Morpho-syntactic Parsing for a Text Mining Environment: An NP Recognition Model for Knowledge Visualization and Information Retrieval

Sahbi Sidhom * and Mohamed Hassoun **

* Doctor-engineer. Department SII-enssib Lyon & ICom University Lyon2, e-mail: sidhom@enssib.fr
** Pr. Department SII-enssib Lyon, e-mail: hassoun@enssib.fr
ENSSIB, 17 Bd. 11 novembre 1918, 69623 Villeurbanne – France

Sahbi Sidhom is a Distinguished Engineer of the Practice of Computer Science at IRSIT-Tunis(1993), France Télécom R&D (1998), EZUS-Lyon1 (2000) and serves as an Assistant Professor at University Lyon2 (2001). He earned his doctorate at Claude Bernard University (2002), where he served as an Associated Researcher at SII Laboratory. His Principal research work is in natural language processing, machine translation and knowledge representation.

Mohamed Hassoun is a Principal Scientist and a Distinguished Professor at ENSSIB: National School of Library and Information Sciences, (1992) in Lyon, France. He is the Supervisor of SII Laboratory (1999) and the Specialized Graduate Studies in Information Networks and Electronic Documents (2001). He is internationally known for his research in natural language processing, speech understanding, and knowledge representation.

S. Siddhom, M. Hassoun (2002). **Morpho-syntactic Parsing for a Text Mining Environment: An NP Recognition Model for Knowledge Visualization and Information Retrieval.** *Knowledge Organization, 29(3/4).* 171-180. 11 refs.

ABSTRACT: Sidhom and Hassoun discuss the crucial role of NLP tools in Knowledge Extraction and Management as well as in the design of Information Retrieval Systems. The authors focus more specifically on the morpho-syntactic issues by describing their morpho-syntactic analysis platform, which has been implemented to cover the automatic indexing and information retrieval topics. To this end they implemented the Cascaded "Augmented Transition Network (ATN)". They used this formalism in order to analyse French text descriptions of Multimedia documents. An implementation of an ATN parsing automaton is briefly described. The Platform in its logical operation is considered as an investigative tool towards the knowledge organization (based on an NP recognition model) and management of multiform e-documents (text, multimedia, audio, image) using their text descriptions.

KEYWORDS: Natural language processing (NLP), knowledge extraction, knowledge classification, visualisation (knowledge network), automatic indexing, information retrieval, Noun phrase, SYDO linguistic approach, text-mining, augmented transition network (ATN).

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

1. Motivations

The diversity of the applications joined together nowadays under the term "Human Language Technology" (HLT) covers several paths of thought. Our research task consisted in marking out what is appropriate about multimedia document understanding and information retrieval in the context of "Natural Language Processing" (NLP), "automatic indexing" and "knowledge representation". However, NLP and linguistic processing are within the confluence of several disciplines: linguistics, psycholinguistics, informatics and mathematics. Thus, this work attaches a great importance to the construction of NLP Tools, which allow perfect criteria for knowledge extraction and management of conceptual information retrieval systems.

Precisely within this construction framework of the NLP tools, our Morpho-syntactic Analysis Platform was implemented to cover the automatic indexing and information retrieval topics (SIDHOM, 2002). We use the formalism of "Cascaded Augmented Transition Network" (ATN) to analyse French text descriptions of multimedia documents. An implementation of an ATN parsing automaton is briefly described. A sub-set of automata is to represent the knowledge network (*i.e.*, NP recognition model).

This Platform, logically operating, will be an investigative tool for the knowledge organisation and management of multiform e-documents (text, multimedia, audio, image), using their text descriptions.

2. Text source analysis: Multimedia Corpus

The corpus used to extract grammar rules consists of a set of French text descriptions of documents. It was assumed that it was a free text type and without constraints of style. These texts are represented by text summaries of the contents of multimedia documents. The source of our corpus consisted of bibliographic notes from INA databases (*i.e.,* Institut National de l'Audiovisuel à Paris).

Principal characteristics of text content:

The information registered in bibliographic notes from INA is varied and complementary in its content. The registered data are shared across several criteria concerning: acquisition, production, usage and document contents (chronological document analysis of INA). The last criterion, the document content, is the object of our study. We were interested in the note fields containing texts as the syntax structures. We had quoted, in particular for our study about INA notes, the following fields:

- Title: title(s) of the document,
- Brief abstract: short and general summary of the document.
- Full abstract: detailed summary of the document,
- Sequences: description set of themes by sequence in the document.
- Source abstract: summary of the author or the production company of the document.

3. Linguistic approach for analysing and indexing

The linguistic model developed by the Research Group SYDO (*i.e.*, SYstème DOcumentaire) in Lyon, France, is our central focus for descriptor reformulation; as compared to traditional problems of document indexing and information filtering. We will explain this linguistic model which accounts for the problems posed in general by the documentary information systems and, in particular, the status of descriptor in the automatic indexing process. Taking a linguistic approach to automatic indexing is another view of information system design, because it concerns the phenomenon of reference to extra-linguistic reality (universe and its objects): to distinguish between the words from language and those from speech that represents objects. According to (Bouché, 1988), "the two approaches, indexing language and taking into account a referential value, seem complementary."

The noun phrase (NP) or the nominal syntagma (SN) is indeed "the minimal unit of the speech which makes it possible to indicate an object" (Le Guern, 1989).

