
Synthesizing a large concept hierarchy from French hyperonyms

Jean Fargues, Adeline Perrin - IBM Paris Scientific Center
3 et 5 Place Vendome, 75021 Paris Cedex 01, France

INTRODUCTION

The KALIPSOS prototype we have developed at the IBM Paris Scientific Center is able to analyze texts written in French and to produce a semantic representation of these texts expressed as a set of inter-related Conceptual Graphs [6, 1]. It uses a semantic lexicon which contains, for each word to be defined, one or more Conceptual Graphs corresponding to one or more meanings of the word. The KALIPSOS question/answering system analyzes a Natural Language query, translates it into Conceptual Graphs, performs pattern matching and deduction on these text graphs to select the answer, and finally generates a Natural Language answer from the selected answer graphs.

We do not detail this KALIPSOS system here because many papers have already been published on it (see the references). We have chosen to present recent work [5] which has been done on building a large concept hierarchy from an existing published dictionary. More precisely, we have synthesized a large semantic network by selecting hyperonym definitions from the "Dictionnaire du vocabulaire essentiel - Les 5000 mots fondamentaux" (G. Matoré, Larousse, Paris 1963) and coding them as a set of Prolog clauses available for the KALIPSOS system.

Concept type hierarchy and hyperonymy

First, we must remind you how the concept type hierarchy is the necessary basis for any use of the Conceptual Graph model. The reader may of course refer to [6]. In the Conceptual Graph model, the concept types are not supposed to be words but abstract symbols (atoms) used to denote a concept. For example, we could have COMMUNICATION-PROCESS as the concept type that occurs in the definition of the verbs "to say", "to communicate", "to discuss", etc. In the same way, a polysemic word like "bat" should point on distinct graphs containing STICK and ANIMAL as primitive concept types. It appears that the concept type hierarchy must also contain chains like:

ELEPHANT < MAMMAL < ANIMAL <
LIVING-BEING < ENTITY

Such a concept type hierarchy is necessary to define the pattern matching algorithms on Conceptual Graphs which are used to build a graph from a Natural Language sentence by joining the Conceptual Graphs of its parts. It is also necessary to encode and to verify the preference semantics constraints in the semantic lexicon. The concept hierarchy is the basis for the join and projection algorithms [3, 6] which provide a way to disambiguate the Natural Language complex sentences and to perform query/answering on Conceptual Graphs.

In the present work, we consider that concept types may generally be identified to word senses. Thus, the word "bat" points on concept types BAT.1 and BAT.2 and $BAT.1 < STICK$, $BAT.2 < ANIMAL$ is stored in the hierarchy (STICK and ANIMAL being the concept types associated to the main meanings of the words "stick" and "animal"). This implies that the synthesis of a large concept type hierarchy is related to the selection of correct hyperonyms. We give here the logical interpretation of the hyperonymy relation between the words with meanings w_1 and w_2 , derived from the one given in [4]:

w_1 is hyperonym of w_2 iff, for every sentence S

$$\text{true}[S(w_2)] \supset \text{true}[S(w_2/w_1)]$$

where:

$S(w)$ stands for a sentence containing an occurrence of w ,

$S(w_2/w_1)$ stands for the sentence $S(w_2)$ in which the occurrence of w_2 is replaced by w_1 .

\supset is the usual logical implication.

For example, ANIMAL is hyperonym of DOG because all assertions about a particular dog remain true when we substitute "the animal" for "the dog". Of course, this criterion is not always verified in a such formal way. It is only a guideline.

In a Natural Language dictionary, the Natural Language definitions may be classified into a typology, as in [4]. For example, all the definitions of the form NP VP may be hyperonym definitions, as in:

'elephant': a very large animal with two tusks and a trunk with which...

But NP VP definitions may also be meta definitions, as in:

'beget': old use to become the father of
or, as an example in French: 'être': mot qui dénote la faculté d'exister.

In this paper, we have tried to translate the definitions into English to make it easier to read, but our French dictionary (5,000 entries) uses simpler definitions than the Longman dictionary. This is the reason why the reader will not find a perfect match when referring to the Longman. Furthermore, this work depends on the particular dictionary (Matoré Larousse) we adopted but the important fact is that the result we have built is coherent and correct.

