceso quirúrgico visitaron las urgencias más adecuadamente.	X5-1	1	1	1	1	1	1	1	1	1	1	0	1	1
Estudio descriptivo sobre una muestra aleatoria representativa de los 84.329 pacientes atendidos en urgencias durante el año 1999.	X7-1	0	1	1	1	1	0	0	0	0	0	0	0	0
...

first main dimensions of a *Principal Component Analysis*. These two main dimensions grasp a maximum of variability (of information). Thanks to that, reduction of dimensions is carried out minimizing the loss of information.

The outcome of the MS-procedure is presented in Figures 3 and 4.⁹ The first main dimension of the MS-model is plotted as the vertical (*y*) dimension, the second as the horizontal (*x*) dimension. Figure 3 shows the overall similarity between all summaries (i.e., subject summaries and authors' summaries), or, in other words, the similarity calculated over all 624 variables of the content selection table. According to the diagram in Figure 3, the 6 doctors and the authors of the articles select very similar content for inclusion into the summary, but rather different from the content selected by the linguists (the summaries by the medical doctors are circled). On the other hand, although several linguists coincide largely with respect to their selection criteria, the dispersion among the summaries of the linguists is larger.

Figure 4 shows how similar the subject summaries are to the summaries of the authors. For this latter presentation, MS thus examined to what extent the selection preferences of our subjects coincided with the choices of the authors, without

Figure 3. Overall similarity between the subject and author summaries

Figure 4. Contrast of the content of the authors' summaries to the content marked for inclusion into the summary by the 12 subjects

taking the coincidence between the selection preferences among the subjects into account. This second examination was carried out to verify, on the one hand, to what extent the subject summaries included content chunks also found in authors' summaries, and, on the other hand, to assess to what extent the subject summaries contained further chunks not available in the summaries of the authors (i.e., to what degree the length of the subject summaries exceeded the length of the author summaries).¹⁰ It demonstrates that not only the content chunks selected by the medical doctors tend to coincide, but that these chunks also tend to coincide with the chunks selected by the authors. The summaries of the linguists, on the other hand, again, show a rather different, dispersed pattern.

To confirm the outcomes of *Multidimensional Scaling*, we further assessed the similarity of the subject and author summaries by *Hierarchical Unsupervised Clustering*. *Unsupervised Clustering* (UC) is a technique that groups data in terms of a group hierarchy (instead of a flat classification) according to the distance between them — without using any type of external information to organize the groups: at each level of the hierarchy, pairs of groups that are sufficiently similar to each other form a generalized group (which is then again used as a group).

The resulting classification is shown in terms of dendograms in Figures 5 and 6. The dendograms display group hierarchies built from the 13 group nodes (1 author group + 6 medical professional groups + 6 linguist groups) in our experiment. The length of the branches of the dendograms (the vertical axes) displays the distance between the different group nodes. The left part in Figures 5 and 6 captures the author group and the 6 medical professional groups; the right part — the 6 linguist groups. I.e., UC confirms the similarity between the summaries of the authors and the summaries of the medical doctors on the one hand, and between the summaries of the linguists on the other hand.

As can be observed, the results are also similar to the results obtained by *Multidimensional Scaling*.

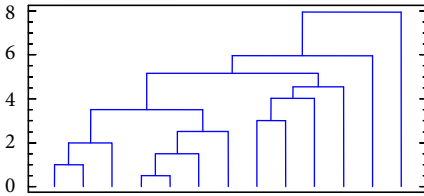


Figure 5. Overall similarity between the subject and author summaries

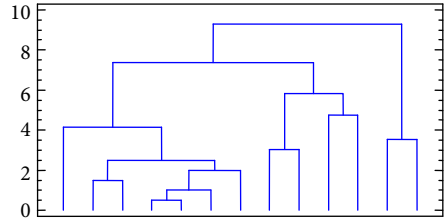


Figure 6. Contrast of the content of the authors' summaries to the content marked for inclusion into the summary by the 12 subjects informants

In the next section, we discuss the very pronounced similarity pattern between the summaries of the medical doctors and the deviances of these summaries from the summaries of linguistic professionals in more detail.

4.1.3 Qualitative Assessment of the Experiment Outcome

The *Multidimensional Scaling* and *Clustering* experiments described above confirm the high degree of coincidence among the summaries by medical professionals (informants and authors) and their deviance from the summaries by linguists. Let us now analyze the content chosen by the two groups.

Firstly, we observe a general trend among linguists to include too much information from the *Introduction* section into their summaries. We hypothesize that this is due to their lack of formation in medicine, which makes them consider term definitions, historical data, confirmation of facts, references to previous works, etc. relevant to the summary. In contrast, medical doctors did not choose this type of content for their summaries because it appears to them either trivial or well-known in the field.

Secondly, we observe that medical professionals select balanced information from each of the four sections (*Introduction*, *Patients and methods*, *Results* and *Discussion*) with the goal to reflect in the summary the IMRD-structure of the article. The linguists, although aware of the fact that a medical summary must follow the IMRD-structure, did not follow this guideline strictly.

Thirdly, we observe that medical doctors usually select numerical information, especially in the section of *Patients and methods* and *Results*, whereas linguists tend to avoid the inclusion of numerical information. Instead, they favour more explanations and fewer figures.

Finally, we clearly see that linguists preferably select content from the *Discussion* section, whereas medical doctors normally include into the summary a much briefer account of the *Discussion*.

Table 2. Cases and significant examples of different content chunks selected by medical doctors and linguists

Case	N	Examples	Drs	Lings	
Introduction	1	- Definitions: Ex. “ <i>C. difficile</i> es un bacilo grampositivo anaerobio, productor de esporas, que puede colonizar a pacientes hospitalizados, en especial a ancianos, tras tratamiento antibiótico.” Ex. “El paraquat es un herbicida bipiridilo.”	0%	100%	
		2	- Historical data: Ex. “Desde que a finales de la década de los años cincuenta se empezaron a utilizar los primeros fármacos antihipertensivos se ha producido una explosión imparable de nuevos fármacos hasta nuestros días.” Ex. “La osteoporosis y su complicación clínica, las fracturas, han despertado en los últimos años un gran interés no sólo por su alta morbi-mortalidad sino también por su estrecha relación con el envejecimiento poblacional.”	0%	100%
			3	- Confirmation of facts: Ex. “La infección por el VIH en España ha estado focalizada principalmente en los usuarios de drogas inyectadas, y a partir de esta población se ha ido extendiendo de forma secundaria por vía sexual y perinatal.” Ex. “En la actualidad, el cáncer de mama constituye un problema de gran importancia sanitaria en los países desarrollados.”	0%
	4	- Previous works: Ex. “La caracterización molecular de la cistinuria comenzó a principios de los noventa cuando Calonge et al. demostraron que varias mutaciones en el gen SLC3A1 estaban asociadas a la cistinuria humana.” Ex. “Sempere et al. han validado un Protocolo de Adecuación de Urgencias Hospitalarias (PAUH) para detectar visitas inadecuadas a este servicio.”		0%	100%
		5	Ex. “Con la RM-mielografía se obtuvo información nueva en 81 casos (32%).” Ex. “La progresión a sida en España hasta 1996 y el efecto de la edad son similares a otras cohortes europeas.”	100%	0%
			...		
	Patients and methods and Results (Numerical information)	6	Ex. “La prevalencia de anticuerpos anti-VIH fue de 0,99 por 1.000 en 1996, 1,29 en 1997, 1,42 en 1998 y 1,54 en 1999.” Ex. “Las lesiones bucales, faríngeas y/o esofágicas fueron visibles en algunos pacientes ya en el momento del ingreso, pero en otros aparecieron días después.”	100%	0%
			7	Ex. “Dado el aumento de la prevalencia del VIH en madres de recién nacidos, son necesarios el consejo y la oferta sistemática de la prueba del VIH a todas las mujeres embarazadas.” Ex. “La seroprevalencia del VIH en madres de recién nacidos obtenida en este estudio no puede considerarse representativa de toda España.” Ex. “Aun así, las seroprevalencias encontradas son superiores a las descritas en otros países europeos.” Ex. “El patrón geográfico es muy similar al de la incidencia de casos de sida.”	100%
	0%	16.6%			
	0%	16.6%			
	0%	16.6%			

Table 2 summarizes the material on which our analysis is based in terms of some representative examples from the corpus. It shows a number of content chunks that have been selected, respectively omitted by medical professionals and linguists. The first column (Case) refers to the case that is illustrated, the second to the number that is assigned to each case (N), the third (Example) offers specific examples of the case in question; the fourth (Drs) and the fifth (Lings) indicate the percentage of medical doctors and linguists who selected these examples.