Indexing can be done by simple words or noun phrases. In the case of the NP or SN, they can be obtained by symbolic techniques (by labelling and filtering on syntactic patron) or by morpho-syntactic analysis. The purpose of syntagma indexing is used with a view to augment the precision of the descriptors by decreasing their ambiguities (Jacquemin, 2000).

3.1 Status of the Descriptor: Noun phrase (SN)

According to M. Le Guern, the noun phrase is an organised predicate: "the word of the lexicon, {maison}, *(i.e., house),* does not mean any house that is, whereas

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

it is enough that speech builds the syntagma {une maison}, *(i.e., a house)*, so that is indicated to be a concrete object. The closing of the predicate by the quantifier {une}, *(i.e., a)*, transforms it into a term or descriptor" (Le Guern, 1989).

The word {house} as a word of the lexicon, is considered by the author as a free predicate, which does not suppose any given universe. The lexicon relates to the words independently of the objects. The passage of the free predicate to the dependent predicate is an operation of a quantification (∀ or ∃), which consists of placing the word from the lexicon into its speech universe.

On the logical level, the human brain has the possibility of functioning according to two different systems: intensional logic and extensional logic.

Intensional logic has the characteristic of being internal logic without reference to a universe. It is the case of the operation of the lexicon in natural language. Consequently, the lexicon becomes a whole of properties (= free predicates), which are not in relation to objects.

In the speech production process, a given topic relates to not only the nature of the free predicates allotted on a subject, but also the transformations which these predicates undergo with the thread of the speech: the comparison of the predicates and their connection. Part of these inter-connected elements constitute the SN. They are the minimal units of speech (= dependent predicates) that can indicate objects of the universe. A dependent predicate is an object in a class.

Extensional logic translates predicates by "taking into account" speech. The elements of this speech are "indexing terms" or "noun phrases".

3.2 Grammar of the Noun phrase

The SN grammar is expressed by means of symbols (terminal and non-terminal) and rules. The terminal symbols represent syntactic categories. The grammar rules can use the associated variables for syntactic, flexional and lexical properties. In addition to morphological categories, some conditions are used to satisfy the rules of usage.

A set of syntactic categories is built for rewriting the rules. The left side of each rule is separated on its right side by an arrow (→), the concatenation is presented by the symbol (+).

- **VT : Terminal Vocabulary of SN**

= {

Symbol	Syntax Category
F-NOM	Noun
F-NOM-PRP	Proper noun
F-NOM-PRO	Pronoun
F-NAN	Noun or Adjective
F-ADJ	Adjective
D	Determiner
D-DEF	Definite Determiner
D-NUM	Numeral, cardinal or assimilated Determiner
D-IND	Indefinite Determiner
W-QUA	Adverb of Quantity
W-AAJ	Adverbs of Intensity (as adjective modifier)
P	Preposition
P-DE	Preposition /de/
CI, LA	words /ci/ and /là/

}.

- **VN : Non-terminal Vocabulary of SN**

= {

Symbol	Structure Category
N", N', N, A', A, D'	N" : represents a Noun Phrase, (or SN = Nominal Syntagma). N' : represents a Noun Expression. N : represents a Noun A', A and D', D : respectively to Adjective and to Determiner
EP	Prepositional Expression
SP	Prepositional Phrase

}.

- **R: Grammar rules of SN**

We use Chomsky's notation rules

(i): X" → spec(X'). X' and (ii): X' → X

= {

N° rule	Illustration	Rule
1	le + président + Chirac	**N"→** D'+N+F-PRP
2	le + président	N"→ D'+N'
3	Lui	N"→ NOM-PRO
4	Chirac	N"→ NOM-PRP
5	(la) vente de produit + à la cantine + de l'université	N'→ N+ SP^n, $n \geq 1$
6	étudiant + <u>assidu aux cours</u>	**N'→** N+A'
7	chien-ci	N'→ N+CI
8	chien-là	N'→ N+LA
9	Ministre	N'→ N
10	les + trois (candidats)	**D'→** D-DEF+D-NUM

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

N° rule	Illustration	Rule
11	de + ces (élections)	D'→ P-DE+D-DEF
12	beaucoup + de + leur (temps)	D'→ W-QUA+P-DE+D-DEF
13	peu + de (résultat)	D'→ W-QUA+P-DE
14	Le	D'→ D
15	particulièrement + fidèle	**A'→** W-AAJ+A
16	rouge + de colère	A'→ A+EP
17	Rectoral	A'→ F-ADJ,REL
18	assidu + aux cours	A'→ A + SP
19	chef + de gare	**N→** N+EP
20	drapeau + blanc	N→ N+A(QUA)
21	grand + sportif	N→ A(QUA) + N
22	Ville	N→ F-NOM
23	Fenêtre	N→ F-NAN
24	Joli	**A→** F-NAN,(QUA)
25	Impartial	A→ F-ADJ,(QUA)
26	(le directeur) de + la gare	**SP→** P+ N"
27	(le chef) de + gare	**EP→** P+N'

}.

The grammar of the noun phrase (SN) was used to support several research tasks within the SYDO group. Within the framework of our research, the grammar was increased by contextual rules of the whole sentence structures (simple or complex components) according to the logico-semantic architecture: *Figure 3.2.*

4. Morpho-syntactic Parser using the ATN Formalism

The application is based on the implementation of the Augmented Transition Network Formalism (ATN) of William Woods (WOODS, 1970-1980), (BATES, 1978), and its extension to Cascaded ATN or CATN. The structures objects generated by the ATN automata are similar to syntagmatic trees and the morpho-syntactic analysis for the sentence with its structures, modalities and components.