The method

The method was mainly empirical: it was not so clear that the information contained in the dictionary would be useful for synthesizing a large and coherent concept hierarchy. We will return to this important point later. But we must add that the building of a large concept hierarchy from natural language definitions has limits. For example, it cannot be a simple hierarchy but a hierarchy in which the links are labeled by conceptual relations like part-of, set-of, etc. Another limit is that the theoretical transitivity of the hyperonymy relation can only be verified on a chain of word senses if the chain is not too long. It should be noted that we were particularly interested in the top part of the hierarchy, i.e. in the list of the basic concepts from which all the others may be derived. The method, a bottom-up one, was carried out in the following stages:

1. The hyperonymy definitions were selected from the dictionary (by hand).
2. The meanings of the words, in the entries and in the definitions, were distinguished by introducing a coherent subscript notation for the current word and the main noun of its definition (by hand).
3. The relation between the current meaning of the word and its hyperonym were encoded as a Prolog clause (by hand).
4. Loops were suppressed by the application of Prolog consistency checking programs that introduced an additional synonymous relation between concepts. We mean here that when $w1 < w2$ and $w2 < w1$ are found, we declare as a Prolog clause that `SYNONYM(w1,w2)`.

5. Prolog programs were applied to the result in order to display it in a suitable way (see appendices A and B), and to have associative access to this data from Prolog.

There is a difference between simple hyperonym definitions and compound hyperonym definitions. A simple hyperonym definition has the syntactic pattern `N + VP...`, or `N + RELATIVE-CLAUSE...` In this case we choose N as the hyperonym of the current word, if it is a correct hyperonym. A compound hyperonym definition has the syntactic pattern:
`NP VP...`, where NP has the form:

1. N ADJECTIVE
2. N PREPD N (PREPD stands for 'de' 'du', 'de la')
3. Absence de N (absence of N)
4. Manque de N (lack of N)
5. Action de V (action of V)
6. Résultat de N (result of N)
7. Ensemble de N (set of N)
8. Masse de N (mass of N)
9. Groupe de N (group of N)
10. Réunion de N (union of N)
11. Fait de V (fact consisting in V)
12. Façon de V (way of V)
13. Manière de V (manner of V)
14. Possibilité de V (possibility of V)
15. État de N (state of N)
16. Art de V (art of V)
17. Quantité de N (quantity of N)
18. Liste de N (list of N)
19. Suite de N (sequence of N)
20. Partie de N (part of N)
21. Morceau de N, pièce de N (piece of N)
22. UNITÉ.I de N (unit.I of N)
23. Division de N (division of N)
24. Élément de N (element of N).

In all these cases, we keep the information contained in the NP and we code it into Prolog as follows:

- Case 1: we include the adjective in the first hyperonym and we derive a secondary hyperonym, for example:

```
Elephant: A large animal ...  
ELEPHANT < LARGE-ANIMAL <  
ANIMAL
```

- Case 2: we keep the compound noun as the first hyperonym and we generate its secondary hyperonym, for example:

```
Doute: Etat d'esprit ...  
(Doubt: State of mind ...)  
DOUTE.1 < ETAT-DE-ESPRIT.2 < ETAT
```

Jeudi: Jour de la semaine ...
 (Wednesday: day of the week ...)
 JEUDI < JOUR.2-DE-SEMAINE <
 JOUR.2

- Cases 3 to 24: we consider that the relation is a primitive conceptual relation which labels the hyperonymy link in the concept hierarchy, for example:

Infanterie: ensemble de troupes ...
 (Infantry: set of troops ...)
 INFANTERIE <
 ENSEMBLE-DE(TROUPE.1) <
 ENSEMBLE

Manche:
 1. Partie d'instrument ...
 Part of an instrument ... (handle)
 2. Partie de vêtement ...
 Part of an item of clothing ... (sleeve)
 MANCHE.1 < PAR-
 TIE.1-DE(INSTRUMENT) < PARTIE.1
 MANCHE.1 < PAR-
 TIE.1-DE(VETEMENT) < PARTIE.1

Décision: Action de choisir ...
 (decision: act of choosing ...)
 DECISION < ACTION-DE(CHOISIR) <
 ACTION-DE < ACTION

This last case implies that the result is more than a simple hierarchy: from a formal point of view it is a semantic network because of the use of primitive relations ACT-OF, PART-OF, SET-OF etc.