The quantitative and qualitative evaluation of summaries provided especially by medical specialists provides some insights that are immediately relevant to our effort to develop a model for automatic summarization of medical articles. Firstly, professionals adopt a different strategy for the selection of summary relevant content than language specialists, who can be assumed to be knowledgeable in the structure of general discourse and linguistically motivated summary writing. Secondly, all professionals tend to choose roughly the same content chunks for the summary — which means that the summaries of the professionals can be drawn upon for distillation of the summarization criteria. Thirdly, if available, the summary of the author (as one of the professionals) can be taken as reference for later evaluation of our model.

In order to work out a summarization model for medical articles in Spanish, we must thus draw upon the specialized discourse knowledge of the professionals, i.e., study the summaries of the professionals.

4.2 Towards an extraction-based summarization model

The summaries of the medical specialists in Sections 4.1.2 and 4.1.3 reveal that:

- the summary must include information from all four sections of the article;
- numerical information from the *Patients and Methods* and *Results* sections must be included;
- definitions, historical data, confirmations of facts and references to previous works should not be included into the summary;
- the *Discussion* part of the summary must not be longer than the parts corresponding to the other sections of the article.

Obviously, such general criteria do not suffice for the construction of a fine-grained summarization model. But they set up a framework which must then be refined. Thus, we must ensure that the four sections are identified. Furthermore, we must identify and assess numerical data (to be kept for the summary). This can be done by searching for specific keywords, numerals, etc.

As a rule, numerical data are embedded into the discourse structure of the text. Thus, in the following example (Example 1) from the *Patients and methods*

section, between the sentence containing numerical information and the succeeding sentence, the discourse relation ELABORATION in the sense of the *Rhetorical Structure Theory* (cf. also Section 2.1) holds.¹¹

Example 1:¹²

[*En marzo de 1997 se produjo, de forma explosiva, en una unidad de geriatría de 24 camas, un brote epidémico de colitis que afectó a 12 (50%) de los 24 pacientes ingresados.*]_N [*Se consideró que los pacientes estaban afectados por diarreas cuando presentaban tres o más deposiciones diarias blandas o líquidas durante un mínimo de 2 días, en ausencia de otra causa que lo justificara.*]_S
 ‘[In March 1997, in a geriatric unit of 24 beds, an epidemic outbreak of colitis, which affected 12 (50%) of the 24 admitted patients, occurred in an abrupt form.]_N [Due to the absence of another cause that justified it, it was considered that the patients were infected by diarrhoea when they presented three or more soft or liquid daily depositions during a minimum of 2 days]’_S

In order to ensure that, on the one hand, the deletion of one element of a discourse structure does not lead to incoherence, and, on the other hand, no summary relevant element of the discourse structure is deleted or no irrelevant element is kept for the summary, the summarization strategy must draw upon the discourse structure.

As already mentioned in Section 2, in the literature, discourse driven techniques received a strong echo (Ono et al. 1994; Marcu 2000; Teufel and Moens 2002). However, it has also been noted that discourse criteria alone often lead to both omission of summary relevant information and inclusion of summary irrelevant information. In the domain of medical articles, we found that in particular discourse tree depth criteria used, e.g., by Marcu for elimination of text chunks, are not reliable.¹³ On the other hand, when used independently of the discourse tree depth, certain RST-relations (such as CONDITION and CAUSE, etc.) can be well used on their own, without reference to other criteria. Consider the following example:

Example 2:

[*Disponer de un amplio arsenal de fármacos antihipertensivos puede constituir un inconveniente en la toma de decisiones terapéuticas*]_N [*si los esquemas son confusos.*]_S
 ‘[To have a wide arsenal of antihypertensive drugs can be an inconvenience for making therapeutic decisions]_N [if the schemas are confusing.]’_S

Between the first and the second sentence, the RST relation CONDITION holds. If the first sentence (the relation’s nucleus) is to be included into the summary, the second sentence (the satellite) cannot be omitted.¹⁴ Other relations such as

ELABORATION, SEQUENCE, etc., require, as a rule, additional communicative and/or syntactic criteria to be of use for summarization; for illustration, consider Example 3.

Example 3:

[La **infección** por *C. difficile* es la causa más frecuente de diarrea nosocomial en nuestro medio y representa el 15–20% de las diarreas asociadas al uso de antibióticos.]_N [El espectro clínico de la **infección** oscila desde la colitis pseudomembranosa hasta la diarrea leve y el portador asintomático.]_S
 ‘[The *C. difficile* **infection** is the most frequent cause of nosocomial diarrhoea in our environment and represents the 15–20% of the diarrhoea associated to the use of antibiotics.]_N [The clinical spectrum of the **infection** oscillates between the pseudo membranous colitis and the slight diarrhoea and the asymptomatic bearer.]_S

In Example 3, between the first and the second sentence, the relation ELABORATION holds. If we know that the satellite sentence elaborates on the *Theme* of the nucleus sentence,¹⁵ we can deduce that it provides further details on already mentioned content and can thus be safely discarded from inclusion into the summary.

In addition to discourse, syntactic and communicative criteria, we must consider lexical summarization criteria. As already in the case of other types of criteria, lexical criteria include *omission* criteria and *inclusion* criteria. The latter are based on the insight that for each thematic area, some cue words are available that mark relevant text chunks. Our study revealed that in Spanish medical articles, the set of the relevance-marking cue words includes the Spanish equivalents of such nouns as

objective, object, summary, purpose, intention, result, etc.

and of such verbs as

[to] carry out, [to] associate, [to] analyze, [to] present, [to] relate, [to] evaluate, [to] contribute, [to] study, [to] value, [to] find, etc.

Cf. two examples (the cue words are in bold):

Example 4:

*El **objetivo** del presente estudio ha sido conocer el patrón nutricional de la población escolar de un núcleo rural con marcado carácter industrial e identificar sus alteraciones nutricionales.*

‘The **objective** of the present study has been to know the nutritional pattern of the school population of a rural centre with a marked industrial character and to identify its nutritional alterations.’

Example 5:

*El estudio descriptivo **analizó** el estado de salud percibido según el tipo de trabajo realizado (trabajadoras o amas de casa) y el resto de variables.*

‘The descriptive study **analyzed** the state of health perceived according to the working type carried out (workers or homemakers) and the rest of variables.’

From the empirical study, we can thus conclude that an adequate extraction-based summarization model must be multidimensional. It must take into account textual, discourse, syntactic, communicative and lexical criteria.

5. Building a motivated summarization model

Our extraction model is a rule-based model that consists of four different submodules. Each submodule contains rules of an individual dimension or a combination of dimensions of the linguistic description: the textual, the lexical, the discourse and the syntactic-communicative dimensions. To derive the initial rules for each submodule, we carried out an empirical analysis of fifty medical articles in Spanish. The submodules are continuously being extended as further material is analyzed. Figure 7 shows the architecture of our model.

Let us briefly discuss each of the stages of our model.

5.1 Primary textual processing stage

The primary textual stage captures the four-partite textual structure of a medical article: (1) *Introduction*, (2) *Patients and methods*, (3) *Results*, and (4) *Discussion*. Remember that medical journals usually require the division of an article into four sections, prescribing, roughly, the headings of each. Furthermore, they require a summary to reflect the content of each section.

Given that the original articles contain predefined section titles, this stage is nearly trivial — although we must take into account that deviations from the predefined titles may well be observed.

5.2 Primary lexical processing stage

The primary lexical processing stage eliminates text passages that are demonstrably not relevant to the abstract. These are, among others, passages that contain:

Figure 7. Architecture of the multidimensional summarization model

- cue words referring to statistical evaluations of the content of the article;
- references to tables and figures;
- cue words identifying definitions (such as the noun *definición* ‘definition’ and its adjectival and verbal derivatives) and explicit new term introductions;
- references to previous and related work identified in terms of explicit bibliographical references and cue term patterns such as (DET/PRON +) *trabajo(s)* ‘work(s)’/*estudio(s)* ‘study(-ies)’/*investigación(es)* ‘investigation(s)’ / ... (+ MODIF), with the optional DET/PRON including *el* ‘the’/*otro* ‘other’ /*algún* ‘any’ /*ese* ‘this’/... and the optional MODIF including *previo* ‘previous’/*anterior* ‘previous’/...;
- historical data (dates, years, etc.).

