

ACQUISITION OF CAUSAL INFORMATION FROM MEDICAL DOCUMENTS IN FRENCH FOR EXPERTS

Wonsil Park, Maryvonne Abraham

ITI, Ecole Nationale Supérieure des Télécommunications de Bretagne, France
LaLIC, University of Paris4-Sorbonne

Abstract – This paper discusses how to select causal information from medical documents, in particular the information structured by verbal locutions in French. Causal relations are represented not only by verbs, *X CAUSES Y*, but also by other different information, for instance a temporal order between cause and effect, different meanings of the effect, and so on. With an example, we propose how to categorize causal relations in order to obtain much more detailed information from medical documents. This method is useful to experts to increase knowledge data base, and it could help students in Medicine to understand and organize their tasks.

Keywords – Semantic analysis of causal relations in French, Natural language processing, Contextual exploration, Internet learning, Assisted education in Medicine

I. INTRODUCTION

The interest of our research is obtaining maximal information from documents that we get everyday. Our research is constituted of two phases: extracting information on causal relations from medical documents in French and representing various consequences of relation cause-effect. French language shows a number of expressions to describe causal relations: simply in temporal order, connectors like *parce que* (because), *car*(since), adverbs like *donc* (so), *alors* (then), or verbal locutions like *provoquer* (produce), *pousser à* (push), *déclencher* (activate), etc. A causal relation is not always uniform as “X causes Y”. Natural languages are very rich in semantic representation. A verb supplies to us not only a causal relation between X and Y, but also how they react each other, and even the consequence generated from this relation: for instance “X is a cause of development of Y”, “X is a cause of generation of Y”, “X is an immediate cause of Y”, etc. We have approached to this aim by analyzing French verbal meanings, since a verb does not have a unique meaning and each meaning relies on relations between arguments and semantic invariant. This research will complete and detail causal relations.

II. CONTEXTUAL EXPLORATION SYSTEM

A text understanding processing is constructed of several procedures. The systems that we present in

this paper are developed by the Laboratory LaLIC¹. The contextual analyzing technique, developed from theoretical researches of LaLIC, allows us to search the semantic contents of texts. We will not discuss in detail about these systems and we will be content with describing them². The architecture of plat-form *FilText* (fig.1) shows the procedure from documents to three different finalities, extracting information, summarizing texts and representing information in graphs.

FilText is a semantic filtering system oriented to users. Its contextual exploration motor calls and executes the rules since an analyzed text. It is a conceptual model relying on the structuration of texts, mainly grammatical structure, because the grammar encodes a part of semantics.

The plat form *FilText* is general in the following meanings: the data base of linguistic markers and rules can be increased so that the quality of semantic filtering becomes better; it is possible to add other modules to the modules already constructed without any change

Our research is especially oriented to extracting information of causal relations from medical texts. Characteristics of the medical domain are more complicated than the mechanics. Because there are much more unknown factors which might influence on results. Symptoms, diagnostics, medication, etc. are different and various according to a patient's conditions – personal (age, genetic factors, etc.) and environmental (aliment, climate, pollution, etc.). Therefore, there should be some unhealed factors which could lead to totally unexpected results. A causal structure in the medical domain stays hypothetical until we obtain a result.

III. LINGUISTIC REPRESENTATION OF CAUSAL RELATIONS

Extracting causal information from texts obliges us several fundamental tasks, which consist to linguistic and philosophic researches. Above all, we

¹ *Langage, Logique, Informatique, Cognition*. We can consult more information about these system in the website www.lalic.paris4.sorbonne.fr/html/contexte.html

² For more information, see Desclés et alii. (1996), S. Ben-Hazez, J.-L. Minel (2000) and G. Crispino, S. Ben-Hazez, J.-L. Minel (1999). It is also available in the website of LaLIC, www.lalic.paris4.sorbonne.fr

Report Documentation Page

Report Date 25 Oct 2001	Report Type N/A	Dates Covered (from... to) -
Title and Subtitle Acquisition of Causal Information From Medical Documents in French for Experts		Contract Number
		Grant Number
		Program Element Number
Author(s)		Project Number
		Task Number
		Work Unit Number
Performing Organization Name(s) and Address(es) ITI Ecole Nationale Supérieure Des Telecommunications de Bretagne France		Performing Organization Report Number
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500		Sponsor/Monitor's Acronym(s)
		Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom. , The original document contains color images.		
Abstract		
Subject Terms		
Report Classification unclassified		Classification of this page unclassified
Classification of Abstract unclassified		Limitation of Abstract UU
Number of Pages 4		

should understand what the causality means and how a causal relation is organized. First of all, the causality is an interpretation of world that we perceive and a relation between events, a reorganized relation in a coherent way in order to give an explanation. We give a first condition to find a causal relation³.