The result shows that there are 57 main hierarchies. We give the corresponding table containing the top concepts and the number of sons they have. In all, more than 3,600 word meanings have been coded into the network. Please see Table 1 Before the Appendices.

This result is not homogenous: some hierarchies contain many nodes and the others a few nodes. We can consider that:

- Some hierarchies correspond to fundamental types:
 - Etre (being)
 - Chose (thing)
 - Fait (fact)
 - Action (action)
 - Substance (substance)
 - Quantité (quantity)
 - Manière (manner)
 - Force (strength)
 - Son (sound)

– Feu (fire)

- Other hierarchies correspond to logical types:
 - Motif (motive)
 - Conséquence (consequence)
 - Fonction.2 (function)
 - Lien (link)
 - Manque (lack)
 - Absence (absence)
 - Élément (element)
 - Nombre (number)
 - degré (degree)
- Other hierarchies correspond to topological concepts:
 - Borne (boundary)
 - Bordure (edge)
 - Direction (direction)
 - Dimension (dimension)
 - Espace (space)
 - Intervalle (interval)
 - Contenu (content)
 - Volume (volume)
- The remaining hierarchies correspond to other isolated types and contain fewer concepts than the preceding ones. They are also pertinent but it is surprising to obtain some of them as basic generic concepts.

Another remark must be made on the transitivity of the '<' hyperonymy relation. It appears that in a chain $w_1 < w_2 < \dots < w_n$, it is possible to consider that each relation $w_i < w_{i+1}$ is justified. Nevertheless, it is more difficult to justify $w_1 < w_n$. For example, consider the chain:

TOMBE < FOSSE < TROU < OUVERTURE.2 < PASSAGE.2 < LIEU < PARTIE.1(ESPACE)

grave < pit < hole < opening < passage < location < part-of(space) (we give this translation, but it is very difficult to keep the exact nuances of the French chain).

In this chain, each contiguous relation is justified, but to justify the link between 'grave' and part-of(space) requires specifying the point of view that is taken. In fact, we have reached the limits of the process of building a concept hierarchy from existing dictionaries.

Appendix A contains an extract of the hyperonym dictionary we obtained (for the meanings of words beginning with G) and Appendix B contains an extract of the hierarchy whose top concept is CHOSE (thing).

Tops	nb of sons	Tops	nb of sons	Tops	nb of sons
'âge' (age)	4	'être' (being)	557	'absence' (absence)	11
'accélération' (acceleration)	2	'action' (action)	637	'activité' (activity)	57
'affaire' (business)	2	'autorité' (authority)	5	'avantage' (advantage)	7
'bâtiment.1' (building)	55	'bien' (good)	2	'blessure' (wound)	5
'boisson' (drink)	4	'bordure' (edge)	2	'borne' (boundary)	2
'chose' (thing)	221	'coiffure' (hairstyle)	7	'conséquence' (consequence)	714
'contenu' (content)	4	'coup.1' (blow)	3	'degré' (degree)	6
'demande' (request)	4	'dimension' (dimension)	15	'direction.1' (direction)	5
'dommage.2' (damage)	2	'effort' (effort)	5	'espace.1' (space)	93
'fait' (fact)	72	'feu.2' (fire)	2	'fonction.2' (function)	5
'force' (strength)	42	'habileté' (skill)	3	'habitude' (habit)	3
'intervalle' (interval)	62	'lien' (link)	33	'ligne.1' (line)	17
'manière' (manner)	545	'manque' (lack)	13	'motif' (motive)	4
'nombre' (number)	7	'nourriture' (food)	14	'obstacle' (obstacle)	7
'organisation' (organization)	2	'quantité' (quantity)	51	'relief' (relief)	3
'religion' (religion)	5	'remarque' (remark)	4	'renseignement' (information)	13
'reste' (remains)	2	'récit' (tale)	7	'réponse' (reply)	2
'règle.2' (rule)	11	'son' (sound)	14	'substance' (substance)	177
'volume.1' (volume)	7	'vue.3' (view)	3	'élément' (element)	663

Table 1. List of top concepts and number of sons

Appendix A

G

gain (gain)

1. somme.2 (sum)
2. quantité d argent (amount of money)
3. quantité (quantity)

gala (gala)