5.3 Discourse + syntactic-communicative processing stages

The discourse and the syntactic-communicative processing stages are intertwined. They adapt general purpose summarization criteria to the domain of medical

articles. As already mentioned above, we use RST, which proved to be suitable for summarization, as discourse representation. However, our use of RST is rather different from the uses discussed in the literature. Thus, while, e.g., Marcu (2000) — see in particular endnote 13 — draws upon the depth of the discourse tree, we assume that it is the nature of each RST-relation, which allows us to judge its importance for the summary. Cf. the following rule that makes use exclusively of the discourse structure:¹⁶

IF S is satellite of a CONDITION relation CO
THEN KEEP S

It indicates that the satellite element must be kept in order not to lose essential information. This rule is applicable to Example 2 above as well as to the following Example 6.

Example 6:

Si los pacientes requieren un flujo superior a 3 litros por minuto, es probable que no sea bien tolerado y seguramente la VAO no es la mejor solución.

‘If patients require a flow superior to 3 liters per minute, it is likely that it is not well tolerated, and VAO is certainly not the best solution.’

Another rule of this type is:

IF S is satellite of a BACKGROUND relation B
THEN IF $|Sec_{SAT,rem}| > 1$ ELIMINATE S

(with $Sec_{SAT,rem}$ as section in which the satellite S of B occurs)

Example 7:

[En la enfermedad isquémica coronaria (EIC) se han descrito diferentes alteraciones en la circulación sistémica del sistema hemostático, existiendo muy poca información de los posibles cambios que pueden acontecer en la circulación coronaria, cerca de la lesión trombótica y sus diferencias con las alteraciones encontradas a nivel periférico.]_N [Nos planteamos este trabajo con el doble objetivo de investigar si en los enfermos con EIC existen diferencias en distintas variables hemostáticas al efectuar sus determinaciones en la sangre obtenida del seno coronario (SC) respecto a la obtenida de la circulación periférica (CP), así como los posibles cambios que la ATPC pudiera producir en la hemostasia en ambos lugares.]_S

‘[In coronary ischemic disease (EIC), different alterations in the systemic circulation of the haemostatic system have been described, the information being scarce of possible changes that may occur in coronary circulation, near the thrombosis injury and its differences with alterations found at the peripheral level.]_N [We address this work with the double objective of investigating whether there are differences in ill people with EIC with respect to different haemostatic variables when

carrying out their identification in blood obtained from coronary sinus (SC) compared to blood obtained from peripheral circulation (CP), as well as possible changes that ATPC could produce in the homeostasis of both locations.]_S'

All RST-relations used in our model are summarized in Appendix 2 of this article.

As syntactic representation, we use the dependency-oriented deep-syntactic structures of the Meaning-Text Theory, MTT (Mel'čuk 1988).¹⁷ As communicative structure (i.e., *information structure*), we also use MTT's communicative structure (Mel'čuk 2001). As already pointed out above, syntactic-communicative criteria and discourse criteria tend to be intertwined in that a syntactic or communicative criterion is required to be valid for one of the elements (nucleus or satellite) of a given discourse relation. Consider two examples:

IF S is satellite of an ELABORATION relation E
 AND there is a syntactic dependency structure such that
 'X -ATTR→S'
 THEN ELIMINATE S
 (with 'X' as an element in the syntactic structure)

and

IF S is satellite of ELABORATION E
 AND S elaborates on the Theme of the nucleus N of E
 THEN ELIMINATE S

The application of both rules leads to the elimination of the S-fragment. The first combines syntactic criteria with discourse criteria, while the second combines communicative criteria with discourse criteria. Cf. an example for the first rule:

Example 8:

Curiosamente, en nuestro trabajo se comportó como factor de riesgo la edad tardía de la menarquía (incluyéndose esta variable en la población general y posmenopáusica, en el modelo final de riesgo).

'Curiously, in our work, the late occurrence of the menarche acted as a risk factor (this variable being included in the general and postmenopausal population, in the final model of risk).'

(the APPEND-dependent is in bold)

The second rule applies to Example 3 above and to Example 9:

Example 9:

[A los portadores de cuerpos extraños intraabdominales que contienen cocaína, con fines de contrabando, se les conoce con el síndrome del body packer.]_N [Los prin-

principales problemas médicos que se plantean en estos pacientes son: la sobredosificación de drogas por la rotura de uno de los paquetes y la obstrucción intestinal por impactación de dichos paquetes en el tubo digestivo.]_S

‘[Bearers of extrinsic intra-abdominal bodies containing cocaine with the purpose of smuggling are known as having the body packer syndrome.]_N [The main medical problems considered with respect to these patients are: overdose of drugs due to the breakage of one of the packets and intestinal obstruction due to the impact of these packets in the digestive tube.]_S’

The combination of communicative structure criteria (and, in particular, of the Theme/Rheme dichotomy) with discourse structure criteria turned out to be a very reliable source of evidence for summarization. In the literature, both types of structures have been applied so far separately (e.g., Marcu 2000; Teufel and Moens 2002) for the use of discourse structure.

5.4 Secondary textual and lexical processing stages

The secondary textual and lexical processing stages serve to assess the relevance of the fragments of the article in question selected for inclusion in the summary. The criteria that contribute to an increase of the relevance score of a sentence are of two kinds:¹⁸

- i. occurrence of lexical units with specific characteristics: (1) verbs in the 1st person form,¹⁹ (2) any content word from the title, (3) any lexical unit from the list of cue words which has been compiled beforehand, (4) numerical information in *Patient and methods* and *Results* sections; cf. examples for each of them:
 - (1) *Nos hemos decidido a comunicar nuestra experiencia con la dieta hipocalórica como tratamiento único en pacientes afectados de OSAS.* ‘We have decided to communicate our experience in the hypo caloric diet as unique treatment of patients infected by BEARS.’
 - (2) *Programas de detección precoz del cáncer de mama y acceso a la mamografía en España.* ‘Programs for the precocious detection of breast cancer and access to mammography in Spain.’
 - (3) *En las clases más altas (I-II), aunque la mujer trabajase exclusivamente en el hogar, presentaba mejor estado de salud que las trabajadoras de las clases más bajas.* ‘In the highest classes (I- II), even if women worked exclusively at home, they showed better health conditions than workers from lower classes.’

(4) *Separando la mortalidad atribuible por sexos, los resultados obtenidos fueron del 21,2% en varones y del 10,0% en mujeres, mientras que los porcentajes de mortalidad atribuible referidos a nivel nacional fueron del 20% y del 5%, respectivamente para varones y mujeres. ‘Separating mortality with respect to gender, the obtained results were of 21.2% in men and 10.0% in women, whereas the percentages of mortality with respect to this criterion at national level were of 20% and 5% for men and women, respectively.’*

- ii. the position of the sentence in the article: (1) the last paragraph of the *Introduction* section, (2) among the first two sentences of the *Patients and methods* section; (3) among the first two sentences of the *Results* section; (4) the first or the last paragraph of the *Discussion* section; cf. an example for case (2):

Los pacientes suicidas que padecían una enfermedad orgánica eran 45, lo que constituye un 17% del total. La edad media de estos pacientes fue de 58,3 años (varones 57,6 años y mujeres 59,2 años) con unos límites de 16 a 90 años. ‘Suicidal patients who suffered an organic illness were 45, constituting the 17% of the total. The average age of those patients was of 58.3 years (males 57.6 years and women 59.2 years) within the limits between 16 and 90 years.’

The last two stages, the secondary textual processing stage and the lexical processing stage, calculate the relevance of the individual sentences within the selected fragments with respect to their inclusion into the summary; cf. a sample textual rule and a sample lexical rule used during these stages:

IF sentence s is one of the 3 last sentences
of the “Introduction”

THEN $\Delta_s := \Delta_s + \delta_s$

IF a sentence s contains a verbal form in 1st person plural

THEN $\Delta_s := \Delta_s + \delta_s$

(with Δ_s as the global relevance weight of the sentence s assessed so far and δ_s as the increment of this weight in case the condition of the corresponding rule is met).

The set of the summarization rules used in our model is given in Appendix 1.