The construction of the sentence (which is noted PHR), according to our study, is articulated around four fundamental syntactic structures, namely: (i) a prefix structure which precedes the sentence (introductive proposition noted PI), (ii) the noun phrase (noted SN with SN_max which incorporates others SN noted SN_inc), (iii) the verb phrase (SV), and (iv) the relative phrase (REL) as an explanatory sentence for the noun or the verb phrase. Each one of these structures is identified in its elements and substructures:

PHR → [PI] + SN + [REL$_{SN}$] + SV + [REL$_{SV}$]
[*x*] *: x is an optional expression*

We also noted, some standard expressions in the corpus: SN without its determiner (zero-D) to express tiles, numeric references, and so forth.

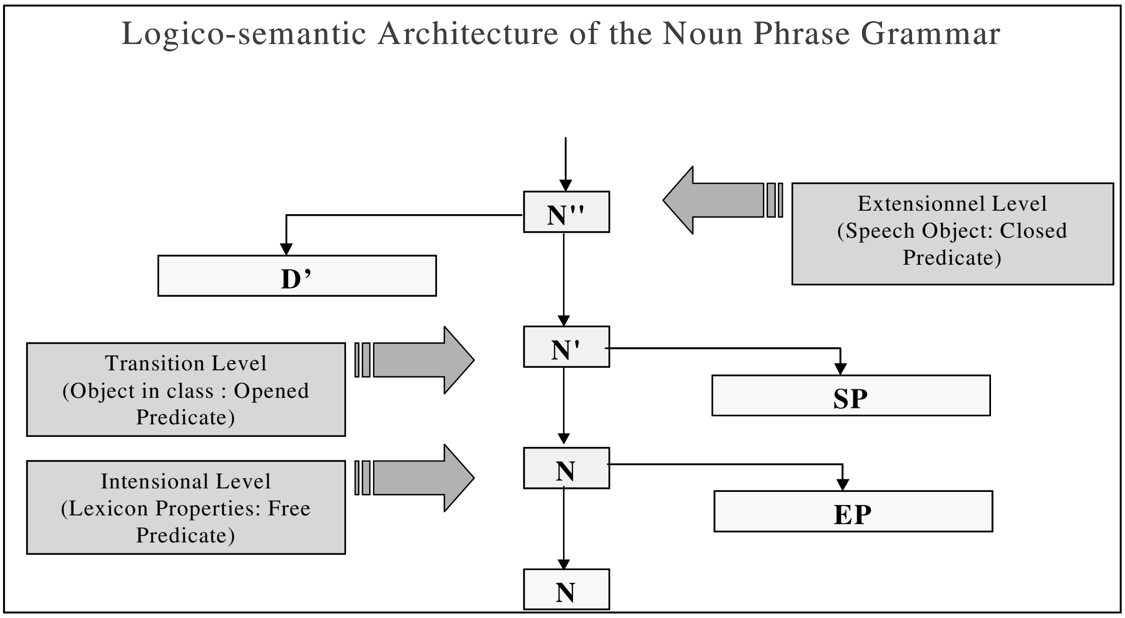

Logico-semantic Architecture of the Noun Phrase Grammar
N''
D'
Extensionnel Level (Speech Object: Closed Predicate)
N'
Transition Level (Object in class : Opened Predicate)
SP
N
Intensional Level (Lexicon Properties: Free Predicate)
EP
N

Figure 3.2. SYDO approach: logico-semantic architecture.

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

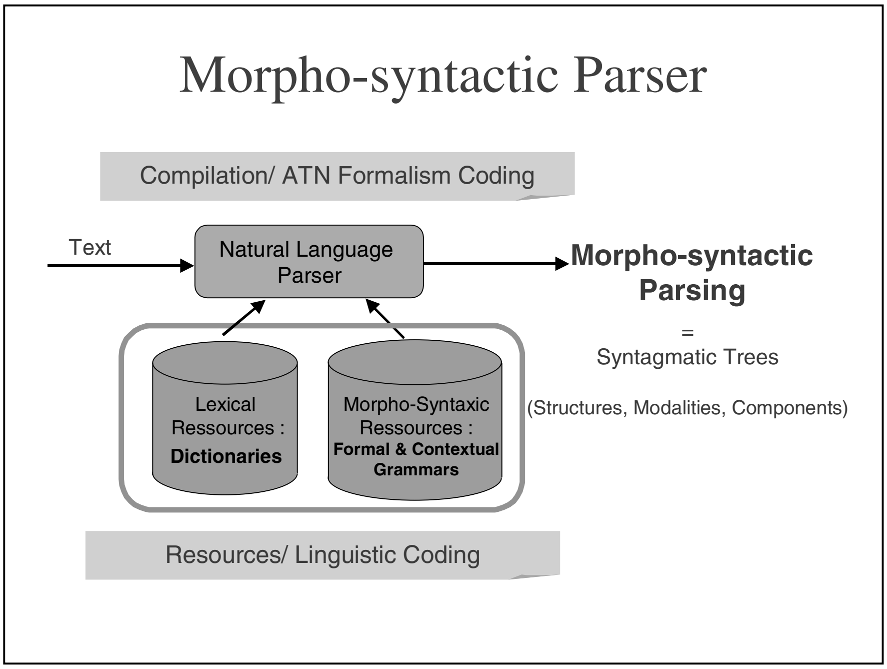

Morpho-syntactic Parser
Compilation/ ATN Formalism Coding
Text
Natural Language Parser
Morpho-syntactic Parsing
=
Syntagmatic Trees
(Structures, Modalities, Components)
Lexical Ressources : Dictionaries
Morpho-Syntaxic Ressources : Formal & Contextual Grammars
Resources/ Linguistic Coding

Figure 4.1. Parsing mechanism: Compilation and Linguistic codings.