Condition: X is a cause of Y, if Y is not chronologically situated before X and if X and Y can be described separately.

However, this condition is not enough to satisfy all descriptions of the cause-effect relation. Natural language demonstrates many ways to describe causal relations between phenomena. In French language, we can describe them simply in temporal order and reorganize with the help of linguistic elements: connectors (*donc* (so), *alors* (then), *si – alors* (if – then), *parce que* (because), *car* (since), *puisque* (because),...), verbs (*provoquer* (produce), *déclencher* (activate), *causer* (cause), *résulter de* (result from), ...), or we can emphasize and detail causal relations: adverbs (*X provoque Y immédiatement, lentement,...* “X produces Y immediately, slowly...”), negation (*X ne déclenche pas Y*, “X does not activate (or start) Y”), modality (*X peut provoquer Y*, “X can produce Y”), tenses (*X provoquerait Y*, “X would produce Y”)... These linguistic elements influence on modification of causal relations.

The causal relations represented by verbal locutions are explained in three categories: the *static relation*, which explains relations between cause and effect, for instance the identification, *X est la cause de Y* (X is the cause of Y); the *kinematic relation*, which encodes spatio-temporal relations, for instance the origin, *Y vient de X* (Y comes from X); the *dynamic relation*, describing the role of cause and the reaction of effect, for instance the starting, *X déclenche Y* (X activates Y).

We start by analyzing the Semantics of verbs which organize the dynamic relation. Because the cause is not considered as a simple origin, but it influences on the construction of effect. Certain verbs describe clearly causal relation and others need contextual information. In order to extract causal information from texts, we need more information than isolated words of the text. We can organize causal temporality as a primal criterion. Because effect cannot happen before Cause.

IV. CHRONOLOGICAL RELATION AND EFFECT MODALITY IN CAUSALITY

First of all, from the semantic analysis of French verbs, which organize dynamic relations, we have obtained three chronological relations between Cause and Effect.

1. **Immediate:** we do not perceive any important temporal intervention or time length between Cause and Effect: *créer, provoquer, produire, déclencher, causer, ...*
2. **Non immediate:** A time length between Cause and Effect is more important. *entraîner, conduire à, aboutir à, ...*
3. **Correlative:** Cause and Effect are not successive as immediate and non immediate. They move progressively on a consequent period. *pousser-à, augmenter, ...*

Effect modality represents in particular the property of effect. It means how the effect reacts to the cause.

1. Starting effect: Cause activates Effect. It can allow to observe successive consequences. *déclencher, créer, provoquer,*
2. Stopping effect: Cause stops activity of Effect. *bloquer, annuler, arrêter, ...*
3. Increasing effect: activity of Effect increases. *pousser-à, augmenter, accélérer, améliorer, ...*
4. Decreasing effect: activity of Effect decreases. *diminuer, ralentir, réduire, ...*
5. Maintaining effect: activity state of effect is stable without any modification. *maintenir, conserver, geler, immobiliser, ...*

A verb can describe a chronological relation and indicate an effect modality. A verb, such *entraîner*, describes a non immediate chronological relation between cause and effect with activating the effect. We construct the three categories of relations between the cause and the effect in the following board.

<i>Semantic relations</i>	• Static relations • Cinematic relations • Dynamic relations
<i>Chronological relations (temporal distance)</i>	• Immediate • Non immediate • Correlative
<i>Effect modalities</i>	• Starting effect • Stopping effect • Increasing effect • Decreasing effect • Maintaining effect

The causal relations, obtained from an analyzed text, will be represented in the following figure.

³ It is very hard to define the causality (causal relation) in one word. Because there are so many various descriptions and definitions about this complex notion. A. Jackiewicz discussed this problem in her Ph.D thesis [A.Jackiewicz, 1998].