6. Validity of the model

For the validation of the performance of our model in praxis, we compare the outcome of its application to five medical articles arbitrarily chosen from our test corpus with the author summaries and the summaries of three medical doctors. The validation is done using two different techniques: ROUGE, which is a stan-

standard evaluation technique of summarization systems, and the Relative Euclidean Distance.

6.1 Evaluation with ROUGE

For the ROUGE-based evaluation,²⁰ we took the summaries of the authors and the summaries of the three medical doctors as models. The summaries of our model were the candidates to be evaluated. As baseline, we used a random selection of the predefined number of sentences from each of the four sections of the original article. The baseline thus incorporated some linguistic information also observed by our summarization strategy; namely that the summary has to include content from each section: *Introduction*, *Patients and methods*, *Results* and *Discussion*.

Given that ROUGE is set up for English and our material is in Spanish, we had to adapt ROUGE for the task. More precisely, we had to (i) replace the English stop word list by a Spanish list that contained the stop words relevant to our topic; (ii) remove the Porter stemmer included in ROUGE, (iii) use lemmatized summaries.²¹

We carried out the evaluation with ROUGE 2 (which uses bigrams) and with ROUGE SU-4 (which uses fourgrams). The results are shown in Tables 3 and 4. The average of both measures is shown in Table 5.

Table 3. ROUGE 2 measures using as models the authors' and medical doctors' summaries and as candidates the summaries provided by our model

ROUGE 2	text 1	text 2	text 3	text 4	text 5
Our model	0.67719	0.73856	0.66617	0.55205	0.65112
baseline	0.5122	0.20822	0.30849	0.17406	0.28731

Table 4. ROUGE SU-4 measures using as models the authors' and medical doctors' summaries and as candidates the summaries provided by our model

ROUGE SU-4	text 1	text 2	text 3	text 4	text 5
Our model	0.64320	0.72189	0.63447	0.52632	0.63476
baseline	0.33333	0.19590	0.30432	0.18221	0.30762

Table 5. Average of ROUGE 2 and ROUGE SU-4 measures

ROUGE average	ROUGE 2	ROUGE SU-4
Our model	0.65702	0.63213
baseline	0.29806	0.26468

As the three tables show, the quality of the summaries of our model is rather high. In the case of both, ROUGE 2 and ROUGE SU-4, it is considerably better than the baseline.

6.2 Evaluation with the Relative Euclidean Distance

For the evaluation with the Relative Euclidean Distance, the results of the experiment are set out in Table 6. Table 6 has the same structure as Table 2. Thus, as already in Table 2, the first column (*Content*) lists all statements that occur in any of the summaries (i.e., either in our summary, in the summary of the author or in a summary of one of the doctors); in total, the table contains 58 statements, of which only six are shown. The second column (*Variable*) contains the identification label of each statement. The third column displays the performance of our model. The remaining columns (*Author*, *Dr.1*, *Dr.2*, *Dr.3*) specify whether the statement in question has been included ('1') or not ('0') by the author or by one of the medical professionals, respectively.

Table 6 shows that the medical professionals and our model equally “under-summarize” somewhat, i.e., include into the summary statements not included by the author. However, in general, the model performs well. Also, we consider it a positive sign that the summary relevance judgements of the medical doctors largely coincide with the judgements of our model.

For a quantitative estimation of the similarity between the alternative summaries of the different subjects, we used the Euclidean distance between them, applying the Pythagoras theorem in the given 58-dimensional space (remember that we are working with 58 statements / variables). Given that the absolute distance values are difficult to assess, we normalize the distances by the distance between two maximally diverging vectors in our vector space: the one with 58 ‘1’s and the one with 58 ‘0’s. The distance between these two vectors is 7.616. In Table 7, the distances are thus distances relative to 7.616. Note that to accommodate for the two normalization vectors, Table 7 has been extended by two lines and two columns.

When reinterpreted in terms of the percentage of relative divergence (i.e., relatively to the maximal divergence), the figures in Table 7 mean that the five summaries of our model diverge from the author summaries in 41% of the cases, from the summaries of Dr. 1 in 22%, and from the summaries of Dr. 2 and Dr. 3 in 26%. In other words, the summaries of our model coincide with the summaries of the authors in 59%, with the summaries of Dr. 1 in 74%, and with the summaries of Dr. 2 and Dr. 3 in 78%. The author summaries and the summaries of the doctors diverge in 34% and 32% of the cases, respectively. Consider a graphical representation of the distances in Figure 8.

Table 6. The inclusion of the statements from a sample article into the summary by the different subjects and our model

Content	Variable	Model	Author	Dr. 1	Dr. 2	Dr. 3
Estudiar los principales factores de riesgo de infección por <i>Clostridium difficile</i> en una unidad de geriatría.	X1-1	1	1	1	1	1
Estudio de casos y controles retrospectivo.	X1-2	1	1	1	1	1
El análisis multivariante confirmó la nutrición enteral por sonda nasogástrica (OR = 6,73; IC del 95%, 1,01–45,35) y los días de tratamiento antibiótico (OR = 1,15; IC del 95%, 1,01–1,28) como factores de riesgo independientes para la infección por <i>C. difficile</i> .	X1-3	1	1	1	1	1
El tratamiento antibiótico, el sondaje nasogástrico y las características de fragilidad de este grupo de pacientes se asocian a la infección por <i>C. difficile</i> .	X1-4	1	1	1	1	1
Se han implicado otros factores no relacionados con el tratamiento antibiótico como factores de riesgo de infección por <i>C. difficile</i> .	X1-5	1	0	1	1	1
La práctica de una política antibiótica restrictiva respecto a la utilización de antibióticos, como se ha llevado a cabo en otros hospitales, reducirá significativamente la colonización y la CACD	X1-6	1	0	1	1	1
...

Table 7. Distance (divergence) matrix

Case	Relative Euclidean distance						
	1:Author	2:Model	3:Dr. 1	4:Dr. 2	5:Dr. 3	6:0-vector	7:1-vector
1:Author	0.000	3.162	2.646	2.449	2.449	7.071	2.828
2:Model	3.162	0.000	1.732	2.000	2.000	7.483	1.414
3:Dr. 1	2.646	1.732	0.000	1.732	1.732	7.550	1.000
4:Dr. 2	2.449	2.000	1.732	0.000	2.000	7.483	1.414
5:Dr. 3	2.449	2.000	1.732	2.000	0.000	7.483	1.414
6:0-vector	7.071	7.483	7.550	7.483	7.483	0.000	7.616
7:1-vector	2.828	1.414	1.000	1.414	1.414	7.616	0.000

For deeper comprehension and further amelioration of our model, it is necessary to analyse cases where the summaries of our model do not coincide with the author summaries. In the experiment discussed in Section 3.1, our model did not include statements found in the author summaries under two concrete circum-

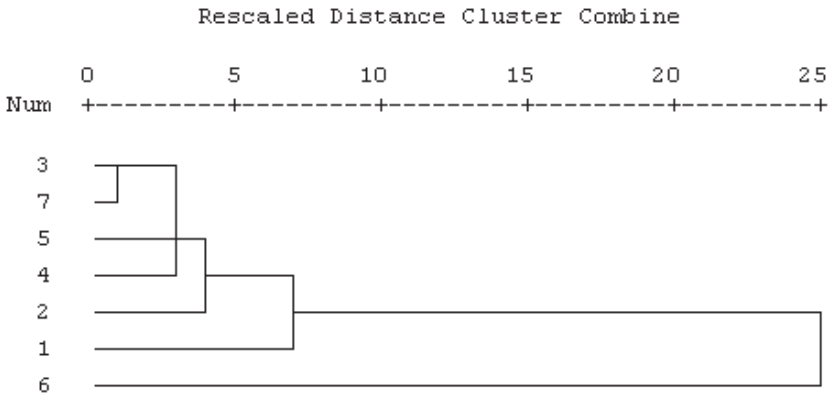


Figure 8. Distance between the summaries of the different subjects and our model (the numbers in the first column refer to the subjects and our model as indicated in Table 4)

stances. Firstly, when the statements contained references to tables such as, e.g., “*Tabla 2*” “Table 2”, “*Tablas 5 y 6*” “Tables 5 and 6”. These statements are eliminated in the primary lexical stage of our summarization.