4.1 Morpho-syntactic Structure

The morpho-syntactic model founded on the CATN formalism and extended to the context of our work (studies on corpus), integrates a robust strategy analysis that gives a good set of results, though incomplete, to parse sentences in texts.
The CATN architecture identifies gradually "well-formed" structures in sentences, and the relations between its structures. Our parser was built accordingly with formal coding compilation in C++ environment and large linguistic resources were joined according to the SYDO Model: *Figure 4.1.*

4.2 Representation of Augmented Transition Networks

The specification of a language for representing augmented transition networks is given in the form of an extended context-free grammar:

```
<ATN>             ::=  (<machinename> ( accepts
                        <phrasetype>*)<statespec>*)
<statespec>       ::=  (<statename>{optional<initialspec>}
                        <arc>*)
<initialspec>     ::=  (initial<phrasetype>*)
<arc>             ::=  (<phrasetype> <nextstate> <act>*)
                  ::=  (<pattern> <nextstate> <act>*)
                  ::=  (J<nextstate> <act>*)
                  ::=  (POP<phrasetype> <form>)
<nextstate>       ::=  <statename>
<pattern>         ::=  (<pattern>*)
                  ::=  <wordlist>
                  ::=  ε
                  ::=  --
                  ::=  <form>
                  ::=  <<classname>>
<wordlist>        ::=  {'<word>|'<word>,<wordlist>}
<act>             ::=  (transmit<form>)
                  ::=  (setr<registername> <form>)
                  ::=  (addr<registername> <form>)
                  ::=  (require<proposition>)
                  ::=  (dec<flaglist>)
                  ::=  (req<flagproposition>)
                  ::=  (once<flag>)
<flagproposition> ::=  <booleancombinationofflagregisters>
<propostion>      ::=  <form>
<form>            ::=  !<registername>
                  ::=  '<liststructure>
                  ::=  !c
                  ::=  !<liststructure>
```

4.3 Illustration by morpho-syntactic analysis

Stating the sentence to be parsed: [Elise LUCETTI donne le theme sur l'émission consacrée aux O.G.M.] *(i.e., Elise LUCETTI gives the topic of the broadcast devoted to the OGM (Genetically Modified Organism).* Its syntagmatic representation is in *Figure 4.3.* The goal of our morpho-syntactic parsing is the decomposition of the sentence in context into syntactic groups, in particular, the recognition of NPs and, inside each NP, other NPs and word roundups.