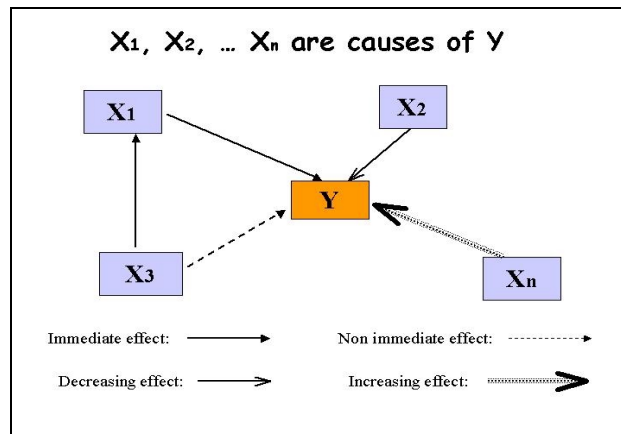


Fig.1 – Different Causal relations represented in graphs

V. EXTRACTION AND REPRESENTATION OF CAUSAL RELATIONS FROM A TEXT: AN EXAMPLE

Now, we show an example⁴, a medical text written in French. It talks about various causes which produce the problem “dysosmy”. We have marked “dysosmie(s)”, considered as the effect. The underlined words designate causal relations; *La ... cause des dysosmies est ...*, *Cette dysosmie est due à ...*, *ces dysosmies induites par ...*, etc. In the first step, we extract the sentences which contain the word “dysosmie(s)”. And then, we analyze verbal locutions.

We extract the sentences which contain the information of “dysosmie(s)”.

- Static relations

identification: *La ... cause des dysosmies est ...* (The ... cause of dysosmies is...)

inclusion: *Cette dysosmie est due à ...* (This dysosmy is due to...)

non immediate: *la dysosmie secondaire à ...* (the dysosmy following to...)

- Dynamic relations

correlative: *Plus ..., plus... augmente* (The more ..., the more... increases)

increasing: *Plus ..., plus... augmente* (The more ..., the more... increases)

We represent the causes producing the problem “dysosmie(s)” in the following graphs.

DYSOSMIES DE PERCEPTION

La première cause (30 % des dysosmies) des dysosmies de perception est la dysosmie secondaire à une rhinite aiguë. La patiente, une femme de plus de 60 ans dans 70 % des cas, constate une anosmie durant une banale rhinite. Mais, fait inhabituel, lorsque la rhinite aiguë guérit après une évolution usuelle d'une quinzaine de jours, la perte de l'odorat persiste. Dans 60 % des cas, une parosmie s'installe, soit immédiatement, soit après quelques mois d'évolution. Cette dysosmie est due à la destruction du neuroépithélium olfactif par les virus de la rhinite aiguë. C'est un diagnostic d'élimination : une maladie chronique des cavités nasosinusiennes peut débuter de la sorte. Un examen tomodynamométrique de la face doit être pratiqué : il est normal dans ces dysosmies induites par une rhinite. Moins d'un tiers des patientes ont une amélioration spontanée de leur odorat dans les mois et les années suivant l'épisode rhinitique. Il n'existe aucun traitement préventif ni curatif.

La deuxième cause des dysosmies est le traumatisme crânien. Le choc est le plus souvent occipital, plus rarement frontal. Plus l'intensité du choc est grande, plus la fréquence des dysosmies augmente. L'association d'un traumatisme facial et d'un traumatisme crânien augmente considérablement le risque de dysosmie. La physiopathologie de cette dysosmie n'est pas univoque : lésions du neuroépithélium olfactif, section des neurones olfactifs primaires au niveau de la lame criblée, contusion des centres olfactifs. Une amélioration peut survenir durant les premières semaines après le traumatisme (par diminution des lésions cérébrales fréquentes). Ces lésions olfactives posent le problème de la réparation juridique du dommage corporel. Si l'imputabilité du trouble est admise, le barème d'invalidité de ces séquelles en droit commun propose un taux de 2 % ou 3 % en cas d'hyposmie, de 5 % en cas d'anosmie avec une majoration de 1 % en cas de parosmie. Il n'existe aucun traitement.

La troisième cause des dysosmies est le vieillissement. La moitié des sujets âgés de 65 ans à 80 ans souffre d'un important déficit olfactif. Après 80 ans, 75 % des sujets sont anosmiques ou fortement hyposmiques. Il semblerait que ce trouble de l'odorat, à l'origine d'une forte hypoguesie, soit responsable du désintérêt des personnes âgées pour l'alimentation. Le déficit olfactif touche aussi bien la sensibilité olfactive (hyposmie, voire anosmie) que la perception de l'intensité d'une odeur, l'identification des molécules odorantes et la mémorisation de nouvelles odeurs. L'ensemble du système olfactif vieillit, du neuroépithélium jusqu'au cortex. Il n'existe aucun traitement.