In all other cases of divergence, our model “undersummarized” i.e., it selected for the summary all statements also selected by the authors, and, in addition, some others not chosen by the authors. These were, first of all, statements from the *Discussion* section, many of which have also been chosen by some or all of the three doctors. *Discussion* is the most complex section of a medical article and thus the most difficult one to summarize. The decision to include a statement from the *Discussion* section into a summary is highly subjective. The fact that other medical professionals tend to select the same statements as our model suggests that these statements are sufficiently relevant to be included in the summary.

7. Summary and conclusions

We believe that we provided evidence that a rule-based summarization model that takes into account the whole range of partly domain-specific linguistic criteria (textual, lexical, discursive, syntactic and communicative) is able to provide high quality summaries of medical articles. The implementation of this model, its further extension and evaluation is subject of ongoing work. For the time being, we are working with a sample corpus annotated semi-automatically with abstract syntactic dependency structures, RST discourse structures and communicative (Theme/Rheme) structures.²² We consider this a valid methodology for a “proof-of-concept” realization of our model. The application of our model to unannotated

corpora would require the availability of a comprehensive parsing/analysis workbench for Spanish, which is still an open problem.

Acknowledgments

This work could not have been carried out without the valuable collaboration of the twelve participants of the experiment. These are six linguists of the Institute for Applied Linguistics of Pompeu Fabra University of Barcelona (Carme Bach, Anna Joan, Ricardo Guantiva, Rodrigo Alarcón, Amor Muntaner and Rogelio Nazar) and six medical doctors from different institutions: Eduardo Barge (Juan Canalejo Hospital, A Coruña), Iria Glez (Teresa Herrera Hospital, A Coruña), Daniel López (Juan Canalejo Hospital, A Coruña), Víctor Vicens (Bellvitge Hospital, Hospitalet de Llobregat), María Giralt (Benito Menni Hospital, Sant Boi de Llobregat) and M^a José Alierta (Spanish and Catalan Association of Psychoanalytic Psychotherapy, Barcelona). We are grateful for their dedication and enthusiasm. Many thanks also to our expert in statistics Jaume Llopis, who was of great help in the statistical evaluation part of our experiment.

Notes

1. Extraction oriented summarization is based on the selection and subsequent potential reduction of summary relevant sentences — in contrast to “abstraction” oriented summarization, which uses knowledge analysis and text generation techniques.
2. In accordance with the linguistic framework underlying our work, we use the term “communicative structure” for what is often referred to as “information structure” (e.g., Vallduví 1990; Kruijff-Korbayová and Steedman 2003).
3. An overview of the state of the art on automatic summarization of general discourse is given in Mani and Maybury (1999) and Mani (2001). Cf. also the introduction to the special issue on summarization in *Computational Linguistics* (Radev et al. 2002) and also the forthcoming special issue on the topic in *Information Processing and Management* 43(4), 2007.
4. Farzindar et al. use the term “thematic structure” to refer to what is commonly known as “text structure.” In order to avoid confusion, we use the term “text structure.”
5. Such surface criteria may well require deep processing — e.g., named entity recognition.
6. IMRD stands for *Introduction, Patients and methods, Results and Discussion*.
7. The 20 articles in our experiment are articles from *Medicina Clínica*, which forms part of the Technical Corpus of the Institute for Applied Linguistics of the Pompeu Fabra University of Barcelona. This corpus is available online and can be accessed via <http://bwananet.iula.upf.edu/>.
8. Note that the purpose of this table is merely to illustrate the similarity of the content choices made by the subjects, not to analyze why the individual subjects have picked one or the other content chunk. Therefore, we do not include into the table such selection criteria as position of the chunks in the text.

9. In the Diagram of Figure 3, the two first dimensions of the Multidimensional Scaling model cover 67.9% of the information; in the Diagram of Figure 4, 61.9%. Taking into account that we are working with 624 variables this means that there is a strong correlation among the variables and that the representation obtained by Multidimensional Scaling is indeed reliable and informative.
10. In general, the subjects tended to select for the summary the maximal amount of chunks permitted in the experiment (i.e., the 20 lines) while the summaries of some authors were considerably shorter.
11. For the sake of the clarity of our presentation, let us mention that in RST discourse relations are binary relations. Asymmetrical and symmetrical relations are distinguished. For instance, ELABORATION, CONSEQUENCE, CAUSE, etc. are asymmetrical relations; LIST, CONTRAST, etc. are symmetrical relations. The head of an asymmetrical relations is called *nucleus* (N), the tail — *satellite* (S). In symmetrical relations, both elements are nuclei.
12. The English translations of the following Spanish examples are very approximate and nearly literal. Therefore some of them may appear on the edge of being ungrammatical.
13. In short, Marcu's strategy prefers discourse structure elements that form part of nuclei of discourse relations at several depths of the discourse tree. Satellite elements or elements that form part of only a few (or one) nuclei are candidates for elimination.
14. In Marcu's model, such a satellite would be a strong candidate for elimination.
15. We interpret the term Theme in the sense of the Prague school (Sgall et al. 1986) and of the Meaning-Text Theory (Meřčuk 2001); cf. also Section 5.
16. For better readability, the rules are cited in a pseudo-code format.
17. Dependency-based syntactic structures proved to be more suitable for the purpose of extraction-based summarization than constituency-based structures since the former provide us with grammatical information whose relative relevance can be assessed (e.g., an attribute is less relevant than an actant of a given lexeme; the first actant is more relevant than the second or third actant, etc.), while the latter does not.
18. Each criterion that is met augments the relevance score of a fragment by a given weight increment.
19. In medical articles, sentences containing 1st person verb forms are, as a rule, comments of the authors on their own work — which makes them relevant for the summary.
20. We cannot go into details of the principles of ROUGE evaluation here. Interested readers are asked to consult (Lin 2004).
21. Instead of using lemmatized summaries, we could have also used a Spanish stemmer. However, as has been shown in IR, for languages with rich inflection (such as Spanish), lemmatization is more appropriate.
22. The annotation of the corpus has been done by several annotators using an interactive annotation editor. Some of the annotators were linguists; others have been trained on the various representation structures specifically for the annotation task. Contrary to the widespread opin-

ion that, for instance, RST-annotation is highly annotator-specific, we did not make this experience: the annotations by different annotators coincided to a large extent.

References

- Abracos, J. and G.-P. Lopes. 1997. "Statistical methods for retrieving most significant paragraphs in newspaper articles." In Mani, I. and M. Maybury (eds.). *Proceedings of the ACL/EACL'97. Workshop on Intelligent Scalable Text Summarization*. 51–57. Madrid, Spain.
- Afantenos, S.D., V. Karkaletsis and P. Stamatopoulos. 2005. "Summarization of medical documents: A survey." *Artificial Intelligence in Medicine* 33(2), 157–177.
- Barzilay, R. and M. Elhadad. 1997. "Using lexical chains for text summarization." In Mani, I. and M. Maybury (eds.). *Proceedings of the ACL/EACL'97. Workshop on Intelligent Scalable Text Summarization*. 10–17. Madrid, Spain.
- Brandow, R., K. Mitze, and L. Rau. 1995. "Automatic condensation of electronic publications by sentence selection." *Information Processing and Management* 31, 675–685.
- Ciapuscio, G. 1998. "Los resúmenes de la revista *Medicina*: Un enfoque diacrónico-contrastivo." *Revista Signo y Señal* 10, 217–243.
- da Cunha, I. and L. Wanner. 2005. "Towards the automatic summarization of medical articles in Spanish: Integration of textual, lexical, discursive and syntactic criteria." In *Proceedings of the Workshop "Crossing Barriers in Text Summarization Research. RANLP-2005 (Recent Advances in Natural Language Processing)*. 46–51. Borovets, Bulgaria.
- da Cunha, I. and J. Llopis. In print. "Constatación de la validez de los resúmenes adjuntos a artículos médicos de investigación de cara a la evaluación de resúmenes automáticos." In *Proceedings of the XXIV Congreso Internacional de la Asociación Española de Lingüística Aplicada*. Universidad Nacional de Educación a Distancia, Madrid, Spain.
- Damianos, L., S. Wohlever, G. Wilson, F. Reeder, T. McEntee, R. Kozierek, L. Hirschman and D. Day. 2002. "Real users, real data, real problems: The MiTAP system for monitoring bio events." In *Proceedings of the Conference on Unified Science & Technology for Reducing Biological Threats & Countering Terrorism*. 167–177. University of New Mexico: Mexico.
- Edmundson, H. P. 1969. "New methods in automatic extraction." *Journal of the Association for Computing Machinery* 16, 264–285.
- Farzindar, A., G. Lapalme and J.-P. Desclés. 2004. "Résumé de textes juridiques par identification de leur structure thématique." *Traitement automatique des langues* 45(1), 39–64.
- Gaizauskas, R., P. Herring, M. Oakes, M. Beaulieu, P. Willett, H. Fowkes and A. Jonsson. 2001. "Intelligent access to text: Integrating information extraction technology into text browsers." In *Proceedings of the Human Language Technology Conference (HLT 2001)*. 189–193. San Diego, USA.
- Goldstein, J., J. Carbonell, M. Kantrowitz and V. Mittal. 1999. "Summarizing text documents: sentence selection and evaluation metrics." In *Proceedings of SIGIR-99*. 121–128. Berkeley, California.
- Johnson, D.B., Q. Zou, J.D. Dionisio, V.Z. Liu and W.W. Chu. 2002. "Modeling medical content for automated summarization." In *Annals of the New York Academy of Sciences*. 247–258. New York: New York Academy of Sciences.
- Kan, M.Y., K.R. McKeown and J.L. Klavans. 2001. "Domain-specific informative and indicative summarization for information retrieval." In *Proceedings of the Workshop on text summarization (DUC 2001)*. 19–26. New Orleans, USA.