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

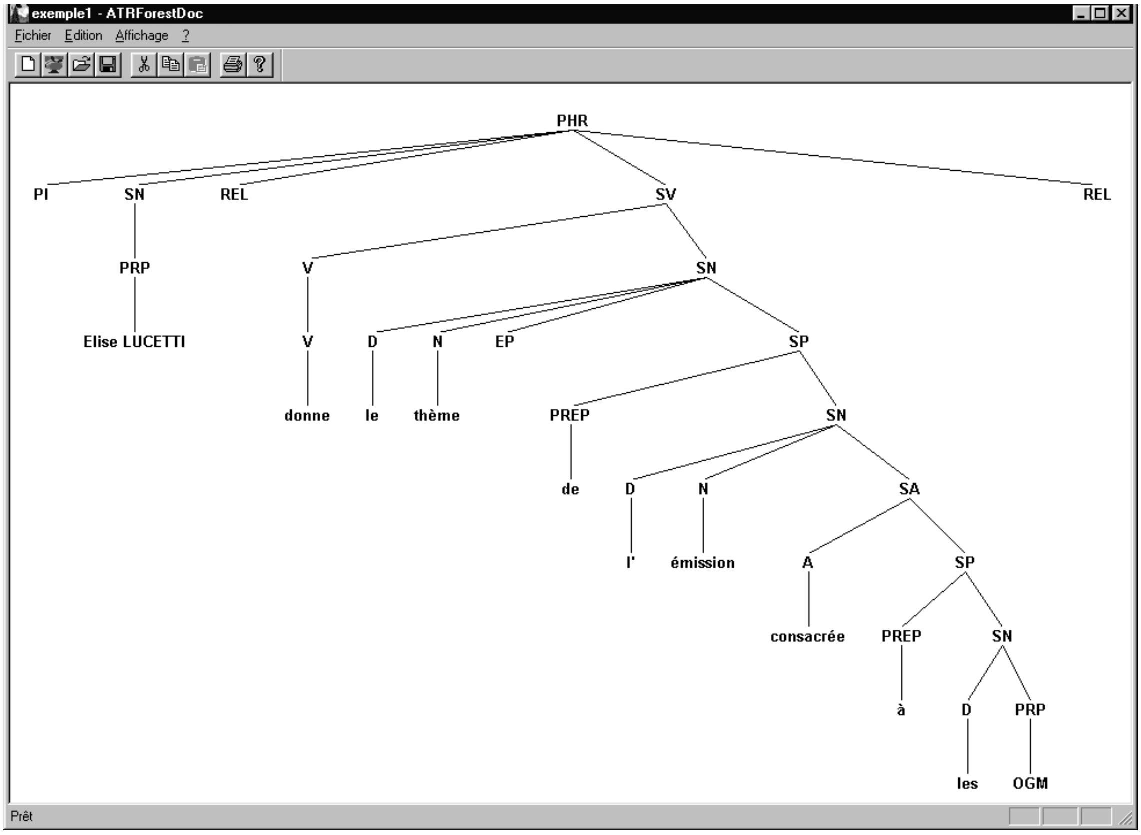


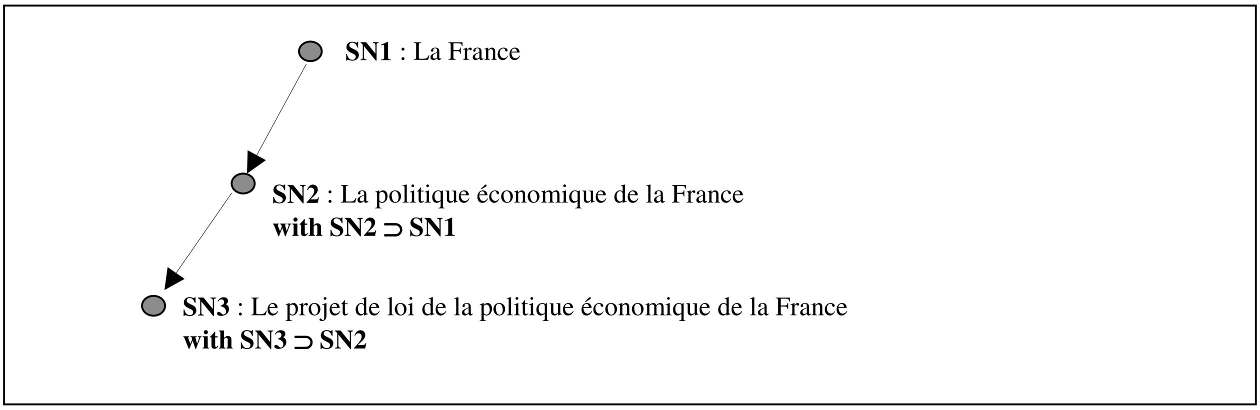
Figure 4.3. Morpho-syntactic parsing: the syntagmatic tree.

5. Noun phrase organization: Classifying process

The noun phrases had their natural organization. In a way, they had a fitting report (SN1 ⊃ SN2 ⊃ SN3 ⊃...), which makes it possible to classify SN in distinct levels: concentric information classes. At the same time, they also had an arborescence report (SN1 ⊃ SN2 and SN1 ⊃ SN3...). This last property allows us to distinguish non-concentric information classes having a common joining SN: Tree information classes. These characteristics make it possible to build a knowledge architecture based on the SN-data, by navigation in tree structures. The superposition of the SN-data with its head (= N), allows navigation in the structures between classes: lattice-data architecture.

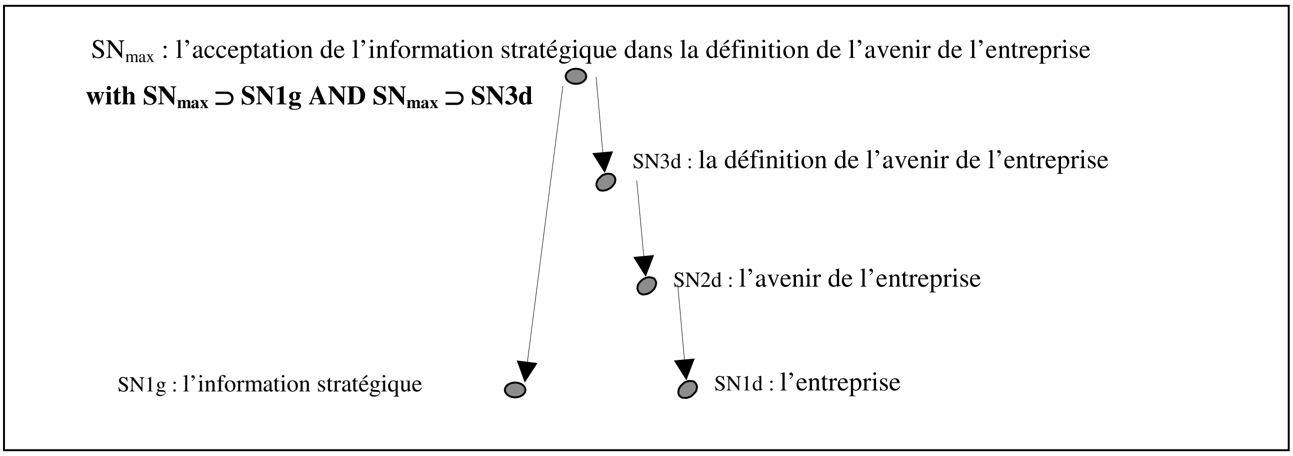
5.1 Organisation Schema in the SN:

- *Instance 1 (fitting report with SN):*
  « Le projet de loi de la politique économique de la France ».
  [Le projet de loi de [la politique économique de [la France]$^{\underline{SN1}}$]$^{\underline{SN2}}$]$^{\underline{SN3}}$


SN1 : La France
SN2 : La politique économique de la France
with SN2 ⊃ SN1
SN3 : Le projet de loi de la politique économique de la France
with SN3 ⊃ SN2

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

- *Instance 2 (arborescence and fitting report with SN):* « L'acceptation de l'information stratégique dans la définition de l'avenir de l'entreprise ». [l'acceptation de [l'information stratégique]$^{\underline{SN1}}$ dans [la définition de [l'avenir de [l'entreprise]$^{\underline{SN1}}$]$^{\underline{SN2}}$]$^{\underline{SN3}}$]$^{\underline{SNmax}}$.