⁴ This text is a part of an article published in www.legeneralisete.presse.fr

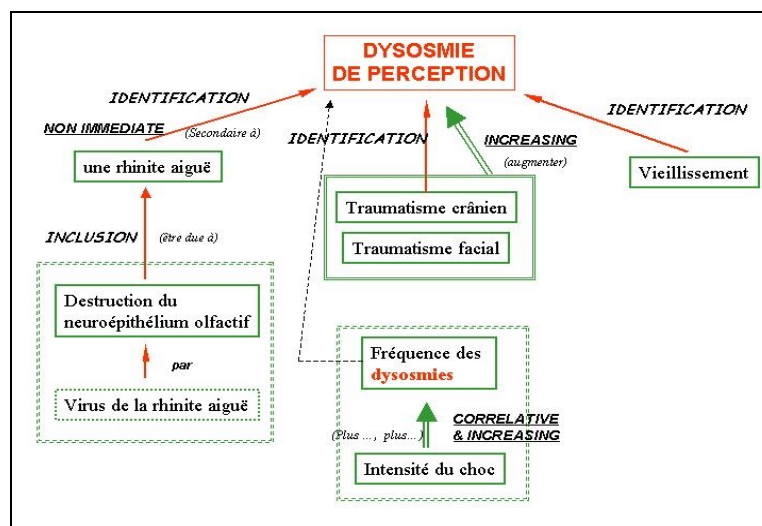


Fig.2 – Representation in graphs: various causes producing “dysosmie de perception”

This representation shows the causal relations strictly related to the problem “dysosmy”. But this text contains other information: 1) problems related to other pathologies (“anosmie”, “hyposmie”, “parosmie”), 2) statistic results (dans 70 % des cas, 30 % des dysosmies, la moitié des sujets âgés de 65 ans à 80 ans,...)

VI. CONCLUSION

In the medical domain, the causality consists to search not only general causes with a certain number of statistics, but also exceptional causes which are rare, but could be very important in certain cases.

Filtering information in a context requires a vigorous work, but it should be relatively dependant of natural languages. The causal relations are complex and even a simple sentence can supply several different relations between the cause and the effect. It is evident that this research of verbal semantic analysis is only a part of the extraction of medical causal information. To obtain the maximal information with completing the knowledge data base, it is necessary to dig in the Semantics of texts. In a context of very large internet information, our method can help to classify the knowledge data concerning specific domains in Medicine. For experts, it is useful to increase knowledge data base, and it could help non experts, students, administrators of hospital, and so on, to understand and organize their tasks.

REFERENCES

- [1] M. Abraham, *Analyse sémantico-cognitive des verbes de mouvement et d'activité, Contribution méthodologique à la constitution d'un dictionnaire informatique des verbes*, Ph.D. Thesis, EHESS, Paris, 1995
- [2] M. Abraham and W. Park, “La polysemie verbale, Méthode de representation du lexique des verbes – un exemple”, Colloque *La polysémie*, University of Paris-Sorbonne, Paris, 2000
- [3] S. Ben-Hazez, J.-L. Minel, “Designing Tasks of Identification of Complex Linguistic Patterns used for Text Semantic Filtering”, *RIAO*, 2000
- [4] G. Crispino, S. Ben-Hazez, J.-L. Minel, “ContextO, un outil du projet FilText orientée vers le filtrage sémantique de texte”, *VEXTAL*, 1999
- [5] J.-P. Desclés, *Langages applicatifs, langues naturelles et cognition*, Hermès, Paris, 1990
- [6] J.-P. Desclés, J. Berri, E. Cartier, A. Jackiewicz, J.-L. Minel, “SAFIR, système automatique de filtrage de textes”, *Actes de TALN*, 1996
- [7] D. Garcia, *Analyse automatique des textes pour l'organisation causales des actions. Réalisation du système informatique COATIS*, Ph.D. Thesis, University of Paris4-Sorbonne, Paris, 1998
- [8] A. Jackiewicz, *L'expression des la causalité dans les texts. Contribution au filtrage sémantique par une méthode informatique d'exploration contextuelle*, Ph.D. Thesis, University of Paris4-Sorbonne, Paris, 1998
- [9] Quintero, José et al. “Collaborative medical reasoning in Telemedicine”, *submitted to EMBC2001*