- Kan, M.Y. 2003. *Automatic Text Summarization as Applied to Information Retrieval: Using indicative and informative summaries*. PhD dissertation. Columbia University, New York, USA.
- Kaplan, R.B., S. Cantor, C. Hagstrom, L.D. Kamhi-Stein, Y. Shiotani and C.B. Zimmerman. 1994. "On abstract writing." *Text* 14(3), 401–426.
- Kruijff-Korbayová, I. and M. Steedman. 2003. "Discourse and information structure." *Journal of Logic, Language and Information* 12, 249–259.
- Kupiec, J., J. O. Pedersen and F. Chen. 1995. "A trainable document summarizer." In *Proceedings of SIGIR-95*. 68–73. New York, USA.
- Lehman, A. 1995. *Le résumé des textes techniques et scientifiques, aspects linguistiques et computationnels*. PhD dissertation. Université de Nancy 2.
- Lenci, A., A. Água, R. Bartolini, S. Busemann, N. Calzolari, E. Cartier, K. Chevreau, and J. Coch 2002. "Multilingual summarization by integrating linguistic resources in the MLIS-MUSI project." In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*. 1464–1471. Las Palmas de Gran Canaria, Spain.
- Lin, C. 2004. "Rouge: A package for automatic evaluation of summaries." In *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*. 25–26. Barcelona, Spain.
- Lin, C. and E. Hovy. 1997. "Identifying topics by position." In *Proceedings of ACL Applied Natural Language Processing Conference*. 283–290. Washington, USA.
- Luhn, H. P. 1958. "The automatic creation of literature abstracts." *IBM Journal of Research and Development* 2(2), 159–165.
- Mani, I. 2001. *Automatic Summarization*. Amsterdam: John Benjamins.
- Mani, I. and M. Maybury. 1999. *Advances in Automatic Text Summarization*. Cambridge: The MIT Press.
- Mann, W.C. and S.A. Thompson. 1988. "Rhetorical structure theory: Toward a functional theory of text organization." *Text* 8(3), 243–281.
- Marcu, D. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge: The MIT Press.
- McKeown, K. and D. Radev. 1995. "Generating summaries of multiple news articles." In *Proceedings of ACM-SIGIR'95*. 74–82. Seattle, USA.
- Melčuk, I. 1988. *Dependency Syntax: Theory and practice*. Albany: SUNY Press.
- Melčuk, I. 2001. *Communicative Organization in Natural Language. The semantic-communicative structure of sentences*. Amsterdam: John Benjamins.
- Ono, K., K. Sumita and S. Miiike. 1994. "Abstract generation based on rhetorical structure extraction." In *Proceedings of the International Conference on Computational Linguistics (ACL 94)*. 344–348. Kyoto, Japan.
- Paice, C.D. 1990. "Constructing literature abstracts by computer: Techniques and prospects." *Information Processing and Management* 26(1), 171–186.
- Pollock, J., and A. Zamora. 1975. "Automatic abstracting research at chemical abstracts service." *Journal of Chemical Information and Computer Sciences* 15(4), 226–232.
- Radev, D., E. Hovy and K. McKeown. 2002. "Introduction to the special issue on Summarization." *Computational Linguistics* 28(4), 399–408.
- Riloff, E. 1993. "A corpus-based approach to domain-specific text summarisation: A proposal." In Endres-Niggemeyer, B. J. Hobbs and K. Sparck Jones (eds.). *Proceedings of the Workshop on Summarising Text for Intelligent Communication — Dagstuhl Seminar Report (9350)*. 13–17. Dagstuhl, Germany.

- Saggion, H. and G. Lapalme. 2000. "Concept identification and presentation in the context of technical text summarization." In *Proceedings of the ANLP/NAACL Workshop on Automatic Summarization*. 1–10. Seattle, USA.
- Sgall, P., E. Hajičová and J. Panevová. 1986. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Dordrecht/Prague: D. Reidel/Academia.
- Silber, H. G. and K. F. McCoy. 2000. "Efficient text summarization using lexical chains." *Intelligent User Interfaces*. 252–255.
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Teufel, S. and M. Moens. 2002. "Summarizing scientific articles: Experiments with relevance and rhetorical status." *Computational Linguistics* 28(4), 409–445.
- Toffler, A. 1970. *Future Shock*. New York: Random House.
- Vallduví, E. 1990. *The Information Component*. PhD dissertation. University of Pennsylvania, USA.

Authors' addresses

Iria da Cunha
Institut Universitari de Linguística Aplicada
(IULA)
Pompeu Fabra University
Pl. de la Mercè, 10–12
08002 Barcelona
Spain
iria.dacunha@upf.edu

M. Teresa Cabré
Institut Universitari de Linguística Aplicada
(IULA)
Pompeu Fabra University
Pl. de la Mercè 10–12
08002 Barcelona
Spain
teresa.cabre@upf.edu

Leo Wanner
Institució Catalana de Recerca i Estudis
Avançats (ICREA) and Technology Depart-
ment
Pompeu Fabra University
Passeig de Circumval·lació, 8
08003 Barcelona
Spain
leo.wanner@upf.edu

About the authors

Iria da Cunha holds a Degree in Spanish Philology from the University of Santiago de Compostela, Spain, and a Masters in Applied Linguistics from the Pompeu Fabra University (UPF), Spain. Currently, she is pursuing her PhD at the Institute for Applied Linguistics (IULA) of the UPF in the area of automatic text summarization in specialized discourse.

Leo Wanner holds a Diploma in Computer Science from the University of Karlsruhe, Germany, and a Doctorate in Linguistics from the University of the Saarland, Germany. After occupying

positions at the German National Centre for Computer Science, University of Stuttgart, University of Waterloo and the ISI, University of Southern California, he is currently an ICREA research professor at the Technology Department of the Pompeu Fabra University (UPF), Spain.

M. Teresa Cabré holds a Degree and a Doctorate in Philosophy and Letters from the Barcelona University, Spain. After occupying positions at Barcelona University, the Catalan Terminology Center, the Catalan's Studies Institute and other institutions, she founded the Institute for Applied Linguistics (IULA) of the Pompeu University (UPF), of which she was the director from 1994 to 2004. She is Professor at UPF since 1994. Currently, she is the President of the Spanish Association of Terminology.

Appendix 1: List of Extraction Rules

Given that the abstract writing instructions provided by the journal from which our articles are taken requires the abstract to contain information from each section of the article, the rules listed below apply to any of the sections — if not explicitly stated otherwise.