SN$_{max}$: l'acceptation de l'information stratégique dans la définition de l'avenir de l'entreprise
with SN$_{max}$ ⊃ SN1g AND SN$_{max}$ ⊃ SN3d
SN3d : la définition de l'avenir de l'entreprise
SN2d : l'avenir de l'entreprise
SN1g : l'information stratégique
SN1d : l'entreprise

The levels (1 to 3) are used to indicate the extraction levels of the noun phrases. Indeed, the size of the level is inversely proportional to the extraction order:

– (l'information stratégique) ^{SN1g}	: *Level 1*
– (l'entreprise) ^{SN1d}	: *Level 1*
– (l'avenir de (l'entreprise) ^{SN1d}) ^{SN2d}	: *Level 2*
– (la définition de (l'avenir de (l'entreprise) ^{SN1d}) ^{SN2d}) ^{SN3d}	: *Level 3*
– (l'acceptation de (l'information stratégique)^{SN1g} dans (la définition de (l'avenir de (l'entreprise) ^{SN1d}) ^{SN2d}) ^{SN3d}) ^{SNmax}	: *Level 0 or maximal*

5.2 Knowledge Classifying Diagram around the SN

The differentiation of the intensional predicates (free predicate = N) from the extensional predicates (saturated predicate or not = SN) makes it possible to solve the major problem involved in knowledge extraction. To distinguish elements, which have intensional predicative properties, from others having referential functions like the noun phrases, makes it possible to provide a new logic approach (intensional and extensional logics) in the management of information system: *Figure 5.2.*

5.3 Construction of the Index File

In our application, the Index File (or index database) will be comprised of the knowledge and facts from the morpho-syntactic parsing by automatic recognition and extraction of SN and its semantic components.

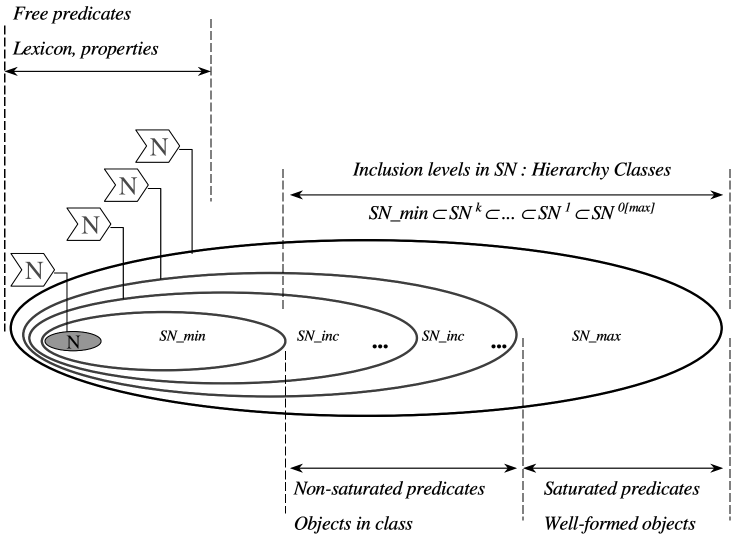


Figure 5.2. SNs' Knowledge classification.

This database consists of:

- Intensional Predicates: SN can be defined as a succession of free predicates built around the head noun (N). N will establish the link with its reference object at the same time of its instanciation (saturation). It constitutes the common feature of all SNs' class: *Figure 5.3a.*
- Non-saturated Predicates: SN is almost the topic by referring to the knowledge extracted in the text. These non-saturated SN (in level i) correspond to { SN+, SN, SN - } in {i+1, i and i-1} levels: *Figure 5.3b.*
- Saturated Predicates: SN_max (or SN+) contains all others SN of lower level. SN of this type represents the generic and complete topic in the text.

6. Visual Information Retrieval

The aim of the project is eminently practical once the noun phrases are recognized, extracted, and then stored in the index database. The list of all SN of the corpora, accompanied by each element in the list of its dependent references (SN saturated or not), and by its head noun (N), will be a fundamental tool to create the knowledge network around the SN.

Indeed, this classification gathers elements (= SN) having common characteristics. Thus, this kind of knowledge can infiltrate easily all the data containing the information. Moreover, this representation benefits its inheritance from the description between classes.