1. fête.1 (commemoration)
2. manifestation.1 (event)
3. action (action)

galerie (gallery)

1. passage.2 (passage)
2. lieu (place)
3. PARTIE1 DE espace.1 (PART1 OF space)
4. partie.1 (part.1)

galon (stripe)

1. bande.1 DE tissu (strip OF material)
2. bande.1 (strip)
3. morceau.1 (piece)
4. PARTIE1 DE objet.1 (PART OF object)
5. partie.1 (part)

galop (gallop)

1. allure (gait)
2. MANIERE DE aller (MANNER OF to go)
3. manière (manner)

gamin (kid)

1. enfant (child)
2. personne (person)
3. être humain (human being)
4. être vivant (living being)
5. être (being)

gamme (scale)

1. SUITE2 DE sons (SERIES OF sounds)
2. suite.2 (series)

gant (glove)

1. vêtement (item of clothing)
2. objet.1 (object)
3. chose (thing)

garage (garage)

1. bâtiment.1 (building)

garagiste (garage owner)

1. homme.3 (man)
2. être humain de sexe masculin (human being of male sex)
3. être humain (human being)
4. être vivant (living being)
5. être (being)

garantie (guarantee)

1. responsabilité (liability)
2. obligation.1 (obligation)
3. devoir (duty)
4. travail.1 (work)
5. activité (activity)

garçon (boy)

1. enfant de sexe masculin (child of male sex)
2. enfant (child)
3. personne (person)
4. être humain (human being)
5. être vivant (living being)
6. être (being)

garde (surveillance)

1. ACTION DE surveiller (ACTION OF to look after)
2. action (action)

Appendix B

chose (thing)

| chose nouvelle (new thing)

| | nouveauté.2 (novelty)

| chose vraie (true thing)

| | vérité.2 (truth)

| énigme (riddle)

| merveille (marvel)

| objet.1 (object)

| | objet.1 creux (hollow object)

| | | moule (mould)

| | | panier (basket)

| | | | corbeille (small basket)

| | | récipient (container)

| | | | récipient petit (small container)

| | | | | tasse (cup)

| | | baignoire (bath)

| | | bol.1 (bowl)

| | | bouteille.1 (bottle)

| | | | carafe (decanter)

| | | | flacon (flask)

| | | cendrier (ashtray)

| | | cuvette (basin)

| | | | lavabo (wash bowl)

| | | | évier (sink)

| | | pot (pot)

| | | poubelle (trash can)

| | | réservoir.2 (tank)

| | | seau (bucket)

| | | tonneau.1 (barrel)

| | | vase.1 (vase)

| | objet.1 fabriqué (manufactured object)

| | | instrument (instrument)

| | | | instrument DE métal (instrument made OF metal)

| | | | | cloche (bell)

| | | | instrument de musique (musical instrument)

| | | | | guitare (guitar)

| | | | | orgue (organ)

REFERENCES

[1] Natural Language Analysis Using Conceptual Graphs A. Berard-Dugourd, J. Fargues, M.C Landau Proceedings International Computer Science Conference '88 Hong-Kong, Dec. 88, pp 265-272.

[2] Natural Language Information Retrieval from French Texts A. Berard-Dugourd, J. Fargues, M.C. Landau, JP Rogala Proceedings Third Annual Workshop on Conceptual Graphs St Paul, Minnesota, Aug. 88

[3] Conceptual Graphs for Semantics and Information Processing J. Fargues, MC Landau, A. Dugourd, L. Catach IBM Journal of Research and Development Vol 30, No 1, pp 70-79.

[4] Pour une logique du sens R. Martin, Linguistique Nouvelle, Presses Universitaires de France, 1983.

[5] Elaboration d'un dictionnaire sémantique informatisé: un réseau d'hyperonymes pour la compréhension du langage naturel A. Perrin, Rapport de DEA, Centre Scientifique IBM de Paris, Sept. 89.

[6] Conceptual Structures: Information Processing in Mind and Machine J.F. Sowa, Systems Programming Series, Addison Wesley, Reading, 1984.

[7] Synonymy and Semantic Classification K. Sparck Jones, Edinburgh Press University, 1986.

[8] Discovering Relationships among Word Senses R.J. Byrd, IBM Research Report Number RC 14799, 1989