For the sake of a more compact and more transparent presentation of the rules, let us introduce first a number of abbreviations:

Δ_s :	global weight assigned to sentence s in the text.
δ_s :	weight increment added to/subtracted from Δ_s during the application of an extraction rule
W :	list of domain-specific prominent words
S_W :	stop word list (words of domain-specific irrelevant words)
T :	title of the article under summarization
P :	list of prominent linguistic patterns that introduce summary irrelevant information
$Sec_{s.rem}$:	section in which the sentence s occurs
$Sec_{SAT.rem}$:	section in which the satellite S of a rhetorical relation occurs
$Sec_{NU.rem}$:	section in which the nucleus N of a rhetorical relation occurs
S_L :	the predefined length (in sentences) of the summary
Δ_T :	predefined sentence weight threshold which a sentence needs to surpass in order to be chosen for inclusion into the summary

A1. Textual Rules

IF sentence s is one of the 3 last sentences of the *Introduction* section
 THEN $\Delta_s := \Delta_s + \delta_s$

IF a sentence s is one of the 2 first sentences of the *Patients and methods* section
 THEN $\Delta_s := \Delta_s + \delta_s$

IF a sentence s is one of the 2 first sentences of the *Results* section
 THEN $\Delta_s := \Delta_s + \delta_s$

IF a sentence s is one of the 3 first sentences OR one of the 3 last sentences of the *Discussion* section

THEN $\Delta_s := \Delta_s + \delta_s$

A1.2. Lexical Rules

A1.2.1. Lexical Rules Scoring Sentences

IF for sentence s the following statement holds: $\exists w: w \in W$ AND $w \in s$

THEN $\Delta_s := \Delta_s + \delta_s$

IF for sentence s the following statement holds: $\exists w: w \in T$ AND $w \in s$ AND $w \notin SW$

THEN $\Delta_s := \Delta_s + \delta_s$

IF a sentence s contains a verbal form in 1st person plural

THEN $\Delta_s := \Delta_s + \delta_s$

IF a sentence s contains any numerical information

AND s is in either in the *Patients and methods* section OR in the *Results* section

THEN $\Delta_s := \Delta_s + \delta_s$

IF |Final summary| > SL

ELIMINATE Sentences S_E from Final Summary with $\Delta_{SE} < \Delta_T$

A1.2.2. Lexical Rules Eliminating Sentences

IF a sentence s contains a reference r to numerical information

THEN IF r appears in parenthesis

THEN ELIMINATE r from s

ELSE IF $|Sec_{s,rem}| > 1$ ELIMINATE s from $Sec_{s,rem}$

IF a sentence s contains d (d being a definition or introduction) of a new concept c

THEN IF d appears in parentheses

THEN ELIMINATE d from s

ELSE IF $|Sec_{s,rem}| > 1$ ELIMINATE s from $Sec_{s,rem}$

IF a sentence s contains a reference r to a table or figure

THEN IF r appears in parenthesis

THEN ELIMINATE r

ELSE IF r is rendered in a linguistic pattern $p \times P$

THEN ELIMINATE p

ELSE IF $|Sec_{s,rem}| > 1$ ELIMINATE s from $Sec_{s,rem}$

IF a sentence s contains a reference r to previous or related work

THEN IF r appears in parenthesis

THEN ELIMINATE r

ELSE IF r is rendered in a linguistic pattern $p \times P$

THEN ELIMINATE p

ELSE IF $|Sec_{s,rem}| > 1$ ELIMINATE s from $Sec_{s,rem}$

IF a sentence s contains no verb

THEN IF $|Sec_{s,rem}| > 1$ ELIMINATE s from $Sec_{s,rem}$

A1.3. Discourse and Syntactic-Communicative Rules

A1.3.1. Discourse Rules

A1.3.1.1. Discourse Rules Eliminating Discourse Elements

A1.3.1.1.1. Discourse Rules Eliminating Satellites

IF S is satellite of a BACKGROUND relation *B*

THEN IF $|Sec_{SAT.rem}| > 1$ ELIMINATE S

IF S is satellite of a PURPOSE relation *P*

THEN IF $|Sec_{SAT.rem}| > 1$ ELIMINATE S

IF S is satellite of a JUSTIFICATION relation *J*

THEN IF $|Sec_{SAT.rem}| > 1$ ELIMINATE S

IF S is satellite of a RESULT relation *RS*

THEN IF $|Sec_{SAT.rem}| > 1$ ELIMINATE S

IF S is satellite of a CONCESSION relation *C*

THEN IF $|Sec_{SAT.rem}| > 1$ ELIMINATE S

IF S is satellite of a REFORMULATION relation *RF*

THEN IF $|Sec_{SAT.rem}| > 1$ ELIMINATE S

IF S is satellite of a CIRCUMSTANCE relation *CI*

THEN IF $|Sec_{SAT.rem}| > 1$ ELIMINATE S

A1.3.1.1.2. Discourse Rules Eliminating Nucleus

IF *N* is nucleus of an INTERPRETATION relation *I*

THEN ELIMINATE *N* AND KEEP the satellite of *I*, *S*

IF *N* is nucleus of an EVIDENCE relation *E*

THEN ELIMINATE *N* AND KEEP the satellite of *E*, *S*

A1.3.1.2. Discourse Rules Keeping Discourse Elements

A1.3.1.2.1. Discourse Rules Keeping Satellites

IF S is satellite of a CONDITION relation *CO*

THEN KEEP S

IF S is satellite of a SUMMARIZATION relation *SU*

THEN KEEP S

A1.3.1.2.2. Discourse Rules Keeping Nucleus

IF *N* is nucleus of a CONTRAST relation *CON*

THEN KEEP *N*

IF *N* is nucleus of a UNION relation *U*

THEN KEEP *N*

IF *N* is nucleus of a LIST relation *L*

THEN KEEP *N*

IF N is nucleus of a SEQUENCE relation SE
THEN KEEP N

A1.3.2. Syntactic Rules

IF the syntactic dependency structure of a sentence s contains a subtree ' $X -APPEND \rightarrow Y$ '
THEN ELIMINATE Y from s

A1.3.3. Discourse-Syntactic Rules

IF S is satellite of an ELABORATION relation E
AND there is a syntactic dependency structure such that ' $X -ATTR \rightarrow S$ '
THEN ELIMINATE S

A1.3.3. Discourse-Communicative Rules

A1.3.3.1. Discourse-Communicative Rules Eliminating Elements

IF S is satellite of ELABORATION E
AND S elaborates on the Theme of the nucleus N of E
THEN ELIMINATE S

A1.3.3.2. Discourse-Communicative Rules Keeping Elements

IF S is satellite of ELABORATION E
AND S elaborates on the Rheme of the nucleus N of E
THEN KEEP S

Appendix 2: RST Relations used for summarization

Table A1.1: RST relations used in the summarization model

Type of relation	Example
MULTINUCLEAR RELATIONS	CONTRAST <i>[Los antecedentes de primer grado se relacionan con un mayor riesgo de aparición del tumor,]_N [mientras que los antecedentes familiares de segundo grado no influyen de manera importante.]_N</i> [The relatives of the first degree are related with a bigger risk of appearance of the tumor,] _N [whereas the family relatives of second degree are not influential in an important way.] _N
	JOINT <i>[En todos los pacientes se realizó un seguimiento radiológico]_N [y fueron dados de alta tras una radiografía del abdomen sin evidencia de cuerpos extraños.]_N</i> [A radiological monitoring of all patients was carried out] _N [and they were discharged from the hospital after an X-ray of the abdomen without evidence of unknown bodies.] _N
	LIST <i>[El 68% de los pacientes eran varones.]_N [El 92% procedían de Colombia.]_N [El 65% ingirieron fármacos antidiarreicos.]_N</i> [68% of the patients were male.] _N [92% were from Colombia.] _N [65% took anti diarrhoea drugs.] _N
	SEQUENCE <i>[A todos ellos se les realizaron una historia clínica y un examen físico.]_N [Se les preguntó por el país de procedencia.]_N [Se registraron la frecuencia cardíaca, la temperatura y la presión arterial.]_N</i> [A clinical case record was compiled and a physical examination of all of them was carried out.] _N [They were asked about the country of origin.] _N [The heart frequency, the temperature and the pressure arterial were registered.] _N
NUCLEUS-SATELLITE RELATIONS	BACKGROUND <i>[A los portadores de cuerpos extraños intraabdominales que contienen cocaína, con fines de contrabando, se les conoce con el síndrome del body packer.]_N [Hemos estudiado la aparición de complicaciones en el seguimiento de individuos que ingieren estos paquetes de droga, con el fin de poder dar unas normas de actuación en estos casos.]_S</i> [Bearers of unknown bodies in the abdomen which contain cocaine for the purpose of smuggling are known by the syndrome of the body packer.] _N [We have studied the occurrence of complications in the treatment of individuals who ingest these packets of drugs with the goal to be able to provide some norms of action in such cases.] _S