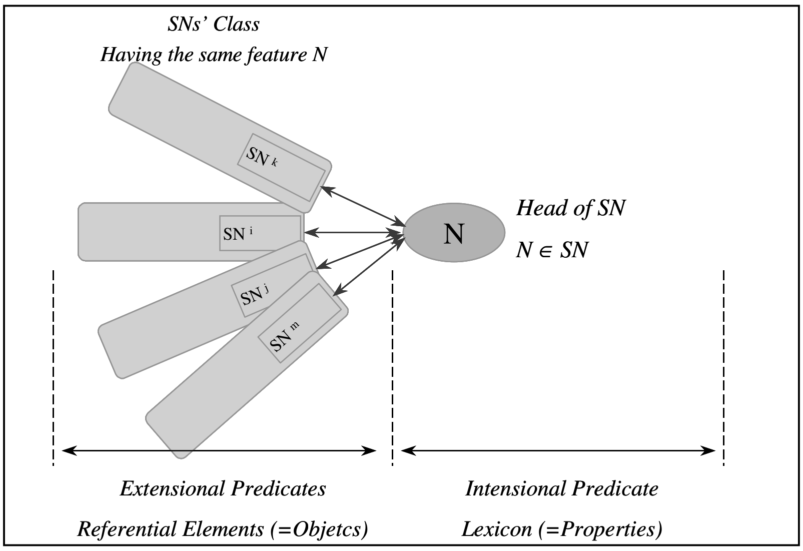


Figure 5.3a. Compositionality Relation in the SN.

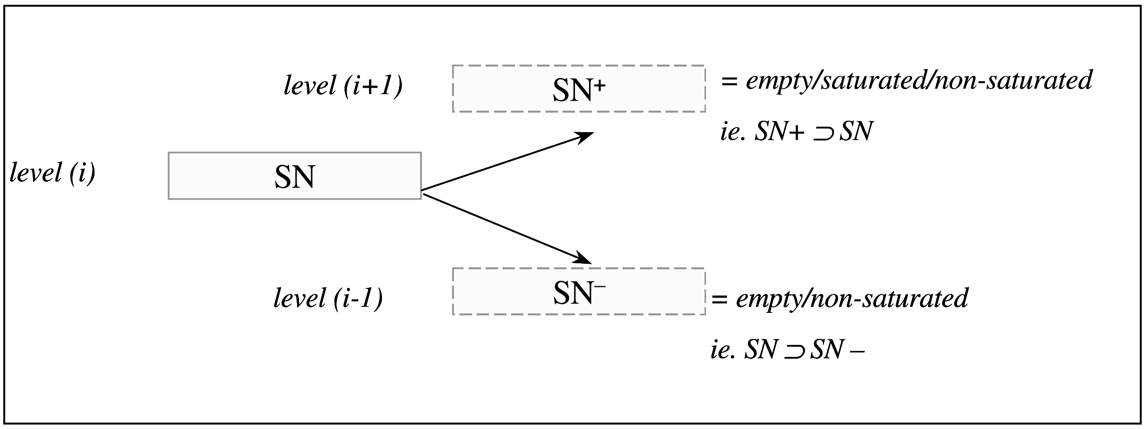


Figure 5.3b. Arborescence and linearity Relations in the SN

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

Our objective was to define an automatic classification approach resulting from the natural classification of SN, which not only makes it possible to build knowledge mapping, but also to refine the relationships between knowledge elements. At this point, we define the Classification Process for Knowledge Organisation in the following way:

- *1 Seeing that:*
  - a set of SNs and their associated descriptions (structuring),
  - knowledge (= SN) of the required classifying structure,
  - knowledge to evaluate the quality of classification (knowledge classes),
- *2 To find a classification of SN elements, i.e.:*
  - a set of classes which gather these SN elements,
  - an intensional definition of each one of these classes R(N,SN),
  - an extensional organization of these classes R(SN+,SN,SN-)
- *Visual illustration:*

We extract a set of elements in the index database. The classification represents the relationships at first (N, SN), and at second (SN+, SN, SN -). The chart (knowledge network) is created with the software AVRILECO v.3. (CNRS-irpeacs), which makes it possible to show the lattice between the various textual elements carrying on the theme "the robots": *Figure 6.*

7. Conclusion

By systematic interaction between the NLP, the knowledge extraction from text and the practical visualisation of knowledge in the information retrieval system, we managed to distinguish, at first, the natural language mechanism and its complexities, and second, the contextual element in the process of knowledge representation. Third, we brought back to

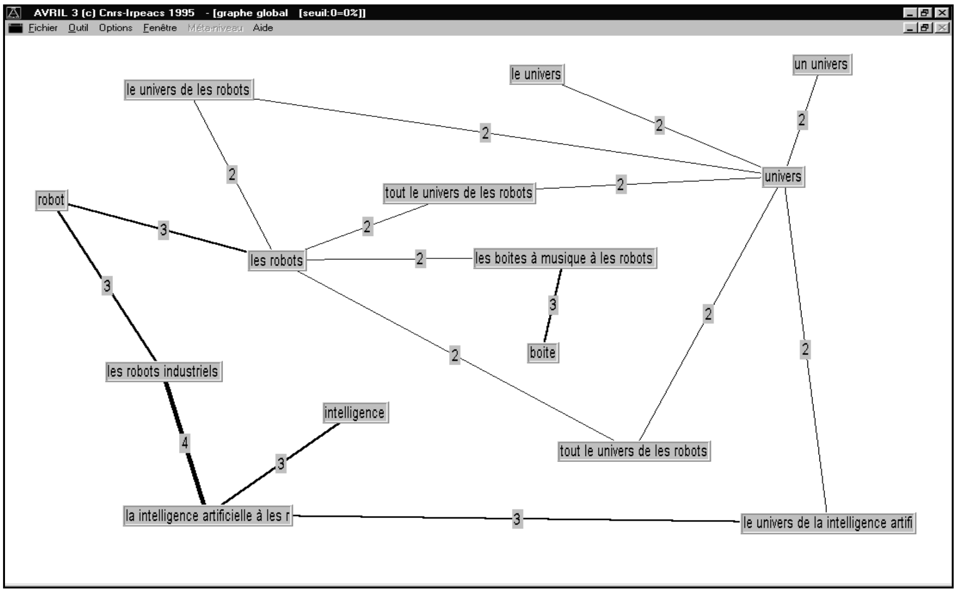

AVRIL 3 [c] Cnrs-Irpeacs 1995 - [graphe global [seuil:0=0%]]
Fichier Outil Options Fenêtre Méta-niveau Aide
le univers de les robots
le univers
un univers
univers
robot
tout le univers de les robots
les robots
les boites à musique à les robots
boite
les robots industriels
intelligence
tout le univers de les robots
la intelligence artificielle à les r
le univers de la intelligence artifi
2
3
4

Figure 6. Knowledge network around the term "robot".

N	SN
Robot	– les **robots** – les **robots** industriels
Univers	– le univers – un univers – le univers de les **robots** – tout le univers de les **robots** – le univers de la intelligence artificielle à les **robots** industriels
Intelligence	– la intelligence artificielle à les **robots** industriels
boite à musique	– les boites à musique à les **robots**
…	– …

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -

the language properties an experimental point of view relating to the natural classification of textual knowledge (SN classification).

The implemented system contains a set of facilities for semantic interpretation. The major motivation for the implementation is to explore the interaction between the syntactic and logico-semantic aspects of the text process "understanding and representation".

Several grammars have been developed and tested on our system using the CATN formalism. Facilities for lexical, morphological and syntactic analysis have been explored.

This analysis Platform in its logical operation is a tool for investigation directed towards the organization and management of knowledge.

In our research, this aspect of knowledge organization guided our aim to make emerge the linguistic properties and the natural language processing in an experimental practice of automatic indexing. We showed the need to coordinate other sources and strategies in the browsing of these properties. It is the mode of reasoning and the operating technique of the speech objects specifically to knowledge visualisation (a preliminary step in information retrieval).

These two last aspects (mode and technique) integrated in the process of the presentation and the organization of the noun phrase (SN) offer relevant scenarii for information retrieval.

Bibliography

Amar, Muriel. (1997). *Les fondements théoriques de l'indexation : une approche linguistique*. Thèse de Doctorat en Science de l'Information et de la Communication : Université Lumière Lyon 2, Lyon, France.

Bates, M. (1978). The theory and practice of augmented transition network grammars. In *Natural Language Communication with Computers* (pp. 191-259), L. Bolc, Ed., Springer-Verlag.

Bouché, Richard. (1988). Valeur référentielle et langage d'indexation. *Colloque Archives et temps réel, CEDRO-Université Lille III*, Novembre 1988, Ed. ADBS Nord.

Jacquemin, C. & Zweigenbaum, P. (2000). Traitement automatique des langues pour l'accès au contenu des documents. In *Le document Multimédia en Sciences du traitement de l'Information* (pp.71-109), Ed. CEPADUES EDITIONS, Toulouse, Editors : J. Le Maitre *alii*.

Le Guern, M. (1989). Sur les relations entre terminologie et lexique. In *actes du colloque: les terminologies spécialisés - Approches quantitatives et logico-sémantique*, et *Meta 34*, (3), (pp. 340-343). Retrieved April 24, 2003, from http://www.erudit.org/revue/meta/1989/v34/n3/.

Le Guern, M. (1994). Les Classes de Mots. In *Parties du discours et catégories morphologiques en analyse automatique* (pp. 207-215). Lyon, France: Presses Universitaires de Lyon.

Sidhom, Sahbi. Mohamed, Hassoun. & Richard, Bouché. (1999). Cognitive grammar for indexing and writing (pp.11-16). *ISKO-España Conference Proceedings*, 22-24 April 1999, Granada.

Sidhom, Sahbi. (2002). Plate-forme d'analyse morpho-syntaxique pour l'indexation automatique et la recherche d'information : de l'écrit vers la gestion des connaissances. *Thèse de Doctorat à l'université Claude Bernard Lyon1*, Lyon, France. Retrieved April 24, 2003, from http://infographie.univ-lyon2.fr/~ssidhom/recherche/THESE_Sidhom2002.PDF.

Woods, William A. (1980). Cascaded ATN Grammars. *American Journal of Computational Linguistics*, *6*(1), (pp. 1-12). Retrieved April 24, 2003, from http://acl.ldc.upenn.edu/J/J80/J80-1001.pdf.

Woods, William A. (1986). *Transition Network Grammars for Natural Language Analysis. in : Natural Languge Processing*. San Mateo, Canada : Morgan Kaufmann Ed. Retrieved April 24, 2003, from http://research.sun.com/people/wwoods/.

Woods, William A. (1997). *Conceptual Indexing : a better way to organize knowledge.* (Technical Report SMLI TR-97-61): SUN Micosystems Lab. Mountain View Canada, April 1997. Retrieved April 24, 2003, from http://sun.com/research/techrep/1997/.

https://doi.org/10.5771/0943-7444-2002-3-4-171 https://www.inlibra.com/de/agb - Open Access -