- CIRCUMSTANCE [*Parece necesario propiciar algún tipo de campaña informativa para sensibilizar a la población femenina ante el cáncer de mama,*]_N [*mientras no se dilucidan las incógnitas que plantean las costosas campañas de detección temprana.*]_S
 [It seems necessary to promote some kind of information campaign to sensitize the female population with respect to breast cancer,]_N
 [as long as the unknown factors that the costly campaigns of early detection imply are not clarified.]_S
- CONCESSION [*El porcentaje de curación fue algo menor en los obesos que en los no obesos,*]_N [*aunque esta diferencia no ha sido estadísticamente significativa.*]_S
 [The percentage of recovery was somewhat smaller in the case of obese than in the case of no obese,]_N [although this difference was not statistically significant.]_S
- CONDITION [*Si los pacientes requieren un flujo superior a 3 litros por minuto,*]_S [*es probable que no sea bien tolerado y seguramente la VAO no es la mejor solución.*]_N
 [If patients require a flow superior to 3 liters per minute,]_S [it is likely that it is not well tolerated, and VAO is certainly not the best solution.]_N
- ELABORATION [*La infección por C. difficile es la causa más frecuente de diarrea nosocomial en nuestro medio y representa el 15–20% de las diarreas asociadas al uso de antibióticos.*]_N [*El espectro clínico de la infección oscila desde la colitis pseudomembranosa hasta la diarrea leve y el portador asintomático.*]_S
 [The infection by the problematic C. is in our view the most frequent cause of nosocomial diarrhoea and represents the 15–20% of the diarrhoea associated with the use of antibiotics.]_N [The clinical spectrum of the infection oscillates from the pseudo membranous colitis to the light diarrhoea and the asymptomatic bearer.]_S
- EVIDENCE [*Presentaron datos clínicos de obstrucción intestinal 11 pacientes.*]_N [*En todos ellos se observaron signos radiológicos de obstrucción.*]_S
 [11 patients showed clinical data of intestinal obstruction.]_N [In all of them, radiological signs of obstruction were observed.]_S
- INTERPRETATION [*La utilización de técnicas como el lavado gástrico, la endoscopia, la extracción manual transanal o el uso de laxantes por vía rectal para intentar extraer los paquetes aumenta el riesgo de rotura de los mismos,*]_N [*por lo que se desaconseja su uso.*]_S
 [The application of techniques such as gastric cleansing, the endoscopy, the transanal manual extraction or rectal use of laxatives in order to try to extract the packets increases the risk of their breaking,]_N
 [therefore their application is not advised.]_S

JUSTIFICATION	<p>[Se realizó cirugía en 7 pacientes (3.3%),]N [en cinco de ellos porque presentaban obstrucción, en uno por rotura de uno de los paquetes y en otro por ausencia de progresión de dos de los paquetes que eran de tamaño superior al resto.]S</p> <p>[Surgery was carried out to 7 patients (3.3%),]N [to five of them because they presented obstruction, to one because of breaking of one of the packets and to another because of the lack of progression of two of the packets that were bigger than the others.]S</p>
PURPOSE	<p>[Para que puedan cumplir su función con eficacia.]S [los SUH precisan que exista un equilibrio apropiado entre la demanda asistencial y su capacidad de respuesta.]N</p> <p>[To enable them to comply with their function with efficiency.]S [the SUH require that an appropriate balance between the demand of assistance and their reaction capacity exists.]N</p>
RESTATEMENT	<p>[Se han tenido en cuenta sólo los días efectivamente abstinentes:]N [es decir, en caso de abandono del tratamiento, se considera que todo el tiempo restante, hasta el día 180, se está consumiendo alcohol.]S</p> <p>[Only really abstinent days have been taken into account:]N [that is, in the case the treatment is given up, it is assumed that all the remaining time, until the 180th day, alcohol is being consumed.]S</p>
RESULT	<p>[Se practicó una radiografía simple del abdomen en todos los enfermos.]N [Se observaron cuerpos extraños intra-abdominales en el 98,6% de los enfermos.]S</p> <p>[A simple X-ray of the abdomen was performed on all patients.]N [Unknown intra-abdominal bodies were observed in 98.6% of the patients.]S</p>
SUMMARY	<p>[Se realizó una radiografía simple.]N [También se llevó a cabo una radiografía combinada mediante varias técnicas.]N [En resumen, se han aplicado diferentes pruebas radiológicas.]S</p> <p>[A simple X-ray was carried out.]N [A combined X-ray was also carried out via several techniques.]N [In short, different radiological tests have been applied.]S</p>

Appendix 3: Sample Texts and Summaries

This Appendix contains the first and the second section of a medical article, the author abstract and the abstract obtained by the application of the model proposed in this article.

Factores de riesgo de infección por *Clostridium difficile* en pacientes ancianos. Estudio de casos y controles.

Fundamento.

La infección por *C. difficile* es la causa más frecuente de diarrea nosocomial en nuestro medio y representa el 15–20% de las diarreas asociadas al uso de antibióticos. El espectro clínico de la

infección oscila desde la colitis pseudomembranosa hasta la diarrea leve y el portador asintomático. La administración de antibióticos constituye el principal factor de riesgo y el más conocido, aunque otros factores, como la toma de fármacos antiulcerosos o la gravedad de la enfermedad, también han sido involucrados en esta infección.

C. difficile es un bacilo grampositivo anaerobio, productor de esporas, que puede colonizar a pacientes hospitalizados, en especial a ancianos, tras tratamiento antibiótico. Entre el 7 y el 14% de estos pacientes son portadores del microorganismo, aunque raramente existe una producción suficiente de toxina A o B, exotoxinas responsables de las manifestaciones clínicas de la enfermedad. Los pacientes ancianos, por sus especiales características (inmunosenescencia, comorbilidad, fragilidad, ingresos hospitalarios frecuentes y polifarmacia) constituyen un colectivo predisuesto al desarrollo de esta infección.

El objetivo de este trabajo es estudiar los principales factores de riesgo asociados a la colitis por *C. difficile*, en un colectivo de ancianos ingresados en una unidad geriátrica de agudos.

Pacientes y método.

En marzo de 1997 se produjo, de forma explosiva, en una unidad de geriatría de 24 camas, un brote epidémico de colitis que afectó a 12 (50%) de los 24 pacientes ingresados. Se consideró que los pacientes estaban afectados por diarreas cuando presentaban tres o más deposiciones diarias blandas o líquidas durante un mínimo de 2 días, en ausencia de otra causa que lo justificara. En todos los pacientes con diarrea se practicó la determinación de la toxina A de *C. difficile* por el método de enzimoimmunoanálisis (ELFA, enzyme linked fluorescent assay, CDA2, VIDAS, bioMérieux). Por cada caso diagnosticado de colitis asociada a *C. difficile* (CACD) se recogieron de forma retrospectiva 3 controles ingresados en la misma unidad, de igual edad (± 2 años) y sexo y en la misma fecha (± 3 meses). Se cumplimentó un protocolo de recogida de datos a partir de la revisión de historias clínicas. El estudio estadístico se realizó con el programa SPSS. Para el estudio de las variables categóricas se aplicó la prueba de la χ^2 con la corrección de Yates, o la prueba exacta de Fisher cuando los valores esperados en una o más casillas eran inferiores a cinco. Para estudiar la asociación entre variables numéricas y categóricas se utilizó la prueba de la t de Student o la prueba no paramétrica de la U de Mann-Whitney cuando las variables no seguían una distribución normal. Se consideró que existía significación estadística con una $p < 0,05$. Se realizó un análisis multivariante mediante un modelo de regresión logística paso a paso que incluía las variables que resultaron significativas en el análisis univariado.

Autor abstract:

Fundamento: Estudiar los principales factores de riesgo de infección por *Clostridium difficile* en una unidad de geriatría.

Pacientes y método: Estudio de casos y controles retrospectivo.

Our abstract:

Fundamento: El objetivo de este trabajo es estudiar los principales factores de riesgo asociados a la colitis por *C. difficile*, en un colectivo de ancianos ingresados en una unidad geriátrica de agudos.

Pacientes y método: En marzo de 1997 se produjo, de forma explosiva, en una unidad de geriatría de 24 camas, un brote epidémico de colitis que afectó a 12 (50%) de los 24 pacientes ingresados. Por cada caso diagnosticado de colitis asociada a *C. difficile* (CACD) se recogieron de forma retrospectiva 3 controles ingresados en la misma unidad.